Exploring women's preferences towards ovarian cancer testing: applications of discrete choice experiments

Submitted by Rebekah Hall
to the University of Exeter as a thesis for the degree of
Doctor of Philosophy in Medical Studies
October 2022

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

Abstract

Background

Despite an increasing emphasis on shared decision-making in healthcare settings, judgements relating to diagnostic testing for cancer remain largely provider-driven. Evidence suggests involving patients in care and treatment decisions improves outcomes, boosts satisfaction and increases knowledge, self-advocacy and adherence. This thesis used ovarian cancer as an exemplar to demonstrate how discrete choice experiments (DCEs) can be used to understand preferences towards cancer testing with the purpose of improving diagnostic outcomes in primary care.

Methods

The thesis began with a systematic review of existing DCEs in the field of cancer testing. Proceeding chapters described the development and implementation and of a DCE eliciting preferences towards ovarian cancer. Barriers to ovarian cancer diagnosis mean the development of an ovarian cancer screening programme continues to be a priority, despite a lack of efficacy within clinical trials to date. In response, a further DCE investigated preferences towards ovarian cancer screening, specifically focusing on the benefit-harm trade-offs. A final component of this thesis sought to investigate methodological challenges relating to the application of DCEs in the field of cancer testing. Specific investigations included the presence and impact of indifferent preferences and stated attribute non-attendance.

Results

The systematic review demonstrated the current neglect of preferences in diagnostic settings and highlighted several methodological challenges that may limit the application of discrete choice findings to clinical and policy-related settings.

Demand for diagnostic testing was high, even when the risk of cancer was as low as 1%. Preferences appeared to centre around the trade-off between accuracy and timeliness. Although test accuracy was consistently found to be the most important attribute to respondents even where additional waiting times have a substantial impact on survival.

Preferences towards ovarian cancer screening were more heterogeneous. Latent class analysis revealed approximately half of respondents strongly prioritised mortality

reduction while remaining respondents placed low importance on this aspect of testing and instead largely focused on the high presence of false positive results.

Conclusions

This thesis demonstrates women's willingness and ability to engage in diagnostic decision-making. Preference heterogeneity further highlights the importance of an individualised approach to care. Comparisons of preferences in screening and diagnostic settings demonstrate fundamental differences, suggesting the large body of cancer screening DCEs are not automatically transferable to diagnostic settings. A greater understanding of preferences and priorities about testing in symptomatic populations is needed. In response, this thesis provides an insight into some of the challenges when conducting DCEs in this domain and offers suggestions for future researchers.

Contents

List of tables

List of figures

List of appendices

List of abbreviations

ANOVA	Analysis of variance
AIC	Akaike Information Criterion
AML	Antonieta Medina-Lara
ANA	Attribute non-attendance
AS	Anne Spencer
ASC	Alternative specific logit
BIBD	Balanced incomplete block design
BIC	Bayesian Information Criterion;
BFI-10	Big Five Inventory-10
BWS	Best-worst Scaling
CA125	Cancer Antigen-125
CAIC	consistent Akaike information criterion
CL	Conditional logit
DCE	Discrete choice experiment
GP	General practitioner
HCL	Heteroscedastic conditional logit
HCP	Health care provider
HRAS	Health Risk Attitude Scale
ISPOR	The Professional Society for Health Economics and Outcomes Research
LCM	Latent Class Model
LL	Log-likelihood
LR	Likelihood ratio
ML	Mixed logit
MMS	Multimodal screening
MNL	Multinomial logit
MRS	Marginal rate of substitution
NHS	National Health Service
NICE	National Institute for Health and Clinical Excellence
NSC	National Screening Committee
OC	Ovarian Cancer
PPV	Positive predictive value
RH	Rebekah Hall
RUT	Random utility theory
SD	Standard deviation
SDM	Shared decision-making
SE	Standard error
SG	Standard gamble
TTO	Time trade-off
TVUS	Transvaginal ultrasound
UK	United Kingdom
UKCTOCS	The United Kingdom Collaborative Trial of Ovarian Cancer Screening
USA	United States of America
VAS	Visual analogue scale
WH	Willie Hamilton

WTA	Willingness to accept
WTP	Willingness to pay

Acknowledgments

This PhD was not completed in isolation and there are a number of people I would like to thank and acknowledge for their support throughout the past few years and even further back.

First things first, I would like to thank my supervisors: Professor Anne Spencer, Professor Antonieta Medina-Lara and Professor Willie Hamilton for their expert guidance and support and being the perfect balance of personalities throughout.

Thank you to Cancer Research UK and specifically the CanTest Collaborative for the funding necessary to complete the research within this thesis and providing invaluable access to expertise in the field on primary care diagnostic research.

I would also like to thank all the people who have volunteered their time to participate in the studies within this thesis. I am particularly grateful to the women who went out of their way to take part in the qualitative aspects of this research in the height of a global pandemic.

A huge thank you to my parents for their unwavering support and encouragement. And last but not least, thank you to Alex for all that you have done and continue to do…now let's talk about something else.

1 Introduction

1.1 Chapter introduction

This chapter provides an introduction to the thesis, which demonstrates an example of how discrete choice experiments can be applied to cancer testing with the purpose of improving diagnostic outcomes. The chapter begins by describing the research problem to contextualise the motivation and purpose of the research questions addressed throughout the thesis. Next, the chapter describes how the research evolved over the course of the PhD. Finally, towards the end of the chapter the research aims and outline for the remainder of the thesis are described.

1.2 Background

Shared decision-making (SDM) is a central principle of healthcare delivery in the UK (Department of Health, 2015). There is evidence to suggest that involving patients in their care and treatment decisions improves outcomes, boosts satisfaction, and increases knowledge, self-advocacy and adherence to treatment plans (Bechel *et al.*, 2000; Fremont *et al.*, 2001; Stevenson *et al.*, 2004). In response, in recent years there have been increasing efforts to improve the collaborative process between healthcare professionals and patients in all aspects of healthcare provision, from treatment decisions during one-to-one consultations through to policy-level decision-making regarding service provision (Department of Health, 2010).

Due to the high incidence and complexity of the disease, SDM relating to the delivery of cancer care has become a key area of research. There is an increasing body of literature aiming to understand patient desires for SDM, optimising and implementation of SDM processes and current satisfaction and shortcomings (Katz *et al.*, 2014; Tamirisa *et al.*, 2017; Waddell *et al.*, 2021). When focusing on cancer care, SDM has become central component of both screening and treatment decisions following diagnosis. To date, decisions regarding investigative testing have been comparatively neglected and remain

largely paternalistic, driven by doctors (Davey *et al.*, 2004; Polaris & Katz, 2014).

Earlier diagnosis of cancer is associated with better outcomes and can reduce the high costs of complex end-stage treatments. However, all tests require a balance of benefits and harms: problems may arise when the preferences of patients misalign with the care they are offered. Misunderstanding of preferences or a mismatch between preferences and practice relating to diagnostic procedures may present a barrier to early diagnosis. Unfulfilled preferences impact the efficiency of testing by reducing help-seeking behaviours, increasing missed appointments and reducing overall satisfaction (Cronin *et al.*, 2018; Kim *et al.*, 2020). Misalignment of preferences may be particularly prevalent in instances where there is uncertainty around testing, where there are multiple investigative options available or where diagnostic guidelines are produced without the incorporation of patient views.

In the field of health economics, preference elicitation has become a prominent area of research, driven by the need to optimise the allocation of scarce resources whilst faced with increasing demand. Preference elicitation methods aim to measure patient perspectives and priorities regarding their care (Ali & Ronaldson, 2012). Findings can be used to increase the role of patients in healthcare provision on a macro-level by informing policies or approaches to care. Discrete choice experiments (DCEs) are a particularly popular preference elicitation method, due to their ability to value aspects of healthcare not easily captured using conventional quality of life measures – by allowing the trade-off between process attributes, non-health outcomes and health outcomes simultaneously, in a relatively cognitively straightforward task (Bridges *et al.*, 2011).

To date, there has been extensive research within the discrete choice literature on preferences relating to cancer screening (Mansfield *et al.*, 2016) and treatment (Bien *et al.*, 2017). Comparatively, very little attention has been paid to preferences relating to cancer diagnostic strategies (Howard *et al.*, 2011;

Ellimoottil *et al*., 2018)[1]. Screening and diagnostic procedures often employ similar test strategies but are contextually different. Asymptomatic and symptomatic individuals have different priorities and motives when seeking healthcare: therefore, preferences, particularly relating to the understanding and acceptability of risks associated with test strategies (e.g. inaccurate results, side effects, etc.) and service delivery (e.g. waiting times), may differ when thinking about diagnostic versus screening choices even when the specific tests involved remain unchanged.

1.3 Purpose of the thesis

In 2010, the Department of Health outlined plans to "introduce choice for diagnostic testing and choice post-diagnosis, from 2011" (Department of Health, 2010). Since publication, little has been done to advance the role of patient decision-making towards testing decisions in the field of cancer. Instead, decisions regarding diagnostic investigations continue to be led by doctors.

There is a substantial and growing body of evidence relating to testing in other fields such as sexual health testing, genetic testing and even cancer screening. By contrast, delayed progress in improving patient autonomy is echoed by a current lack of evidence relating to service users' preferences and priorities regarding diagnostic investigations in the field of cancer. The absence of evidence relating to diagnostic testing is reflected by recent research recommendations outlined by the National Institute for Health and Care Excellence (NICE), the body responsible for producing evidence-based guidance on healthcare provision, including guidance on investigative testing for cancer in the UK:

"studies are needed to assess the key issues in patient experience and patient information needs in the cancer diagnostic pathway, particularly in the interval between first presentation to primary care and first appointment in secondary

[1] Preferences for diagnostic testing have been explored to some extent within the wider preference literature (Banks *et al.,* 2014; Hollinghurst *et al.,* 2016). Studies demonstrated a high demand for testing using a vignette-based approach, however, this methodology does not allow for the estimation of the relative weighting of risks/benefits presented within the choice task meaning the motivators of demand were not clear.

care. Outcomes of interest are patient satisfaction, quality of life and patient perception of the quality of care and information" (NICE, 2015).

In response, the purpose of this thesis was to provide an example of how discrete choice methods can be used to understand preferences towards cancer testing with the purpose of improving diagnostic outcomes. To do so, this thesis specifically focused on preferences relating to the diagnosis of ovarian cancer as an exemplar. Ovarian cancer represents a site which has been underexplored within the preference elicitation literature to date and where early diagnosis is paramount for survival (5-year net survival is 90% when diagnosed at stage I compared to 18.6% at stage III) (Cancer Research UK, 2019).

1.4 Genesis of this thesis

The research questions and methods that form the basis of this thesis evolved substantially over the course of this PhD. The motivation for the PhD was the continued doctor-led nature of decision-making relating to cancer testing. As such, the intention to investigate the role of preferences in diagnostic testing was founded on the current evidence gap in this area and was a clear research aim from the start. The decision to utilise discrete choice methods was also made early due to the ability to measure trade-offs between attributes and implicitly measure welfare outcomes such as willingness to trade and uptake. Discrete choice experiments also benefit from a strong theoretical basis (Lancsar & Louviere, 2008) (Chapter 2). The remaining aspects of the PhD were developed concurrently as the research progressed and further adaptions were made in response to Covid-19 pandemic.

Healthcare setting—Primary care

This thesis focused on preferences towards investigative testing decisions within a primary care setting. The majority of people who are subsequently diagnosed with cancer initially present with symptoms in primary care (Hamilton, 2010). General practitioners (GPs) have an essential role of balancing who and when to investigate to minimise missed or delayed

diagnoses whilst avoiding over-testing, which may overwhelm secondary services, cause unnecessary stress for patients and exacerbate delays for high-risk patients. In practice, this judgement is challenging as symptoms of cancer are often also indicative of many benign conditions and presentation of true cancer cases is infrequent (on average a full-time GP will diagnose fewer than 10 people with cancer each year (Harker, 2017)).

The primary care diagnostic interval represents a critical period where diagnosis may be expedited or delayed (Round *et al.*, 2020). The incorporation of patient views on who, when and how to test into diagnostic guidelines may lead to more efficient and effective care resulting in increased patient satisfaction.

Selection of an exemplar cancer site—Ovarian cancer

A single cancer site was selected to provide a case study of how preference-based studies can be applied to diagnostic testing to help improve diagnostic outcomes. This approach allowed studies to be meaningfully designed around the policy questions surrounding the specific cancer type.

Several criteria were considered when selecting an appropriate cancer site during the early stages of the thesis, including:

i. Current diagnostic outcomes and the need for earlier diagnosis
ii. Current availability and/or feasibility of tests within a primary care setting
iii. Existing debates or uncertainty relating to current testing procedures e.g. competing tests, diagnostic pathway outdated, new or developing test strategies
iv. Existing evidence relating to patient and public preferences
v. Generalisability of research questions and results

In addition to ovarian cancer, several alternative cancer sites were considered and narrowed down to:

- Colorectal cancer (Tests: Faecal tests vs colonoscopy)

- Lung cancer (Tests: CT vs chest x-ray)
- Prostate cancer (Tests: Prostate specific antigen (PSA) testing plus biopsy vs ultrasound testing vs PSA & MRI)

Ultimately, ovarian cancer was selected as the focus for this thesis. Currently in the UK, over 50% of ovarian cancer cases are diagnosed at a late stage (III or IV) leading to poor survival outcomes. This emphasised the need for improved diagnostic outcomes (criterion i).

Preliminary scoping of existing evidence revealed a current lack of evidence relating to preferences for ovarian cancer testing in both screening and diagnostic contexts (unlike prostate, colorectal and lung cancer, where preference-based studies exist, albeit primarily in a screening context) (criterion iv).

Currently in England and Wales investigations of ovarian cancer in primary care involve a dual testing process, consisting of a blood test and ultrasound scanning performed sequentially (NICE, 2015). However, there are many gaps and uncertainties in the evidence surrounding the performance of tests (criterion iii) (Funston *et al.*, 2019) resulting in debates about the appropriateness of the existing diagnostic pathway. NICE guidance relating to ovarian cancer diagnosis has remained unchanged since 2011[2]. Emerging evidence means guidelines may require updating in the near future.

Many of the debates and uncertainties surrounding ovarian cancer testing were considered generalisable to the broader cancer context (criteria v). These issues are discussed in further detail throughout the thesis (see chapter 3): however, a fundamental starting point of the thesis was the patient's preference surrounding the use of triage testing (quick but less accurate test) versus more accurate tests with longer waiting times to rule out cancer. Other considerations included trade-offs between invasiveness (blood tests vs intimate

[2] NICE guidance for Lung cancer was updated in 2019 (NICE, 2019), colorectal guidelines were updated in 2017 (NICE, 2017) and prostate cancer guidelines were updated in 2021 (NICE, 2021).

examinations), convenience (GP surgery vs hospital/clinic setting) and costs to the NHS.

Finally, primary care investigation of ovarian cancer is readily available and advocated within current guidance, (NICE, 2015) allowing for an exploration of preferences surrounding demand for investigative testing in primary care where symptoms do not meet the current threshold for urgent referral threshold (i.e. at what level of risk do patients want to undergo testing and what type of tests would they ideally want?).

Identifying the target population

The debate between the role of patient or public preferences within policymaking is a longstanding question in health economics. Both have merits as well as limitations which are discussed in further detail in Chapter 2. The research in this study aims to understand preferences in order to inform service provision with the intension of improving cancer outcomes and patient satisfaction. It was considered important to understand potential barriers to help-seeking and future testing decisions based on preconceptions of people who may develop symptoms and require testing in the future. As such, a mixed population of women with and without test-experience was chosen for this thesis. Investigative testing is a discrete event rather than a health condition that requires ongoing care. Focusing exclusively on those with previous test experience. Focusing on women diagnosed with ovarian cancer (i.e. patients) was considered, however, it was decided this may lead to biased results since the outcomes of testing are likely to affect views and preferences.

The terms “public” or “general public” are used throughout the thesis when describing the samples of primary research. These terms are used as a means of differentiating between patients and respondents recruited in non-clinical settings rather than intended to indicate a representative sample of the public.
The thesis originally aimed to examine and compare the preferences of GPs alongside women. This aim was based on the important role of GPs in the investigation of cancer symptoms and the collaborative relationship required for

effective shared decision-making. This remains an important area for future research. Unfortunately, the Covid-19 pandemic meant this work became unfeasible within this thesis due to the suspension of non-Covid related research involving NHS staff during the data collection period of this PhD.

Extension to cancer screening

Revising diagnostic guidance to better align with patient preferences is one potential avenue to improve diagnostic outcomes. However, throughout the thesis it became clear that there were barriers to early diagnosis in primary care beyond inefficiencies in current testing procedures. For instance, there is delayed help-seeking due to a lack of symptom awareness from a patient perspective and delays in testing due to the vague nature of symptoms and failure to recognise symptoms on healthcare provider side.

Screening asymptomatic people may help to improve diagnostic outcomes by identifying cancers before symptoms arise. Previous attempts to identify an appropriate screening programme for ovarian cancer have been unsuccessful, however, due to the challenges surrounding early diagnosis of the disease, efforts to identify a suitable test are ongoing (Nash & Menon, 2020). Evidence relating to the acceptability and drivers of uptake for a potential screening programme are limited despite the reliance on voluntary participation. Therefore, following completion of the DCE on preferences for diagnostic testing, the scope of the thesis was extended to explore preferences towards screening programmes for ovarian cancer. This extension was motivated by the emphasis on the role of preferences in improving early diagnostic outcomes for ovarian cancer.

Methodological questions

An early finding of the thesis was despite the growth in DCE publications in recent years, the application of findings to policymaking currently appears limited (Vass & Payne, 2017). Shortcomings or inconsistencies in the application of DCE methods may be a potential barrier to the utilisation of DCE results in clinical settings. As such, this thesis also aimed to extend existing

knowledge surrounding methodological uncertainties which may currently hinder the application of DCEs in broader healthcare context. These methodological questions are discussed in detail throughout the thesis and included:

- The presence of indifferent preferences
- The impact of the number of DCE choice tasks per respondent
- Stated attribute non-attendance
- Communication of risky attributes

1.5 Aims and objectives

The overall aim of the thesis was to provide an example of the application of preference elicitation techniques, specifically discrete choice experiments to the field of cancer testing with the aim of improving diagnostic outcomes. To do so, the thesis uses the case study of ovarian cancer testing.

To achieve this, the main objectives of thesis were to:

(i) summarise and critically appraise the current evidence relating to the use of discrete choice experiments to estimate preferences in the field of cancer testing

(ii) estimate women's preferences towards different **diagnostic** strategies for ovarian cancer and assess the willingness to make trade-offs between different aspects of **diagnostic** tests

(iii) estimate women's preferences towards different **screening** strategies for ovarian cancer and assess the willingness to make trade-offs between different aspects of **screening** tests

(iv) highlight and investigate some of the methodological challenges relating to the application of discrete choice experiments to testing for cancer

Each primary objective consisted of several sub-research questions aiming to achieve a rich understanding towards ovarian cancer testing and investigate the application of DCE methods in this domain. Figure 1.1 summarises the research questions addressed in the fulfilment of the primary research

objectives. The figure contextualises the exploration of preferences in this thesis in relation to the broader diagnostic process from asymptomatic to symptomatic to final diagnosis. These research questions are discussed in detail in later chapters of the thesis.

1.6 Outline of the thesis

Figure 1.2 provides an overview of the remaining chapters of the thesis and how each chapter relates to the primary aims of the thesis. A summary of publication plans is also provided. Chapters flow from one another with earlier chapters building on from previous chapters to ensure a rigorous and thorough development process for the DCEs in this thesis.

Figure 1.1 Summary of the research questions

The research questions aim to investigate preferences at different stages along the diagnostic pathway and/or investigate methodological challenges relating to the elicitation of testing preferences

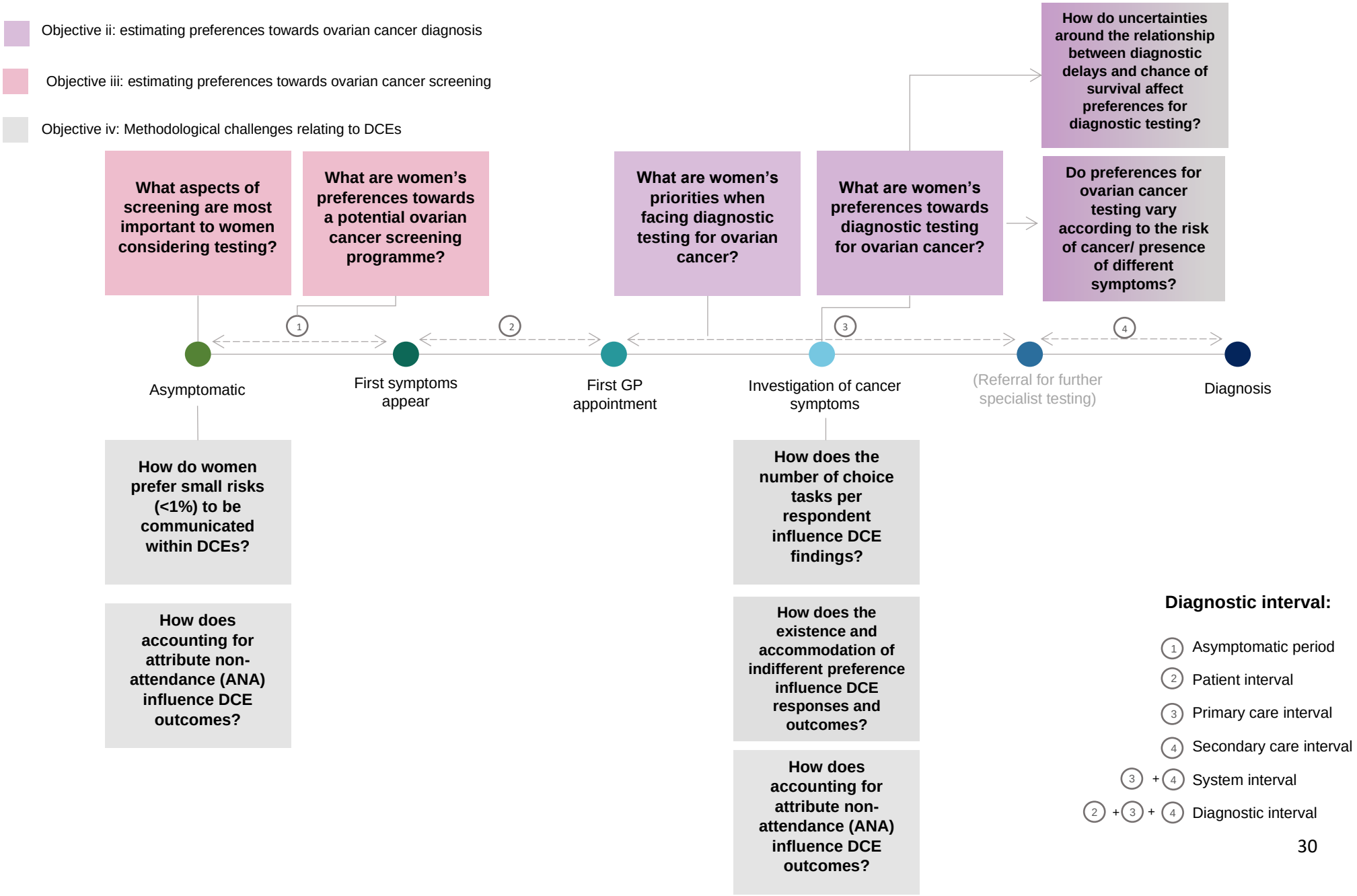


Figure 1.2: Overview of chapter aims, methods and results

Thesis objectives:			Literature	Diagnostic preferences				Screening preferences			
	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6	Chapter 7	Chapter 8	Chapter 9	Chapter 10	Chapter11	Chapter 11
Aims:	To describe the theoretical foundations of preference elicitation	To introduce ovarian cancer and the role of preferences in cancer testing	To systematically review the literature of DCEs on cancer screening and diagnosis	To identify attributes most important to women facing diagnostic testing for OC	To design a DCE to elicit preferences towards diagnostic testing for OC	To present the results the diagnostic of a DCE eliciting preferences towards diagnostic testing for OC	To present methodological extensions on indifference alternatives and task length	To describe the design process of a DCE eliciting preferences towards OC screening	To present the results of a DCE eliciting preferences OC screening	To present a methodological extension on attribute non-attendance(ANA)	To summarise and discuss the findings of the thesis
Methods:			Systematic review	Best worst scaling (BWS), online workshops	Experimental design generation, survey design, think aloud interviews, quantitative piloting	Data analysis-logistic regression models	Data analysis-logistic regression models	BWS, experimental design generation, survey design, think-aloud interviews, quantitative piloting	Data analysis-logistic regression models	Data analysis-logistic regression models	
Results:			• Fifty-two studies were identified, 49 related to preference for cancer screening and only 3 related to diagnostic testing • No studies relating to testing for ovarian cancer were identified	• Women demonstrated a willingness and ability to express preferences toward diagnostic testing • Choices were primarily driven by the trade-off between accuracy (i.e. test sensitivity) and time to diagnosis • Preferences towards testing did not vary based on level of cancer risk as indicated by the severity of symptoms • Demand for testing was high even at low risk levels demonstrating the value of primary care testing • The introduction of survival further strengthened the focus on accuracy vs waiting times • There was substantial preference heterogeneity but sociodemographic associations were limited.	• DCE estimates do not appear to vary based on the inclusion or exclusion of an indifference alternative • Exclusion of indifference responses may have detrimental impact on efficiency of experimental design. • No differences in error variance, MRS estimates or irrational responses were identified between earlier and later choice tasks within the experiment.	• BWS results showed a clear prioritisation of test performance characteristics (e.g. mortality reduction, accuracy) over service delivery characteristics (e.g. location, waiting times, staff attitude). • Mortality reduction was the most important characteristic overall. • Screening decisions largely appeared to depend on the trade-off between false positive results and ovarian cancer deaths avoided. Results estimate a maximum screening uptake rate of 66% even when a 100% reduction in mortality is achieved. • 19% of respondents indicated they would undergo screening based on current test performance despite no improvement in mortality	• Rates of self-reported ANA were high in both screening and diagnostic settings (~70%) • Estimated rates of ANA were higher when respondents were asked which attributes they “considered” rather than those they “ignored” • Adjusting for stated-ANA had no impact on aggregate MRS estimates	

Chapter 2 describes the theoretical foundations of preference elicitation within health economics and provides an overview of the leading methods. This chapter discusses the motivations for the choice to use discrete choice experiment as the primary method for this thesis and provides an introductory summary of the key components of DCEs providing a foundation which is expanded throughout the course of the thesis.

Chapter 3 provides background on ovarian cancer as the chosen case study for this thesis. The chapter highlights current challenges to early diagnosis of ovarian cancer and describes how a greater understanding of women's may help to improve diagnostic outcomes is both a diagnostic and screening setting.

Chapter 4 addresses the first aim of the thesis and presents the results of a systematic review of DCEs on cancer testing. There are several existing reviews on the topic: therefore the chapter begins by summarising existing systematic reviews in an overview of reviews to summarise existing evidence and identify any potential gaps. A systematic review focusing on methodological aspects of cancer testing DCEs is then conducted.

Chapter 5 describes the attribute development process for the DCE relating to preferences towards diagnostic testing of ovarian cancer in primary care. The chapter utilises a multi-method approach by combining evidence from the systematic review in chapter 4 with a best-worst scaling questionnaire and interactive online workshops with the target population to identify the most relevant and important aspects of ovarian cancer testing. The chapter also describes the process of assigning levels to the attributes selected for inclusion.

Chapter 6 describes the remaining DCE development process for the diagnostic DCE. The chapter draws upon leading guidance to create an effective DCE design embedded within an online survey. The chapter describes the development process beginning with defining the research questions, through to DCE development, survey design and piloting. The chapter utilises a number of technical methods including survey design and programming, efficient experimental design generation and cognitive interviewing.

Chapter 7 describes data collection, analysis and results of the final version of the diagnostic DCE designed in Chapters 5 and 6. DCE diagnosis results were used to calculate several post-estimation measures to aid interpretation and contextualise results within policy-related debates. These measures included willingness to trade between attributes, relative importance scores and demand for testing.

Chapter 8 describes the findings of two methodological extensions applied to the DCE of diagnostic preferences. These methodological extensions focused on the influence of including an indifference alternative on DCE outcomes and how the number of choice tasks completed affects estimates based on the potential for learning and fatigue effects as DCE tasks progress.

Chapter 9 describes the design of a DCE to investigate women's preferences towards ovarian cancer screening. The development process builds on methods used in earlier chapters when developing the diagnostic DCE. This allowed descriptions of the development process to be streamlined, whilst still maintaining transparency in reporting. Due to the extensive presence of risk-based attributes this chapter pays close attention to risk presentation during the development process.

Chapter 10 presents the results of the final data collection for the DCE on ovarian cancer screening preferences. The DCE focuses on aspects of test performance and results were used to estimate the willingness to trade between the benefits (lives saved) and harms of cancer screening (e.g. false positive results). This chapter also builds on Chapter 8 by providing further analysis on self-reported attribute non-attendance as related to DCEs on cancer testing.

Chapter 11 is the second methods focused chapter. The chapter combines evidence from both the diagnostic and screening DCEs relating to attribute non-attendance. The chapter investigates rates of self-reported attribute non-attendance, how stated non-attendance influences model estimates and how the framing of stated attribute non-attendance questions influences findings.

Chapter 12 consolidates the evidence presented throughout the thesis and discusses the implications of the results. The chapter then outlines the strengths and limitations of the thesis and highlights areas for future research.

1.7 Ethical approval

A large portion of the research activities within this PhD involved primary data collection. This meant ethical approval was required. Ethical approval for the primary data collected as part of the research conducted in this thesis was covered by two separate applications:

i. Ethical approval for the online surveys with embedded best-worst scaling studies conducted in chapter 5 and 9 as part of the attribute selection process was granted by the University of Exeter Business School (UEBS) ethics committee in collaboration with a colleague from the University of Exeter Economics department (Application number: eUEBS003725). The UEBS ethics committee was identified as the most appropriate based on the strong methodological foundations of the survey methods in the business field (e.g. marketing and economics). The approval certificate for the application is provided in Appendix 1.1.

ii. Ethical approval for remaining primary research conducted in chapters 5-7 and 9-11 was granted by the University of Exeter Medical School Research Ethics Committee (UEMS REC) (Application number: 20/09/261). This included quantitative and qualitative piloting activities as well as final data collection using an online survey instrument with DCE questions embedded. The approval certificate for the application is provided in Appendix 1.2.

All study plans were subject to peer review prior to submission of the ethics application and steps were taken to mitigate any of the potential risks associated with the planned research.

Online workshops conducted as part of the attribute development process in Chapter 5 were not subject to ethical approval as they were classified as Patient and Public Involvement and Engagement (PPIE) activity. Sessions were not recorded and are not intended for publication beyond inclusion in this thesis. The format of sessions and any ethical considerations (e.g. consent, right to withdraw, recruitment, renumeration) were designed based on consultation with PPIE colleagues at the University of Exeter.

1.8 Note on terminology

The thesis aimed to elicit the preferences of those who may be offered testing for ovarian cancer in the future (i.e. people with ovaries). Based on the advice received during the peer-review process conducted during the ethical approval application, all participant facing documentation referred to "people with ovaries" to ensure inclusivity of language. To maximise inclusivity whilst maintaining relevance to the research topic, participation in online surveys was limited using a biological sex filter ("female") but was not limited by gender. No information was collected on the gender of respondents during the survey; however, at the time of data collection the pool of eligible participants (~8,000) included fewer than 25 transgender males or non-binary people[3].

Study instruments and all participant-facing documents were developed in consultation with cisgender women. Existing research on uptake of gynaecological services including cancer screening and diagnostic testing is an ongoing and currently limited area of research. Existing evidence suggests the barriers and facilitators of access to healthcare differ between cisgender patients and those whose gender identity does not conform with their sex assigned at birth (Dhillon *et al.*, 2020; Gatos, 2018; Price *et al.*, 2019; Seay *et al.*, 2017). Therefore, it would be misrepresentative to assume the findings of this thesis are transferable to these populations. For these reasons, the word "women" is used throughout this thesis. It is acknowledged there is a very small chance that this may misgender a very small proportion of the sample.

[3] For privacy reasons Prolific does not provide an exact number of matching participants when the pool is lower than 25

Studies relating to the barriers and facilitators for access to gynaecological care (including investigative testing) for transgender and non-binary people is an important area for future research.

2 Preference elicitation methods: why and how preferences are elicited in healthcare

2.1 Introduction

The aim of this chapter is to provide an overview of the use of preference elicitation methods in healthcare. The chapter begins by providing an explanation of the importance of understanding preferences when considering healthcare provision. Next, a brief summary of current practices and ongoing debates is introduced, including an overview of methods that may be used to elicit preferences in healthcare. The later sections of the chapter focus on the primary elicitation method used throughout this thesis, discrete choice experiments. The chapter explains why discrete choice experiment methods are most appropriate for the research aims of this thesis before providing an introductory overview of the theoretical basis and key components of the methodology.

2.2 Role of preferences in healthcare

Preferences can have an impact on the satisfaction gained from an acceptance of health inventions and services, which in turn impacts outcomes. Increasingly rapid advances in technology have led to a growing number of complex treatment options, often indistinguishable when considering efficacy alone (Mühlbacher & Johnson, 2016). Competing treatment options use different health system resources and are associated with a unique set of risks and benefits, leading to differences in costs and patient burden. As a result, treatment decisions require a trade-off between the risks and benefits of each procedure: the importance of each varies between individual's experiences, values and priorities (Ostermann *et al.*, 2017). Increasing the involvement of patients and the public in healthcare decisions results in more responsive services leading to improved adherence and self-advocacy, resulting in improved health outcomes and increased patient satisfaction (Armstrong *et al.*, 2013; Florin & Dixon, 2004; Kleij *et al.*, 2017).

For this reason, the incorporation of patient and public preferences is crucial to health policy. In the UK, efforts to incorporate preferences are encouraged at every level of healthcare provision, from an individual level where patient-centred care and shared decision-making are central principles (Department of Health, 2015) through to the planning and commissioning level, where preferences are directly incorporated through the inclusion of patient experts on NICE decision panels and implicitly cost-utility analyses. Unfortunately, preferences have been frequently demonstrated to be misunderstood or diminished by healthcare providers and decision-makers (Mulley *et al.*, 2012). Inaccurate measurement or misunderstanding of health users' preferences may lead to inefficiencies and reduced quality in health service delivery (Kleij *et al.*, 2017; Mulley *et al.*, 2012). Therefore, it is necessary that healthcare decisions are motivated by preferences generated from robust, theoretically underpinned methods.

2.3 Current debates

2.3.1 Process vs outcome utility

Traditionally, particularly in relation to economic evaluations of health interventions typically limit the valuation of benefits to the health outcomes only (Drummond *et al.*, 2015) - an approach grounded in the neo-classical framework of welfarism (Hurley, 1998). Welfarism is the normative approach to economics which seeks to achieve the most socially desirable outcomes based solely on the maximisation of utility obtained by individuals under the principle of Pareto efficiency (i.e. a scenario where it is not possible to improve one person's situation without making someone else worse off) (Brouwer *et al.*, 2008). Consequentialism, a key tenet of welfarism, states utility is only derived from the consumption of goods and services, therefore outcomes generate utility but the processes that lead to such outcomes are neutral (Hurley, 2000).

Extra-welfarism moves beyond Pareto efficiency, allowing for considerations beyond the maximisation of individuals' utility. In theory, extra-welfarism allows for an expansion of the evaluative space in endless ways allowing for the inclusion of factors beyond utility that may be important to social as well as individual welfare (Culyer, 1989; Sen, 1993).

The enduring emphasis on outcomes as the main (or only) benefit of the healthcare interventions ignores the potential importance of the process of care in determining the value of interventions to individuals (Mooney, 1994; Mooney & Lange, 1993). “Process utility” may cover any aspect of the health intervention that influences an individual’s utility function but is not considered a final health outcome, for example waiting times or gender of the provider (Birch *et al.*, 2003). Any aspect of an intervention that affects an individual’s utility will have important consequences for their choices, particularly where outcomes may not differ between competing interventions. It is therefore important that the valuation of healthcare interventions incorporates elements of process where these differ between the alternative interventions under consideration, especially for cases such as diagnostic testing where participation and compliance is reliant on the voluntary actions of recipients (Brouwer *et al.*, 2005; Donaldson & Shackley, 1997; Howard *et al.*, 2008).

2.3.2 Whose preferences matter?

The provision and delivery of healthcare involves the interaction between multiple agents including patients, carers, health professionals, funders and policy makers, all of whom have their own set of preferences. Whose preferences to consider when allocating healthcare resources remains an area of ongoing debate (Dolan, 1999).

The welfarist approach, favours the concept of individualism, viewing patients as best placed to make value judgements relating to healthcare. When making policy decisions, it is the patients who will directly benefit or suffer as a result, it is therefore reasonable to argue policy decisions should be based on the direct experiences of such individuals (Brouwer *et al.*, 2008). Furthermore, patients are viewed as experts of their own experience meaning their preferences are likely to be well-formed and based on actual experiences even when responding to hypothetical scenarios, as required by most elicitation methods (discussed further in section 2.5) (Brazier *et al.*, 2005). Indeed, studies have demonstrated the preferences of individuals with direct experience of the

disease or intervention of interest appear more stable or well-formed when compared with the general public (Neuman *et al.*, 2010).

Opponents of welfarism have criticised the reliance on patient preferences on several grounds. Ethically, it may not be appropriate or possible for patients currently living with conditions, many of whom may be very unwell, to complete complex preference elicitation tasks, especially those which involve valuing scenarios incorporating death (Ubel *et al.*, 2003). Additionally, patient responses may not be entirely accurate or reliable. For instance, patients intentionally provide strategic responses to ensure continued access to resources or unintentionally provide skewed responses due to adaptation and acceptance of their current situation (Menzel *et al.*, 2002).

Many argue resource allocation using general population preferences may be fairer, particularly for countries such as the UK, where healthcare is publicly funded through taxation (Brazier *et al.*, 2005). When specifically considering preferences within economic evaluations, leading experts (Weinstein *et al.*, 1996) and decision-making bodies (NICE, 2018), advocate the use of general population values, citing the 'veil of ignorance' argument, which postulates aggregated general public responses to elicitation tasks are not biased by self-interest since there is no vested interest in any particular disease or intervention (Weinstein *et al.*, 1996). However, lack of self-interest is of course gained at the expense of first-hand experience, increasing the likelihood of hypothetical bias (i.e. responses that do not reflect the choices if faced in real life) within observed choices during elicitation tasks (Hensher, 2010).

2.4 Defining preferences in economics

Economists assume that when making decisions between competing goods or services (including health interventions or even health states), individuals' decisions are driven by one objective – utility maximisation (Morris *et al.*, 2012). Utility refers to the level of satisfaction an individual gains by consuming a good or service, it is a subjective concept driven by individual preferences (i.e. a greater preference for a good results in a higher utility gain from its

consumption) (Hensher *et al.*, 2005c). Utility is a latent concept meaning it cannot be directly observed or measured, however, choices made between alternative options can reveal the underlying utility an individual associates with each alternative – this is the basis of utility-based preference methods (Samuelson, 1938).

2.4.1 Choices, preferences and utility

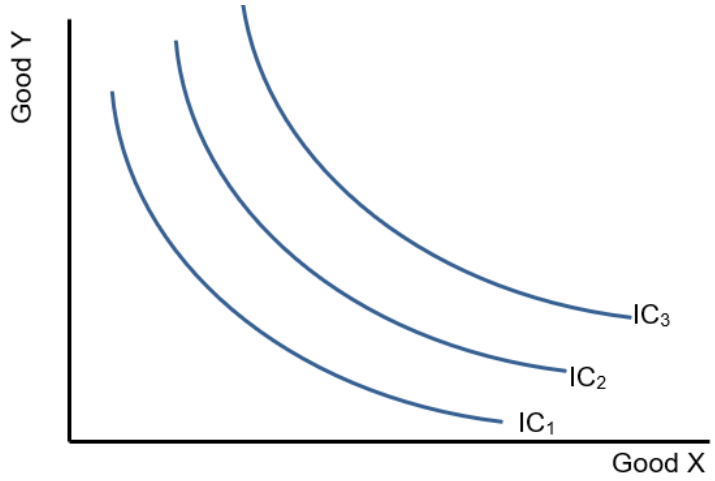
The relationship between utility, U and consumption of goods, X is represented by a utility function:

$$U= U(X_1, X_2,\ldots X_n)$$

The utility assigned to each good is subjective and unique to each individual based on their preferences.

Choice problems between alternative goods and services can be represented diagrammatically using indifference curves. Indifference curves show the combinations of goods where consumption yields the same level of utility. The shape of indifference curve (i.e. steepness/convexity) is determined by preferences. Figure 2.1 shows an indifference map, a series of indifference curves where increased distance from the origin indicates higher levels of utility. The example shows just two goods for simplicity, however, the number of goods and combinations of goods may be infinite (Morris *et al.*, 2012).

Figure 2.1: Indifference map



2.4.2 Marginal rate of substitution

Marginal rate of substitution (MRS) refers to the amount of one good that would be required as compensation for a one-unit loss of another good in order to maintain the same level of utility. MRS is represented by the slope of the indifference curve. Indifference curves are assumed to be convex as a consequence of diminishing marginal utility, where the additional utility gained decreases as consumption of one good increases (Good X). Conversely, as consumption of the other good (Good Y) decreases, the loss in utility from additional sacrifices increases (Morris *et al.*, 2012).

2.4.3 Axioms of choice/preference

Economic theory assumes preferences of individuals are rational in accordance with four axioms of Ordinal Utility theory (Hausman, 2011):

1. Completeness- individuals are able to establish a preference ordering when comparing bundles of goods i.e. individuals are always able to establish whether they prefer or are indifferent to competing consumption options
2. Continuity- the preferences of individuals are continuous (i.e. there are no jumps in people's preferences)
3. Transitivity- For any consumer if A ≥ B and B ≥ C then A must always be preferred to C. i.e. consumers are consistent with their preferences
4. Non-satiation- Individuals always place a positive value on more consumption

2.5 Preference elicitation methods

Wide scale and formal incorporation of patient and public preferences in healthcare decision-making requires the development of robust and reliable methods of preference elicitation.

Figure 2.2 gives a simplified view of the main methods used to measure preferences within healthcare. Classifications within Figure 2.2 are adapted from Ali and Ronaldson (2012).

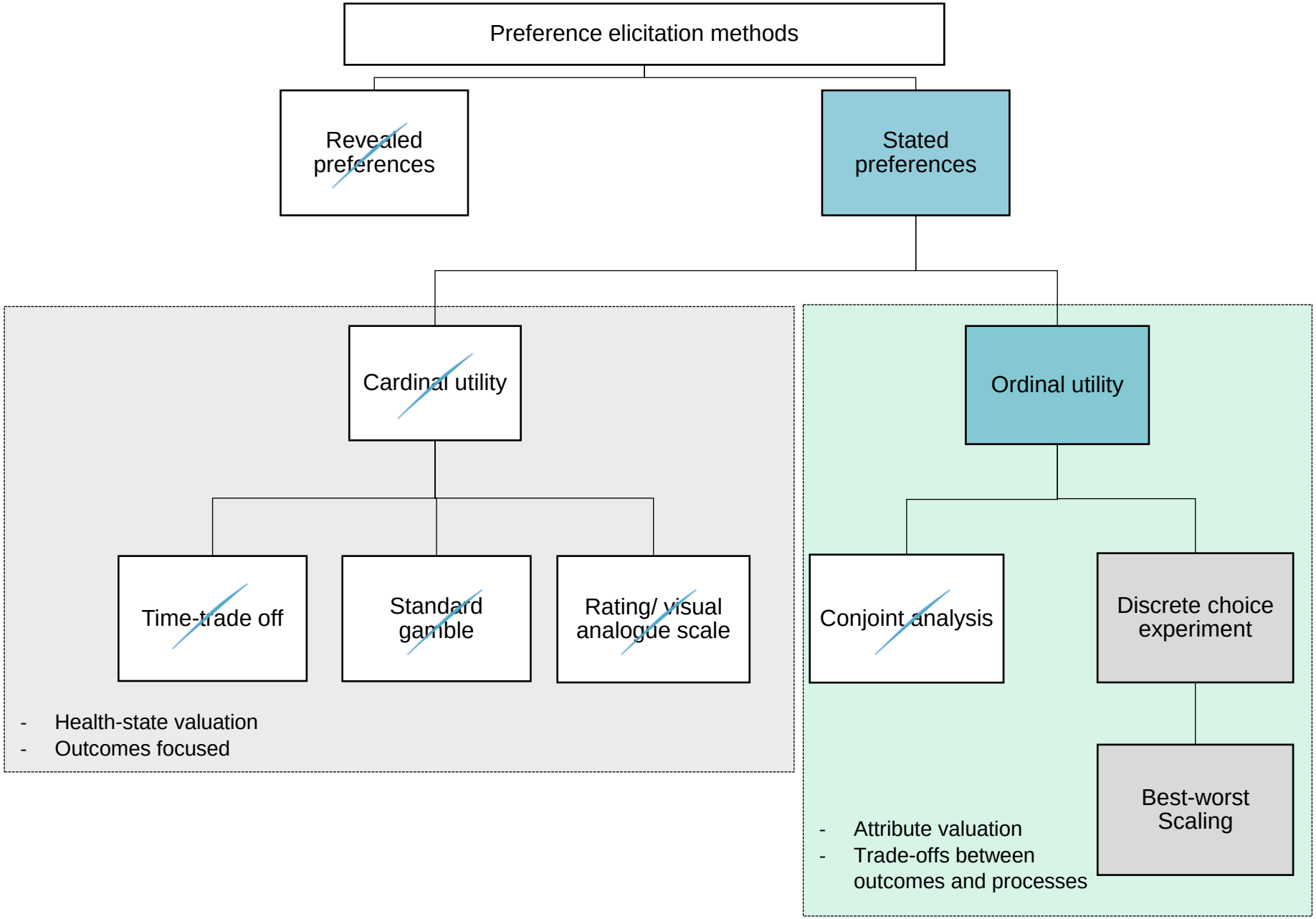
2.5.1 Revealed vs stated preferences

Revealed preference data relates to the observed choices of individuals in real-world situations (Adamowicz *et al.*, 1994). The big advantage of revealed preference methods is that data reflects actual choices, avoiding well-documented problems associated with the use of hypothetical decisions such as strategic responses or failure to fully consider the constraints of a hypothetical decision, which may lead to a gap between what people say they will do and what they actually do (Hensher *et al.*, 2005b). However, reliance on observable data means revealed preference studies are limited to the analysis of situations that currently exist and where there is some degree of observable variation in the attributes of alternatives. In practice, this typically restricts revealed preference data to market goods, limiting potential applications to healthcare problems, particularly in countries with universal, free at the point of use healthcare systems like the UK (Boyle, 2003; Maclennan & Williams, 1980).

Stated preference methods are a response to the limited availability and shortcomings of revealed preference data. These methods collect data relating to preferences using experimental or survey-based methods where respondents state what their choices would be in hypothetical market situations (Hensher *et al.*, 2005b). Stated preference techniques allow for the analysis of preferences for both market and non-market goods where consumption patterns are currently unavailable or unobservable, including the introduction of new or conceptual products. By allowing researchers to vary attributes without restriction, stated preference methods also allow the responsiveness (elasticity) of demand to changes such as price increases or improvements in quality to be considered more thoroughly (Bridges, 2003).

Stated preference methods fall in to two key categories: cardinal utility and ordinal utility measures.

Figure 2.2: Overview of key preference elicitation methods in healthcare



2.5.2 Cardinal utility methods

Cardinal methods assume that preferences can be assigned a value according to a predetermined interval scale (Brazier *et al.*, 2017). Cardinal methods are historically seen as the gold standard in the estimation of HSUVs used to calculate QALYs for use in cost-utility analysis – a cornerstone of economic analysis of healthcare interventions in England and Wales (NICE, 2018).

Three of the most used cardinal preference elicitation methods within healthcare are rating tasks (and visual analogue scale in particular), standard gamble (SG) and time-trade off (TTO) techniques.

2.5.2.1 Rating tasks and visual analogue scales (VAS)

Rating tasks require participants to indicate the value they place on a series of items of interest (e.g. health states, health outcomes) with respect to a clearly defined scale. In healthcare, the visual analogue scale (VAS) is the most commonly used rating method (Drummond *et al.*, 2015). Units on the scale range from 0-100 where zero typically represents death and 100 represents full health. The distance between items on the scale is intended to have interval properties, allowing quantitative estimation of preferences between items (Brazier *et al.*, 2017). Rating tasks, including VAS are based on pragmatism as opposed to being theoretically driven. As a result, methods are subject to widescale criticism. In particular, respondents are not forced to discriminate between items, limiting the information that can be gained from responses (i.e. all items can be assigned the same utility). Furthermore, rating does not consider opportunity costs since respondents are not required to consider any trade-offs between items (Torrance *et al.*, 2001).

2.5.2.2 Time-trade off (TTO)

TTO involves asking a respondent to make trade-offs between the length of life and quality of life by stating the point of indifference between time spent in a diminished health state and a reduced time period spend in full health (Dolan *et al.*, 1996; Torrance, 1976).

2.5.2.3 Standard Gamble (SG)

Standard gamble (SG) replaces time considerations with a risk component by requiring respondents to state the point of indifference between a certain diminished health state and a gamble between an improved health state and immediate death (Drummond *et al.*, 2015).

Despite their dominance, cardinal methods are subject to several shortcomings. For instance, matching methods may be inappropriate for elderly or vulnerable groups due to their complexity and the explicit consideration of death (Coast *et al.*, 2008). Crucially, such methods tend to focus solely on health outcomes (typically health states), neglecting the potential implications of process utility (Ali & Ronaldson, 2012). Cardinal methods are capable of measuring process utility but applications are infrequent and the ability to measure trade-offs between outcomes and processes is limited (Howard *et al.*, 2008).

2.5.3 Ordinal utility methods

Ordinal methods focus on the relative order of preferences for two or more alternatives without directly establishing the degree of preference of one alternative over the other (i.e. the magnitude of preferences is not meaningful) (Board, 2009; Brazier *et al.*, 2017). Ordinal methods allow for the full consequences of health interventions (i.e. health outcomes, non-health outcomes and process attributes) to be considered simultaneously (Ali & Ronaldson, 2012). Ordinal methods are sometimes used to value health states, most notably being used in the estimation of QALY weights for the EQ-5D-5L (Bansback *et al.*, 2012; Devlin *et al.*, 2018). In general, ordinal preference elicitation is more commonly used to evaluate healthcare services, products and policies due to the ability to estimate a wide range of welfare measures including demand, elasticity of demand and the willingness to trade (i.e. marginal rate of substitution) between different aspects of interventions (Brazier *et al.*, 2017).

Following Carson and Louviere (2010), ordinal methods can be separated in to two distinct categories; conjoint analysis and discrete choice experiments.

2.5.3.1 Conjoint analysis

Conjoint analysis requires respondents to rate or rank a series of 'profiles' that describe alternatives (e.g. goods, services, health interventions, health profiles) in terms of multiple attributes. Attributes capture the key features of the alternatives (in the case of healthcare interventions), this may be a combination of outcomes and process characteristics. Economic modelling is used to analyse the ordering of profiles allowing the importance of each attribute to be inferred (Boyle *et al.*, 2001). A further extension of this method is adaptive conjoint analysis, where choice sets are updated and adapted according to previous choices in attempt to improve the efficiency and increase the potential number of attributes within surveys without the associated increase in cognitive and time burden placed on respondents (Green *et al.*, 1991).

Conjoint analysis is commonly used in marketing but is not widely used within healthcare and is subject to criticism. For instance, conjoint analysis requires the simultaneous evaluation of a large number of alternatives (typically between 12-30 profiles), an activity which is not only highly burdensome for respondents but also not reflective of decision-making in the real world (Louviere, 1988). Furthermore, conjoint analysis is fundamentally a ranking or rating exercise and is therefore subject to all the shortcomings associated with these methods (section 1.5.2), meaning responses may not align with the theories of economic demand underpinning the economic models used to explain choices and analyse responses (Louviere *et al.*, 2010).

2.5.3.2 Discrete choice experiments (DCE)

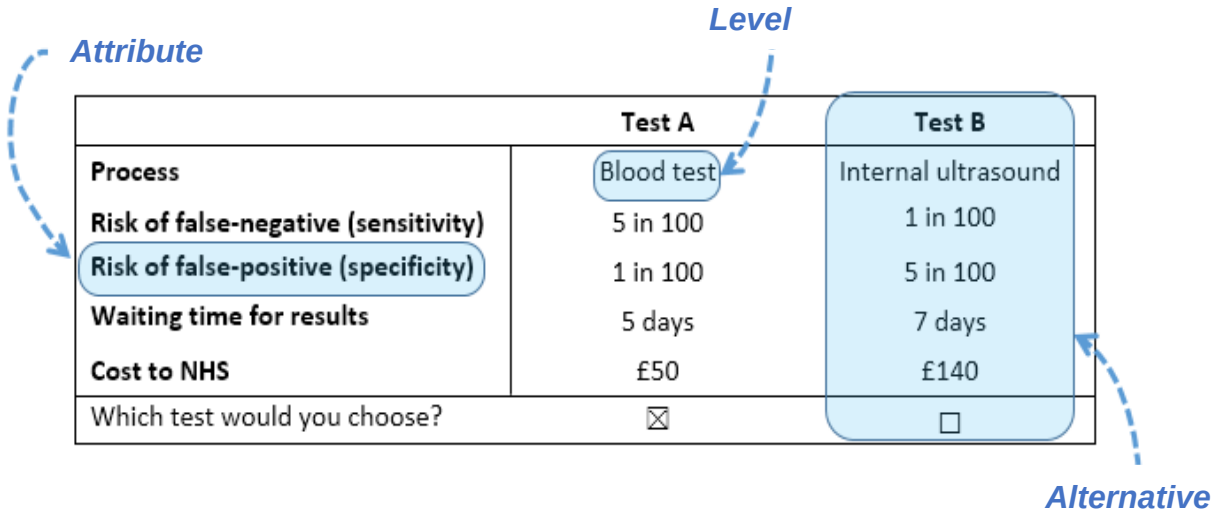
DCEs require respondents to complete a series of choice tasks (Figure 2.3) where they must state their choice between two or more hypothetical alternatives (or profiles) used to describe the competing options of interest (e.g. diagnostic tests, treatments etc.) as specified within the research question (Ryan *et al.*, 2008).

Similar to conjoint analysis, profiles (or alternatives) describe alternative interventions according to key characteristics ("attributes"), each of which may consist of a number of variations ("levels"). By considering all the information and selecting the alternative that yields the highest utility, respondents implicitly make trade-offs between the perceived advantages and disadvantages of each alternative when choosing between

profiles. Through the completion of a series of choice tasks where attribute levels are varied systematically according to an underlying experimental design structure, it becomes possible to measure the influence of each attribute on the choice of the decision maker (Ryan *et al.*, 2001).

In comparison to conjoint analysis techniques, DCEs represent a relatively straightforward task that is more representative of real-world decisions since the comparison of profiles is limited to the identification of the most preferred option amongst a small number (usually limited to two or three) of alternatives (Louviere *et al.*, 2010). The combination of reduced cognitive burden for respondents and strong theoretical foundations (see section 1.6) mean discrete choice experiments are a leading choice-based method within healthcare to date (de Bekker-Grob *et al.*, 2012).

Figure 2.3: Example of a DCE choice task



2.5.3.3 Best-worst scaling

Best-worst scaling (BWS) was introduced by Finn and Louviere (1992) as an extension of, or alternative to, discrete choice experiments. During BWS tasks, participants respond to a series of tasks where they are shown a subset of items and asked to select the “best” or “most important” item and the “worst” or “least important” item. Simultaneously examining items selected as “most” and “least” concern provides

greater information than examining most important items alone, allowing the underlying scale to be inferred (Louviere *et al.*, 2015).

For example, if a choice set contains 4 items; A, B, C and D, and a respondent identifies A as most important; we know that A is preferred to all other items in this set; however, the overall ordering is unknown with six potential relationships (ABCD, ABDC, ACBD, ACDB, ADBC or ADCB). However, if the respondent also indicates that D is least important, then according to the property of transitivity, from a single choice task we can infer the ranking of these for attributes for this individual can only be either ABCD or ACBD.

The completion of several choice tasks with varying subsets of items allows the importance of attributes to be estimated according to an underlying latent scale. A benefit of BWS is a full, proportionally-scaled ranking of a large number of attributes can be achieved with a relatively small amount of choice tasks (Finn & Louviere, 1992).

Summary box 2.1: Rationale for using discrete choice experiments to investigate preferences within this thesis

The aim of this thesis was to understand preferences towards ovarian cancer testing with the fundamental aim of highlighting improvements that may lead to earlier diagnosis. The ability to include both process and outcome attributes was a key motivation for the decision to use DCE methods within this thesis.

The investigation of preferences within this thesis was not limited to outcomes alone. Preferences towards diagnostic processes and experiential factors such as waiting times are also important aspects of healthcare delivery and impact adherence and help-seeking behaviour. As such, the ability to measure trade-offs between process and outcome characteristics was a crucial component of this thesis. The use of DCEs allows for an exploration of the marginal rate of substitution between the favourable and unfavourable aspects of testing whilst avoiding explicit elicitation. This indirect approach utilised within DCE methods means strategic behaviour and associated biases (e.g. protest bids) can be avoided (Hanley *et al.*, 2001).

Finally, repeated choice tasks with varying attributes across a range of levels increases the richness of data and the potential for transferability of findings across healthcare contexts. The focus on attribute-level rather than intervention-level preferences means that whilst choice tasks can be contextualised to the specific issues relating to ovarian cancer testing, findings are more likely to be applicable to other disease areas (Rolfe *et al.*, 2015). For example, providing insights into the trade-off between waiting times and accuracy in cancer testing more generally.

2.6 Theoretical foundations of DCEs

Discrete choice experiments have a strong theoretical foundation based on several normative economic theories that describe how people are assumed to make decisions with Lancaster's demand theory forming the foundations of DCEs..

2.6.1 Lancaster's theory- demand theory

Prior to the publication of Lancaster's seminal 1966 paper: A New Approach to Consumer theory, leading theory on consumer behaviour could be described as a "goods are goods" approach, whereby consumption decisions were exclusively based on preferences for goods themselves rather than the distinguishing, intrinsic properties that define every good/service (Debreu, 1959). Despite the prevalence of this theory, it failed to explain many aspects of observable consumer behaviour such as variations in taste across individuals.

In response to the shortcomings of this goods-focused theory, Lancaster (1966) proposed an alternative theory of demand which now forms the theoretical basis for any attribute-based valuation method. The theory consisted of three main assumptions:

1. Goods, *per se*, do not provide a consumer with utility, instead goods possess characteristics, and it is these characteristics that derive utility
2. Goods will generally possess more than one characteristic and characteristics may be shared by more than one good.
3. Goods consumed in combination may possess different characteristics compared to when they are consumed separately

Although this theory may seem quite straightforward, it accounted for many economic phenomena that were previously observable but unexplainable under the 'goods are goods' approach. For example, assumption one explains the willingness to pay a premium for luxury brands. Assumption two provides an explanation for why some goods are considered close substitutes (e.g. paracetamol and ibuprofen to relieve mild to moderate pain) whereas assumption three explains how some goods are

considered complementary (e.g. Antihistamine tablets and Sodium Cromoglicate eye drops to relieve hay fever symptoms), demonstrating increased utility when consumed together.

When considering preference estimation, Lancaster's theory allows for several important inferences. Firstly, the values placed on each characteristic can be summed to estimate the value of a good or service as a whole (Ryan, 2004), meaning preferences can be examined at both an aggregate 'good' level and characteristic level. This approach allows us to predict how preferences will change when we modify the options or baskets presented to consumers by studying how these vary according to the change in the characteristics that make them up. By relying on a study of the characteristics rather than the goods or service involved, we can predict how consumer behaviour is expected to change when new goods are introduced into the marketplace by considering the preferences for the underlying characteristics that make up the new item.

2.6.2 Random Utility Theory (RUT)

Random utility theory (RUT) introduced by Thurstone (1927) and developed by McFadden (1974; 1986) forms the theoretical basis of most DCEs (and BWS studies) conducted in healthcare to date. Random utility theory still assumes that consumers act rationally according to the axioms of choice, so does not attempt to explain irrational behaviour but rather the unavailability of underlying information required by researchers.

RUT assumes that individuals derive a certain level of utility from consumng any good and when presented with competing options individuals make selections according to utility maximisation subject to any constraints (e.g. budgetary, time). In reality, utility is a latent concept that cannot be fully observed. Instead, utility consists of two components; a deterministic element that can be observed and measured through the choices of individuals and a random, unobservable and/or unmeasured element.

Therefore, in the context of DCEs, the latent utility, U for individual, n for a given profile or scenario, i can be estimated by the following equation:

$$U_{ni} = V_{ni} + \varepsilon_{ni}$$

(eq. 1)

V, represents the deterministic component of utility constructed of all the attributes and levels included within the DCE (X_i) combined with any explicitly measured covariates (i.e. personal characteristics) (Z_i). ε, represents the random component of utility consisting of all other factors that influence the preference of individual, n toward scenario, i. This random component, ε, may consist of observable factors that are not captured by the DCE (i.e. unobserved personal characteristics) and/or fundamentally unobservable/inconceivable factors relating to variations in utility/preferences.

The deterministic element of utility is typically modelled according to equation 2:

$$V_{ni} = \beta X_{ni} + \gamma Z_n$$

(eq.2)

where β is a vector that represents the weight of each attribute on the overall utility of alternative i and γ is a vector of respondent characteristics (i.e. covariates). β-coefficients are also known as part-worth utilities and their size indicates importance of a particular attribute on an individual choice and the direction indicates whether this effect is positive or negative. Part-worth utility estimates can be used to calculate the relative importance of attributes and the willingness to trade between attributes based on the marginal rates of substitution. Additionally, part-worth utilities can be summed to calculate the total utility of competing alternatives (e.g. different diagnostic tests) or predict uptake rates.

Since the actual distribution of this random component is unknown, a probabilistic function is used to estimate choices within a DCE. The probability of selecting profile, i from a set of alternatives, j can be expressed as:

$$P_i = P(U_{ni} > U_{nj})$$

$$= P(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj})$$

$$= P(V_{ni} - V_{nj} > \varepsilon_{ni} - \varepsilon_{nj}) \; \forall \; j \neq i$$

(eq. 3)

The probability of choosing profile i from alternatives is directly observed during the DCE. The equation demonstrates that a high probability of choosing alternative i, implies a high probability of the deterministic utility of i (allowing for random errors), is greater than the deterministic utility of alternatives, n. RUT assumes a joint distribution for ε_i and estimation models are derived by assuming a distribution for this random component of utility.

2.6.3 Analysis of choice – modelling approaches

Multinomial logit model (MNL)

The multinomial logit model (MNL) (or Conditional Logit model) is the most commonly used discrete choice model (Soekhai *et al.*, 2019) and is used as a starting point for analysis throughout the empirical chapters of this thesis.

MNL assumes errors follow an independent and identical (i.i.d) type 1 Gumbel distribution meaning the error term has a mean of zero and there is no correlation in the error term across alternatives or across choices. This means the probability of individual, n, choosing alternative, i can be represented by the following equation:

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^{J} e^{V_{nj}}}$$

(eq.4)

Replacing V in eq. 4 using eq. 2 gives:

$$P_{ni} = \frac{e^{(\beta X_{ni} + \gamma Z_n)}}{\sum_{j=1}^{J} e^{(\beta X_{nj} + \gamma Z_n)}}$$

(eq.5)

Mixed logit (ML)

The mixed logit model (also known as random parameters logit) is used to accommodate and evaluate the extent of preference heterogeneity across choice data. The ML introduces a greater level of flexibility by allowing some or all model parameters (which are fixed within MNL models) to become random across respondents, based on a specified distribution (Train, 2009).

In practice, this means the utility function (eq. 1) relating to the utility that decision maker, n obtains from alternative, i becomes:

$$U_{ni} = \beta_n X_{ni} + \varepsilon_{ni}$$

(eq.6)

This function is almost identical to the utility function underpinning the MNL model except β-coefficients are now indexed by n (i.e. coefficients become individual-specific, reflecting personal tastes and preferences).

Utility is now dependent on an individual-specific set of coefficients. If β_n was observable for each individual, then the choice probability becomes a standard MNL model (i.e. the probability is *conditional* on β_n):

$$L_{ni}(\beta_n) = \frac{e^{(\beta_n X_{ni})}}{\sum_{j=1}^{J} e^{(\beta_n X_{nj})}}$$

(eq. 7)

However, in practice β_n is unobservable. Instead, β_n is modelled as a random variable which differs between respondents according to a density $f(\beta)$, which is a function of parameters θ. This means an unconditional choice probability must be estimated. This is achieved by integrating the conditional choice probability, $L_{ni}(\beta_n)$ over all possible variables of β_n:

$$P_{ni}(\beta_n) = \int \frac{e^{(\beta_n X_{ni})}}{\sum_{j=1}^{J} e^{(\beta_n X_{nj})}} f(\beta) d(\beta)$$

(eq. 8)

Put simply, the mixed logit probability is achieved by taking an average of all possible conditional choice probabilities (for all possible values of β_n) weighting by the likelihood of observing a particular β_n in the population according to the density function. In practice, rather than estimating fixed model coefficients, the research specifies a distribution for the coefficients and model estimates relate to the parameters for this distribution (Hensher *et al.*, 2005c; Train, 2009). The requirement to make distributional assumptions about the distribution of heterogeneity is the primary limitation of the ML model. However, Greene and Hensher (2003) argue this limitation is somewhat offset by the flexibility of the model and the ability to specificity unobserved heterogeneity at the individual level. A secondary limitation is the computational power required to estimate ML models, particularly where interactions are included to assess the influence socio-demographic characteristics on preferences.

Latent class logit

The final model used throughout this thesis is a latent class model (LCM). Similar to the mixed logit model, LCM models are used to explore preference heterogeneity within models. However, unlike mixed logit which estimates the random parameters following a continuous joint distribution, the LCM assumes preference heterogeneity can be captured within a discrete number of preference classes. Each with a separate utility function meaning unobserved heterogeneity is captured by these preference classes. Respondents are probabilistically assigned to classes based on the underlying latent structure of preferences (Train, 2009).

The LCM model simultaneously estimates attribute parameters using a MNL model for each class and predicts the probability of an individual belonging to a specific preference class (Shen, 2009).

The probability that individual, n chooses alternative i, given the individual is a member of preference class, c is given by equation 9:

$$P_{ni|c} = \frac{e^{(\beta_c X_{ni})}}{\sum_{j=1}^{J} e^{(\beta_c X_{nj})}} \quad c=1,\ldots C$$

(eq. 9)

Class membership is also modelled in MNL form:

$$H_{nc} = \frac{e^{(\gamma_c Z_n)}}{\sum_{c=1}^{C} e^{(\gamma_c Z_n)}} \; c=1,\ldots C, \; \gamma_{C=0}$$

(eq. 10)

Where H_{nc} denotes the probability that individual, n belongs to class c. Where γ_c is the parameter vector for Z_n, a set of observable characteristics included in the model as explanatory variables determining class membership (i.e. sociodemographic differences between preference classes). The parameter vector for the final preference class C is normalised to zero to allow model identification to be achieved.

The semi-parametric approach of the LCM overcomes from of restrictive distribution assumptions required in ML model (Greene & Hensher, 2003) . Instead, the primary obstacle when using LCM is the specification of the number of classes. The number of classes included in the LCM must be imposed by the researcher. The appropriate number of classes is typically determined by estimating multiple models with different numbers of classes and examining several post-estimation criteria used to assess model fit. Most frequently log-likelihood (LL) and Bayesian information Criteron (BIC), although Akaike Information Criterion (AIC) and Consistent Akaike Information Criterion (CAIC) are also commonly considered (Shen, 2009). Each criteria assigns different penalties based on model characteristics based on the number of parameters to be estimated and sample size. In all cases, models with lower criteria values indicate a better fit. However, a key caveat is class sizes may need to be restricted in order to aid interpretability as increasing numbers of classes may become indistinguishable differences in preferences, and significance of parameters may be impacted if sample size is not sufficient to accommodate the segmentation.

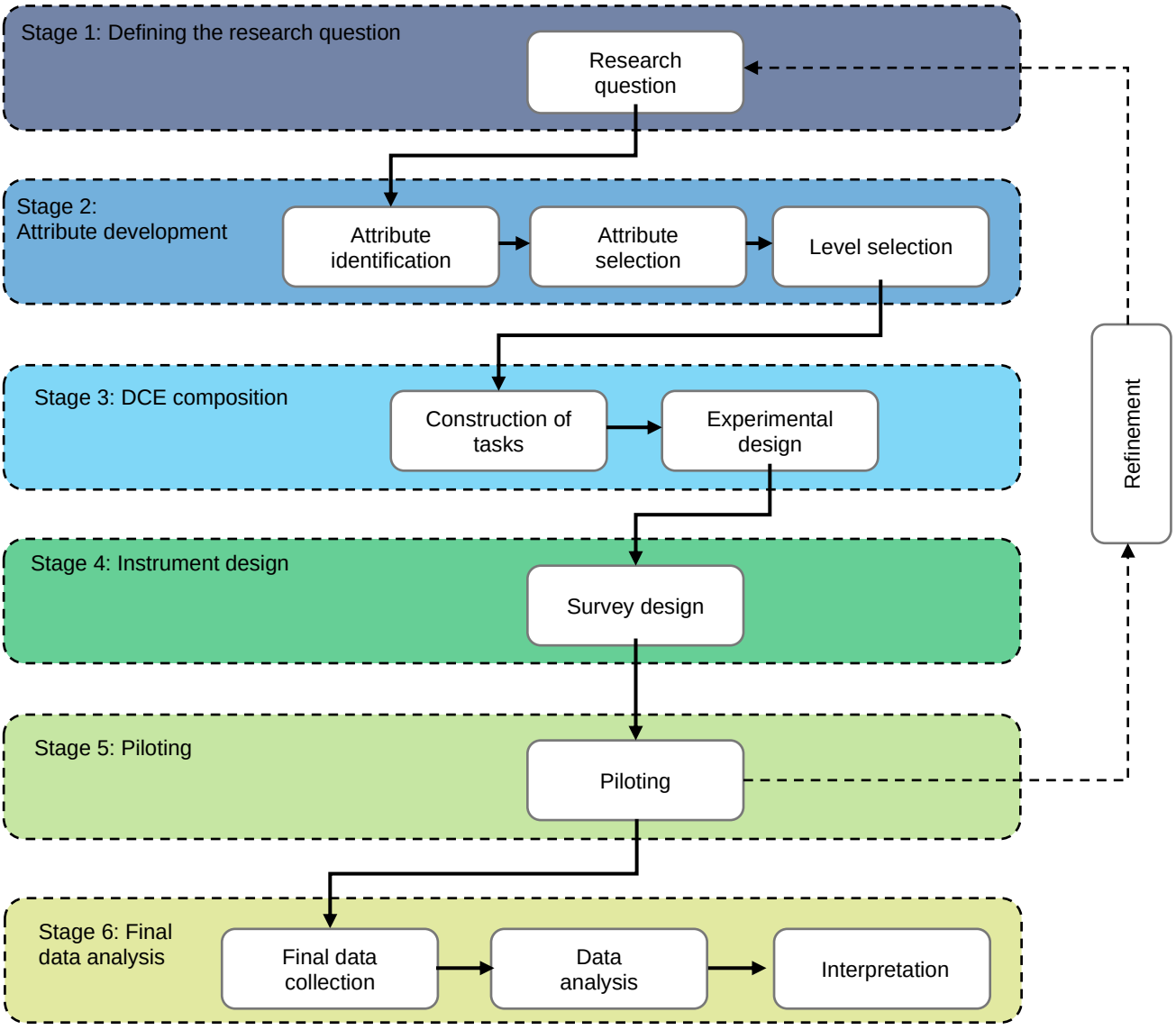
Overall, both ML and LCM provide useful and acceptable approaches to accommodating preference heterogeneity (Greene & Hensher, 2003). Comparisons of the two approaches are inconclusive, although a limited number of studies suggest LCM may marginally outperform ML (Shen, 2009).

When exploring heterogeneity in this thesis, ML is used to initially understand the scope of heterogeneity within samples. LCM is used as follow up model to understand sociodemographic associations with preference heterogeneity.

2.7 Key steps of DCEs

DCE methods have been outlined in detail within several methodological guidelines (Bridges *et al.*, 2011; Coast *et al.*, 2012; Hauber *et al.*, 2016; Johnson *et al.*, 2013; Lancsar & Louviere, 2008). The key steps involved in undertaking a DCE are summarised in this section (Figure 2.7). Specific details relating to each step are discussed in greater detail in proceeding chapters.

Figure 2.4: Key steps required to design and conduct a discrete choice experiment



A well-conceptualised research question with a clearly defined purpose and perspective are essential for meaningfully designed and implemented DCEs (Bridges *et al.*, 2011). The format of choice tasks should also be determined early on, including decisions relating to the number of alternatives per choice tasks, whether alternatives are labelled (e.g. 'blood test', 'CT scan', 'ultrasound') or generic (e.g. 'Test A', 'Test B') and whether respondents are forced to choose between alternatives or are provided a 'opt-out' option (Lancsar & Louviere, 2008). To maximise external validity, choice formats should simulate the actual choice if faced in real life as closely as possible and based on the outcomes of interest. For example, where estimating uptake, a labelled alternatives and an opt-out option may be most appropriate for accurately modelling participation. Alternatively, if measuring the marginal rate of substitution between attributes is the primary object then an unlabelled, forced choice format may be more beneficial.

The choice of attributes and associated levels within DCEs ultimately determines the legitimacy of results implying attribute selection requires a rigorous approach (Bridges *et al.*, 2011). The importance of any given attribute is calculated relative to the other included attributes within the study, therefore if central characteristics of an intervention are excluded or misunderstood by respondents the results are likely to be biased and less valid in the context of the research question. As a minimum, the selection of attributes should be based on a review of existing literature or policy questions with guidance typically advocating the use of qualitative methods such as interviews/focus groups with industry experts, patients and a sample of the target audience as several key points throughout the process to ensure the comprehensiveness and completeness of final attributes (Coast *et al.*, 2012). The total number of attributes is context-specific but is typically limited to between 4-9 (Soekhai *et al.*, 2019).

Similar to attributes, levels should be evidence-based and may be defined according to specific, currently available options or more generically to allow for the incorporation of future options within the existing framework. Levels may be defined categorically (e.g. blood test, ultrasound), continuously (e.g. number of days, costs) or binarily (e.g. yes/no) (Lancsar & Louviere, 2008). Levels should be plausible (i.e. the total range should only cover what is meaningfully required) whilst also remaining distinguishable

and interpretable to respondents (i.e. limited number of levels not too close together) (Bridges *et al.*, 2011).

Choice sets are typically limited to between 8-32 questions per respondent and are chosen according to an experimental design which aims to maximise the preference information that can be gained from respondent choices. For the majority of DCEs including all possible combinations of attributes and levels ('full factorial design') is not achievable (Johnson *et al.*, 2013). Instead, a fractional factorial design using a sample of choice sets is generated using statistical software such as Ngene or SAS.

Survey instrument design is also an important step in the development process. When designing choice tasks, attribute wording, ordering and framing are important consideration and may influence responses (Howard & Salkeld, 2009; Kjær *et al.*, 2006). Additionally, it is important to consider the sampling strategy and mode of administration which may vary from self-completed or interviewer-administered tasks either electronically or pen-and-paper. Each method having benefits and limitations in terms of costs and number and quality and rate of response. Well-designed training materials, practice questions and contextual information are also important for maximising the quality of responses. Finally, researchers usually wish to include additional background questions to provide context and aid the analysis and interpretation of results (e.g. sociodemographic information, attitudes and beliefs).

Following the formative stages of DCE design, a pilot study should be conducted (Lancsar & Louviere, 2008). Piloting should mirror the intended final data collection process and aim to test task complexity, understanding of attributes and levels and response rates. Findings from piloting inform sample size calculations and may highlight flaws in the existing DCE design which can be amended before proceeding to final data collection (Hensher *et al.*, 2005b).

Data are analysed by applying econometric modelling techniques. Model selection should be informed by economic and behavioural theory and statistical considerations relating to model fit (Bridges *et al.*, 2011). Conditional logit is the default model for many studies, before fitting alternative models based on different assumptions about the distributions and properties of the random error component of utility and underlying

relationship structure of included attributes (Hauber *et al.*, 2016). To aid interpretation, model estimates are then used to calculate additional welfare outcomes such as uptake, elasticity of demand and marginal rates of substitution, commonly in the form of willingness to pay (WTP) estimates.

2.8 Chapter summary

This chapter introduced the economic theory underpinning preference elicitation and described leading preference elicitation methods. Preferences are the underlying determinant of the of utility gained from consuming a good or service. Rational consumers make choices with the objective of maximising their utility. This means preferences can be inferred by examining the choices of consumers.

Revealed preference methods are the optimal method for estimating preferences, however, measurement is reliant on availability of observable data which is often not possible in healthcare settings due to the non-market nature of medical goods. Stated preference methods provide an alternative approach to preference elicitation and involve observing choices in hypothetical market scenarios.

The discrete choice experiment was identified as the most appropriate stated preference method for this thesis due to the ability to measure trade-offs between outcomes and process attributes. This feature is less common in other leading methods within health economics (e.g. TTO and SG) which typically focus on outcomes only.

The final sections of the chapter introduced the theoretical foundations and key stages of DCEs. These concepts are expanded during later chapters.

3 Ovarian cancer: current challenges and the role of preferences

3.1 Introduction

This chapter provides an overview of ovarian cancer including key statistics (relating to incidence, diagnosis, and prognosis), diagnostic guidance in primary care and current challenges and barriers to earlier diagnosis - which can lead to improved outcomes. The chapter presents examples of how women's preferences may be useful in addressing some of the challenges to earlier diagnosis, summarising existing evidence about preferences towards ovarian cancer and highlighting key evidence gaps.

3.2 Ovarian cancer statistics

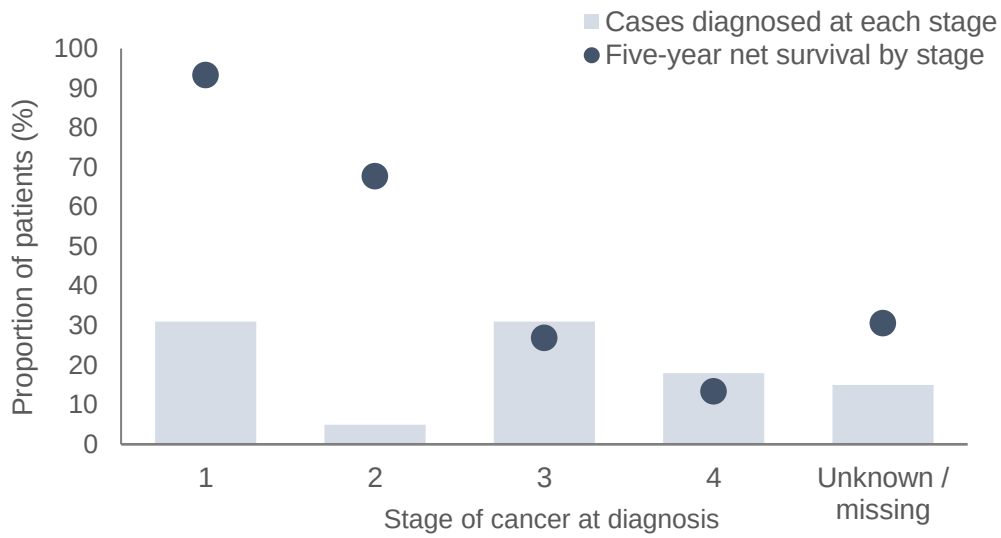
3.2.1 Ovarian cancer incidence

Ovarian cancer is the 6th most common cancer in females in the UK, with approximately 7,500 new cases diagnosed annually, accounting for 4% of all new cancer cases in females (Cancer Research UK, 2019). Ovarian cancer incidence is strongly correlated to age, with incidence rates increasing steeply from the age of 40 and peaking in women aged 75-79 (72 cases per 100,000 compared to overall incidence rate of 22.7 cases per 100,000).

3.2.2 Stage of diagnosis and prognosis

At present, over 50% of ovarian cancer cases with a known stage are diagnosed at a late stage (III or IV) (Figure 3.1) (Cancer Research UK, 2019). Average survival for ovarian cancer is low: on average - 46% of women in England and Wales will survive five or more years. Stage of diagnosis has huge implications on survival. Five-year survival for women diagnosed with stage IV cancer is just 13%, as opposed to 93% for women diagnosed at stage I (ONS, 2019b). Alongside prognosis, stage at diagnosis also has cost implications, with one study estimating average treatment costs for stage IV are three times higher than costs of treating stage I ovarian cancer (Incisive Health, 2014).

Figure 3.1: Proportion of ovarian cancer cases diagnosed at each stage and five-year survival statistics based on stage at diagnosis



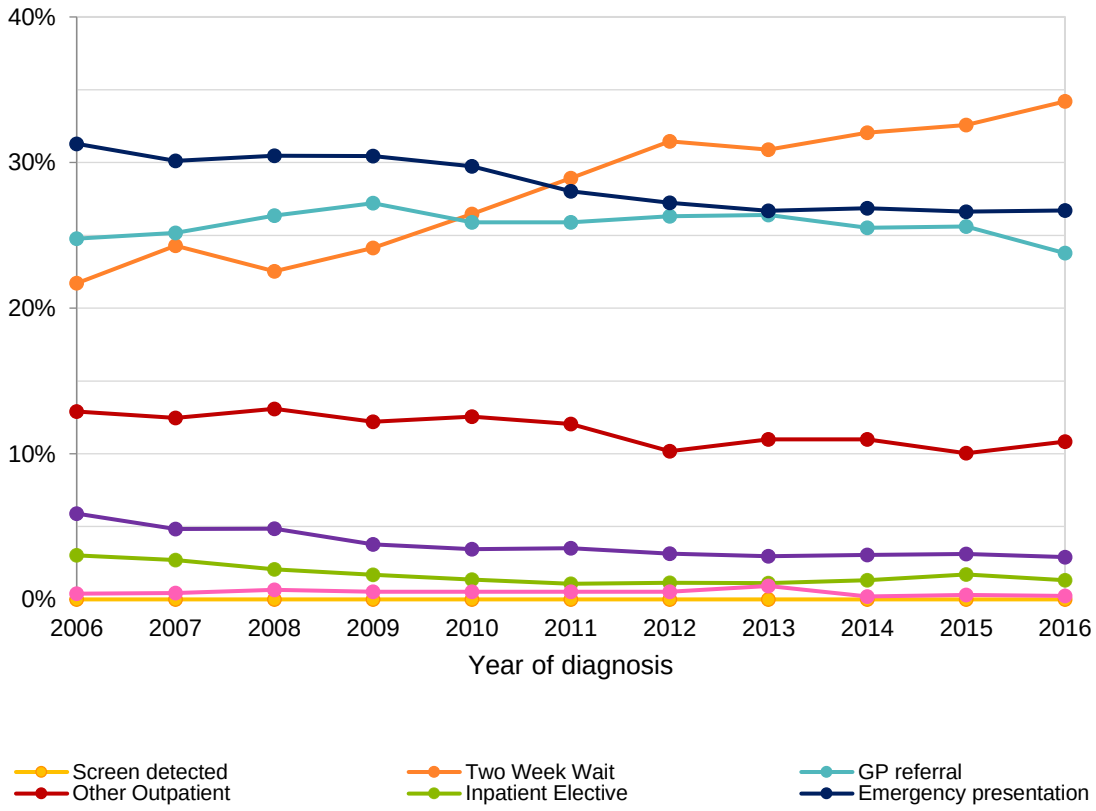
Source: Cancer Research UK, 2019

3.2.3 Routes to diagnosis

The route to diagnosis has important implications on stage at diagnosis and ultimately survival. Figure 3.2 demonstrates trends in routes to diagnosis between 2006-2016, which is the most recently available public data (NCRAS, 2019a). Notably, reduced access to primary care during the Covid-19 pandemic is likely to have had an impact on routes to diagnosis. In England, a large proportion of cases are diagnosed through referral from primary care, either following the two-week wait pathway (~35%) or routine referral (~25%). Despite a trend towards increasing diagnosis via referral in primary care over time, around a quarter of cases are diagnosed following emergency presentation (for example, via Accident and Emergency admission). Eighty percent of cases diagnosed following emergency presentation are diagnosed at a late stage, whereas 6 out of 10 cases identified via GP referral are diagnosed at stage I or II[4] (NCRAS, 2019b). Improved prognosis based on route to diagnosis highlights the important role of early identification of ovarian cancers in primary care.

[4] Calculations are based on cases where the stage of diagnosis was known

Figure 3.2: Percentage of ovarian cancer diagnosis by presentation route over time

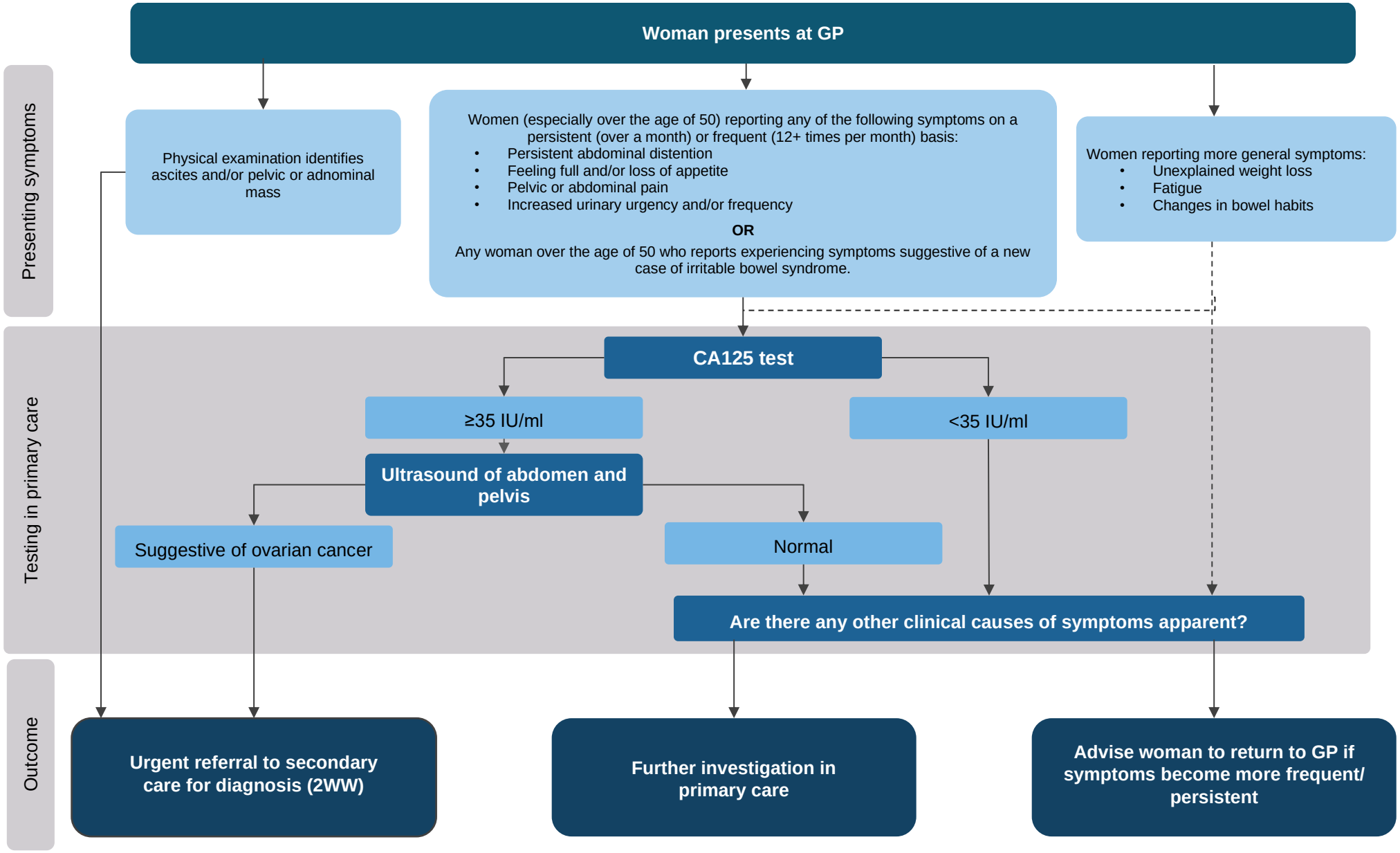


Source: National Cancer Registration and Analysis Service, 2019

3.3 Investigation of ovarian cancer in primary care

In England and Wales, recommendations on how to recognise and manage patients with suspected ovarian cancer in primary care are outlined by NICE Suspected cancer: recognition and referral (NG12) guidance (NICE, 2015), summarised in Figure 3.3. Initial recommendations for testing are based on presenting symptoms. Urgent referral under the two-week wait pathway is advocated for any woman if physical examination identifies ascites (a build-up of fluid in the abdomen) and/or a pelvic or abdominal mass. Women, particularly those over 50, who report experiencing symptoms commonly associated with ovarian cancer (e.g. bloating, pelvic pain, new onset IBS symptoms) on a frequent or persistent basis should be investigated in primary care. For women with more generic symptoms such as unexplained weight loss or fatigue, ovarian cancer may also be considered as a possible cause.

Figure 3.3: NICE Guidance on investigation for suspected ovarian cancer in primary care (NICE, 2011)



In England, testing in primary care is performed on a sequential basis; a serum CA125 (cancer antigen 125) blood test should be requested in the first instance, followed by an ultrasound of the abdomen and pelvis only if results are abnormal. If both tests are suggestive of ovarian cancer, then patients should be urgently referred to secondary care using the two-week wait route. If one or both tests appear normal then alternative diagnoses should be considered before advising women to self-monitor and return to their GP if symptoms persist.

3.3.1 CA125 blood test

The CA125 test is a tumour marker blood test that measures the level of CA125 protein in the blood (Doubeni *et al.*, 2016). Increased levels of CA125 protein (above 35 units per millilitre of blood) may be indicative of epithelial ovarian cancer[5] (NICE, 2011). Benefits of using CA125 blood tests to rule out cancer and triage patients for further investigative testing are that it is cheap and can be performed easily in primary care with no additional training requirements.

3.3.2 Transvaginal ultrasound

Following NICE guidance, ultrasound investigation of suspected ovarian cancer is generally only accessible following an abnormal CA125 result (NICE, 2011). Scans are managed by primary care clinicians but are usually performed in a secondary care setting according to variations in regional access to ultrasound services. Transvaginal ultrasound (TVUS) –internal imaging of the ovaries and fallopian tubes via the vagina, is the most common imaging method used to investigate ovarian cancer but may also be accompanied by an externally performed abdominal ultrasound (Doubeni *et al.*, 2016). Ultrasound testing is more costly than CA125 blood tests and may also be uncomfortable or embarrassing for some women.

3.4 Current challenges/ barriers to earlier diagnosis in primary care

[5] Epithelial ovarian cancers are cancers where the primary tumour forms on the tissue lining the ovary and account for 85-90% of all ovarian cancers. Non-epithelial ovarian are rare (~10% of cases), particularly in women over 40. Additional tumour marker tests targeting non-epithelial ovarian cancers exist but are typically offered in a secondary care setting (NICE, 2011)

This section summarises some of the challenges relating to current diagnostic processes within primary care. Addressing some of these obstacles may help to improve outcomes by improving current rates of early diagnosis and reducing the frequency of diagnoses via emergency presentation.

3.4.1 Demand side- help-seeking by people with symptoms

Symptoms of ovarian cancer are non-specific and indicative of many other conditions and public awareness of symptoms remains low. Pathfinder, a periodically conducted research project by Target Ovarian Cancer (an ovarian cancer charity) which provides a detailed picture of the experiences of people living and working with ovarian cancer in the UK, found less than 30% of the public were able to name any one of the four primary symptoms of ovarian cancer unprompted (i.e. persistent bloating, pelvic/abdominal pain, loss of appetite, increase urinary urgency/frequency) (Target Ovarian Cancer, 2016).

An international survey of 1531 women with ovarian cancer found nine in 10 patients experienced multiple symptoms prior to diagnosis regardless of stage at diagnosis (World Ovarian Cancer Coalition, 2018), however, the time between symptom onset and consulting a health professional varied greatly, with one quarter of people waiting three or more months. The Pathfinder study had similar results, with just 36% of women visiting their GP within a month of symptoms starting (Target Ovarian Cancer, 2016).

A further challenge to help-seeking behaviour is misconceptions around cervical cancer screening leading to false reassurance, with a 2019 YouGov survey finding around 20% of women believed cervical screening is also able to detect ovarian cancer (Target Ovarian Cancer, 2019).

In recent years, there have been substantial efforts to increase public awareness of the symptoms of ovarian cancer and encourage earlier consultation once symptoms arise through public health campaigns (Target Ovarian Cancer, 2022a). Encouragingly, evidence suggests such campaigns are successful in increasing public knowledge of the disease (Target Ovarian Cancer, 2016). However, maintaining and

increasing public awareness remains an ongoing challenge for improving early diagnosis of ovarian cancer.

3.4.2 Supply side- timeliness of testing

Despite direct accessibility to CA125 blood tests and non-urgent ultrasound, delays on the supply side within primary care are common and seemingly exacerbated by gaps in awareness of key symptoms, particularly in the early stages of the disease. Periodic surveys of GPs have demonstrated an improvement in ovarian cancer symptom awareness over time, however, as of 2016 44% of surveyed GPs still believed symptoms only presented in later stages of the disease and 77% agreed with the statement ovarian cancer is a “silent killer” (Target Ovarian Cancer, 2016).

Misdiagnosis or initial investigation for alternative symptoms may be a further source of provider-driven delays. For example, forty-six percent of women from the Pathfinder survey (Target Ovarian Cancer, 2016) were initially referred for testing for conditions other than ovarian cancer. Of these women, 21% over the age of 50 years were told they may have irritable bowel syndrome (IBS), despite NICE guidance clearly stating a possible diagnosis of ovarian cancer should be considered for any woman aged 50 or over presenting with symptoms that suggest a new case of IBS (NICE, 2011) (Figure 3.3).

The Cancer Patient Experience Survey found as of 2019, 51% of women diagnosed with gynaecological cancers visited their GP three or more times before being referred for diagnostic testing. The average across all cancers was 21%, meaning women with ovarian cancer are more likely to face repeat visits to their GP[6].

Lim *et al.* (2016) examined the average diagnostic interval for ovarian cancer based on the primary care records of patients receiving care at 10 UK gynaecological centres. Results suggested that supply-side delays meant the average diagnostic interval (from first presentation to diagnosis) varied between 3-4 months depending on the tumour type and stage of the disease at the time of diagnosis. This finding is

[6] Percentages are calculated based on those who reported visiting their GP at least once about their symptoms prior to diagnosis. Individuals who presented and diagnosed in secondary care exclusively are not included in calculations.

supported by self-reported diagnostic intervals from the 2016 Pathfinder study where 45% of women reported waiting 3 or more months to diagnosis after first visiting the GP, a trend that appeared to be stable across time (Target Ovarian Cancer, 2016).

3.4.3 Limitations of current test modalities

CA125 testing

At the outset of this PhD, high quality evidence on the efficacy of CA125 testing in primary care was non-existent: however, evidence from secondary care demonstrated a number of crucial limitations. Firstly, the test is non-specific; CA125 concentration varies naturally between women and may also be influenced by other conditions such as endometriosis, ovarian cysts, pregnancy, cancers of different origin, liver disease or even menstruation (Moss *et al.*, 2005; Ortiz-Muñoz *et al.*, 2014), such variations mean results need to be interpreted in the context of accompanying symptoms and additional testing prior to referral is recommended (e.g. TVUS). Additionally, 'normal' (<35 IU/ml) results do not necessary ensure the absence of ovarian cancer, with studies demonstrating up to 20% of patients with advanced cancer and 50% of patients with early-stage ovarian cancer will not have elevated levels of CA125 (Dochez *et al.*, 2019). During the course of this PhD, additional evidence exploring the efficacy of CA125 in primary care was published, with results having further implications for the diagnostic pathway (Funston *et al.*, 2020a) (discussed further in section 3.4.4).

Transvaginal ultrasound (TVUS)

There are substantial uncertainties about the efficacy of transvaginal ultrasound used for primary care investigations of ovarian cancer symptoms. Findings from a limited number of studies are mixed. Doubeni *et al.* (2016), indicate TVUS provides good sensitivity, particularly for early stage and non-epithelial ovarian cancers. Alternatively, other studies suggest whilst TVUS is generally effective at detecting pelvic masses, there is limitation in sensitivity to detect small, early lesions (Gilbert *et al.*, 2012; National Academies of Sciences & Medicine, 2016). Test specificity (i.e. the false positive rate) also appears to be an issue, with the majority of ovarian masses

identified using TVUS being benign upon further investigation (National Academies of Sciences & Medicine, 2016). However, existing studies relate to the use of TVUS as a screening modality or investigations in a secondary care setting. Issues of test performance are heightened due to a current lack of guidance on what TVUS abnormalities should prompt referred (NICE, 2011).

Beyond test performance, the use of TVUS is subject to several further challenges that may lead to missed or delayed appointments and/or reduced satisfaction. For instance, many patients may anticipate the test to be uncomfortable or embarrassing (Gentry-Maharaj *et al.*, 2013). Furthermore, in comparison to the CA125 blood test that can be performed by a nurse or phlebotomist in a local GP surgery, TVUS is more expensive for the NHS, has longer waiting times and often requires travel to a specialist unit.

3.4.4 Test strategy/ diagnostic pathway debates

In 2015, NICE updated referral guidelines for all cancers to recommend urgent referral under the two-week wait pathway for patients with symptoms that suggest a 3% or higher chance of cancer (NICE, 2015). Ovarian cancer diagnostic guidance has not been updated to reflect this change (NICE, 2011). Emerging evidence suggests that due to current inefficiencies and uncertainties in the existing diagnostic pathway, patients who exceed the threshold for urgent referral (>3% chance of cancer) based on symptoms or triage testing using CA125 may not qualify for urgent referral or urgent referral may be delayed. Additionally, alternative testing procedures are an area of ongoing research increasing the need for updated ovarian cancer guidelines in the near future.

3.4.4.1 CA125 abnormal cut-off level

Since 2011, NICE guidelines have used a cut-off of ≥35 U/ml to classify “abnormal” or “raised” CA125 levels which qualify for further investigative testing in primary care prior to referral to secondary care (NICE, 2011). However, studies in both symptomatic and asymptomatic populations have demonstrated that a more individualised approach to measuring and monitoring CA125 levels for suspected ovarian cancer could increase

the diagnostic performance of the test compared to a single threshold applied to all women (Cramer *et al.*, 2011; Funston *et al.*, 2020a).

Funston *et al.* (2020a) investigated the diagnostic performance of CA125 testing in primary care at different cut-off points using routinely collected data. Results were also stratified according to patient age. Findings demonstrated the positive predictive value (PPV: the probability that patients with a positive result truly have the disease) at or above the current cut-off point was 10.1% (95% CI: 9.1—11.2%) overall and varied based on patient age from 3.4% (95% CI: 2.5—4.4%) for women under 50 years to 15.5% (95% CI: 13.6—16.8%) for women 50 years or older. A higher proportion of patients under 50 years old are diagnosed with borderline malignancies and tumours which are less likely to elevate serum CA125 levels (e.g. mucinous epithelial and nonepithelial tumours) which may explain the poorer test performance in this group.

Evidence on the performance of CA125 testing has further implications for triaging in primary care. Firstly, the use of CA125 testing as a triage test in younger patients results in reduced sensitivity (i.e. increased false negatives) likely increasing diagnostic delays. As a result, alternative or complimentary testing in younger symptomatic patients may increase sensitivity. Additionally, in all populations, abnormal CA125 levels at the current cut-off of ≥35 U/ml exceed the current NICE threshold for urgent referral (>3% risk of cancer). Despite this, guidelines still require a follow up ultrasound prior to referral regardless of CA125 level, regardless of how much higher than the cut-off results are.

3.4.4.2 Sequential versus simultaneous testing

Given the uncertainty and limitations in performance of both CA125 and TVUS, clinical guidelines for many countries advocate the use of both tests in combination (Funston *et al.*, 2019). However, differences in the timing and sequencing of tests varies and remains an area of debate, with many countries- perhaps most notably, Scotland - performing both tests coincidingly, with referrals made on the basis of abnormal results for either or both tests. Alternatively, guidelines from England, Wales, Ireland and British Columbia, Canada, recommend sequential testing, whereby, in the absence of key symptoms (e.g. abnormal mass and/or ascites) an abnormal result in the first-line

test is necessary to trigger further testing in primary care and both results must be abnormal before being referred for specialist care (Funston *et al.*, 2019).

Sequential testing improves specificity and manages costs, both of which are important especially within publicly funded healthcare systems with resource limitations. On the other hand, simultaneous testing or referral based on an abnormal result in a single test may increase the sensitivity of primary care testing. To date, the appropriate balance between benefits and harms from a patients' perspective remains unclear. A further consideration is whether women are willing to undergo invasive testing as an initial investigation given how non-specific ovarian cancer symptoms are and the low prevalence of disease. However, the avoidance of false negative results appears crucial to increasing early diagnosis, with one study finding the diagnostic interval is doubled for individuals who receive an initial false negative result (Funston *et al.*, 2021).

3.4.4.3 Role of risk prediction tools

High levels of late-stage diagnosis mean research into alternative or complimentary diagnostic approaches is an ongoing area of development. One promising area of research is the use of supplementary risk prediction tools during consultations that can be combined with existing testing methods to improve predictive performance (Funston *et al.*, 2020b; Hippisley-Cox & Coupland, 2012; Nash & Menon, 2020; Sundar *et al.*, 2016; Westwood *et al.*, 2018).

Risk prediction tools aim to avoid missed cases of ovarian cancer but also triage patients and identify high-risk patients for urgent referral leading to earlier diagnosis. However, risk prediction tools vary in sophistication, from symptom checklists to ensure ovarian cancer symptoms are explicitly solicited during consultations to tools using complex algorithms to generate ovarian cancer risk scores based on symptoms, patient characteristics and behavioural risk factors and test results (Funston *et al.*, 2020b). This means tools vary greatly in their ability to predict ovarian cancer, which may impact the likelihood false positive and false negative outcomes. The acceptability of risk-stratified testing is an emerging area of research. Several recent publications relating to cancer screening suggest risk-based screening is viewed favourably by the

majority of respondents. Acceptability within a diagnostic context is less understood (Ghanouni *et al.*, 2020a; Mbuya Bienge *et al.*, 2021; Woof *et al.*, 2020).

3.5 Preferences towards diagnostic testing for ovarian cancer

3.5.1 Role of preferences

Section 3.4 highlighted several shortcomings relating to the investigation of ovarian cancer in primary care. Emerging evidence about the efficacy of test strategies means that NICE guidelines require updating. Accommodating the preferences of people who may undergo testing in the future is essential when designing care pathways, particularly under a system emphasising the importance of shared decision-making. Consideration of preferences means guidelines better align with patient expectations, encouraging help-seeking behaviour, improving compliance and avoiding missed appointments (Cronin *et al.*, 2018; Kim *et al.*, 2020). As a result, delays in diagnosis may decrease and patient satisfaction will be maximised.

Improved understanding of preferences also impacts current clinical practice. Research shows the use of NICE guidance is extremely variable. For instance, a survey of 258 GPs found the majority would refer patients on the basis of raised CA125 regardless of TVUS findings (Moss *et al.*, 2013). Lack of adherence to guidelines further increases the importance of understanding preferences to ensure GP decisions are reflective of their patients' wants and needs.

3.5.2 Existing evidence

Preferences around ovarian cancer testing in a diagnostic setting is an underexplored area. A search of the literature identified just two studies; both qualitative, investigating the experiences and preferences of patients who were subsequently diagnosed with ovarian cancer (Fitch *et al.*, 2002; Jelicic *et al.*, 2019). Fitch *et al.* (2002) conducted 18 open-ended interviews with women living in Canada focusing on their experiences during the diagnostic interval. Jelicic *et al.* (2019) is an Australian-based study. Thirty-four semi-structured interviews were conducted with women living with ovarian cancer. Both studies identified key areas where healthcare could be better aligned with patient preferences, heavily focusing on shortcomings in communication throughout the

diagnostic process. Women often felt symptoms were dismissed and found communication with healthcare professionals (HCPs) confusing and difficult.

Findings from both studies reflected many of the shortcomings discussed throughout this chapter. Women reported feeling uncertain about the severity of their symptoms prior to consulting a professional as well as experiencing delays due to perceived inaction by their GP due to a lack of symptom awareness or dismissal of patients due to the vague nature of symptom presentations. Areas of dissatisfaction during the investigative stage within primary care included being misdiagnosed with common conditions (e.g. constipation, benign cysts and fibroids) and undergoing multiple investigative tests prior to referral, although some women viewed this positively, considering it “thorough”. The importance of timely diagnosis was another recurring theme, with women identifying various stages during the diagnostic process where delays were incurred and expressed anxiety and fears that such delays resulted in disease progression. Findings from both studies help to identify shortcomings in current diagnostic procedures. Studies highlight the importance of communication and the doctor-patient relationship and demonstrate the existence of unmet preferences within a diagnostic context (an area of shared decision-making that has been underestimated to date).

3.5.3 Evidence gaps

Whilst the focus on women living with ovarian cancer provides an experiential perspective, the preferences and preconceptions of women facing testing for the first time are currently unknown. Furthermore, section 3.4 highlighted several shortcomings in current diagnostic procedures, where a clearer understanding of preferences would be beneficial when reconsidering guidance documents. For instance, attitudes towards the concept of triage testing (i.e. how do individuals value the trade-offs between timeliness and accuracy of testing?) and the acceptability of sequential versus concurrent testing procedures.

3.6 Ovarian cancer screening

3.6.1 Purpose of screening for ovarian cancer

Effective cancer screening programmes decrease deaths through detection of pre-cancerous or early-stage disease in asymptomatic individuals (Gupta *et al.*, 2019). In the UK, there are currently three screening programmes: breast, cervical and bowel screening (NHS UK, 2021b). Given the barriers to early diagnosis within primary care, screening in populations before symptoms arise may be a solution to reduce late stage diagnoses and ultimately save lives. There have been efforts to identify an appropriate ovarian cancer screening programme for over three decades, with multiple clinical trials taking place both nationally and internationally, in both average-risk and high-risk populations (Nash & Menon, 2020). To date, trials have demonstrated existing screening methods offer no benefits in terms of survival and are associated with several risks, in particular high levels of false positive results (Gupta *et al.*, 2019) (Table 3.1). However, given the high mortality rates associated with ovarian cancer, research to develop an appropriate screening programme are ongoing (Gupta *et al.*, 2019; Nash & Menon, 2020).

Table 3.1: Summary of large-scale ovarian cancer screening trials targeting general public populations.

Relevant trials were identified based on a summary paper by Nash and Menon (2020)

Trial name	Country	Trial type	Participants	Screening strategy	Key outcomes	References
UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS)	UK	RCT	Total: 202,638, aged 50-74 Arm 1: 50,640 Arm 2: 50,639 Arm 3: 101,359	Arm 1: Multimodal screening: Annual CA125 test interpreted using Risk of Ovarian Cancer Algorithm (ROCA) follow-up TVUS if result was abnormal Arm 2: Annual TVUS Arm 3: No screening (control)	Median follow up 16.3 years Incidence of stage I or II disease was 39·2% (95% CI 16·1 to 66·9) higher in the MMS group than in the no screening group, whereas the incidence of stage III or IV disease was 10·2% (−21·3 to 2·4) lower. There was no significant differences in stage between the TVUS and control group. No significant reduction in ovarian and tubal cancer deaths was observed in the MMS (p=0·58) or USS (p=0·36) groups compared with the no screening group.	(Jacobs *et al.*, 2016; Menon *et al.*, 2021)
Prostate, Lung, Colorectal and Ovarian Cancer	USA	RCT	Total: 78,386 aged 55-74 Arm 1: 39,105 Arm 2: 39,111	Arm 1: Annual CA125 test for 6 years (interpreted using a cut-off of >35 units/mL)	Median follow up 14.7 years no reduction in mortality between arms (RR 1.06; 95%CI 0.87–1.3; p-value=0.16)	(Buys *et al.*, 2011; Pinsky *et al.*, 2016)

Screening Trial (PLCO)				and TVUS for first 4 years Arm 2: No screening (control)	No differences in stage of diagnosis between arms	
Japanese Shizuka Cohort Screening Study	Japan	RCT	Total: 82,487 Arm 1: 41,688 Arm 2: 40,799	Arm 1: Annual combined screens of serum CA125 (interpreted using a cut-off of >35 units/mL) and TVUS Arm 2: No screening (control)	Mean follow up 9.2 years Compared to control arm screening resulted in a non-significant increase in Stage I disease in screen arm (63% vs 38%: p-value = 0.2285) No mortality results published	(Kobayashi *et al.*, 2008)
University of Kentucky Study	USA	Single arm prospective interventional trial	Total: 25,329, aged 25-92	Annual TVUS	Mean follow up 5.8 years Compared to a cohort of unscreened women at the same institution, screening significantly increased 5-year survival (74.8% ±6.6% vs 53.7% ±2.3%; P < 0.001)	(van Nagell Jr *et al.*, 2007; van Nagell Jr *et al.*, 2011)

MMS = Multimodal screening, TVUS = Transvaginal ultrasound, RCT = Randomised controlled trial, CA125 = Cancer Antigen 125

3.6.2 Principles of screening

In their landmark article “Principles and practice of screening for disease”, Wilson and Jungner (1968), on behalf of the World Health Organisation (WHO), set out longstanding guidelines for screening programmes that have been adapted by governing bodies over time, but still broadly apply today (Andermann *et al.*, 2008).
The consolidated ten principles of screening are:

1. The condition should be an important health problem
2. There should be a treatment for the condition
3. Facilities for diagnosis and treatment should be available
4. There should be a latent stage of the disease
5. There should be a test or examination for the condition
6. The test should be acceptable to the population
7. The natural history of the disease should be adequately understood
8. There should be an agreed policy on whom to treat
9. The total cost of finding a case should be economically balanced in relation to medical expenditure as a whole
10. Case-finding should be a continuous process, not just a "once and for all" project

From a UK perspective, each constituent country sets its own screening policies based on the recommendations of the National Screening committee (NSC). The NSC evidence review process provides recommendations for screening based on a pre-determined set of criteria used to appraise the viability, effectiveness and appropriateness of potential screening programmes (National Screening Committee, 2015). Criteria are comparable to the principles of screening outlined above and relate to the importance of the condition, the ability of the test to reduce deaths or disease incidence as well as the efficacy and suitability of the test and accompanying screening programme.

Notably, criteria 4.12 and 4.13 highlight the importance of understanding public preferences when designing screening programmes:

"4.12. There should be evidence that the complete screening programme (test, diagnostic procedures, treatment/ intervention) is clinically, socially and ethically acceptable to health professionals and the public.

4.13. The benefit gained by individuals from the screening programme should outweigh any harms, for example from overdiagnosis, overtreatment, false positives, false reassurance, uncertain findings and complications."

3.6.3 Current status of ovarian cancer screening

The UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS) is the largest randomised controlled trial to date (n=202,638) (Menon *et al.*, 2021). In total, 202,638 postmenopausal women, aged 50-75 with an average risk of ovarian cancer, were randomised. Screening modalities were compared to no screening (n=101.359) and consisted of transvaginal ultrasound only (n=50,640) and multimodal screening (MMS) combining CA125 blood tests with risk algorithms and second-line testing using transvaginal ultrasound where necessary (n=50,639) (Jacobs *et al.*, 2016; Menon *et al.*, 2021). Screening was performed annually between 2001-2011. Results at the end of the initial follow-up period in 2014 were promising, demonstrating an increase of 13% in the absolute proportion of women with ovarian cancer diagnosed with stage I or stage II disease in the MMS arm compared to no screening (no change in the TVUS arm) (Jacobs *et al.*, 2016). However, findings on mortality were inconclusive at this stage. There was no evidence of a reduction in disease-specific deaths in either screening arm compared to the no screening group, however, predictive models suggested any observed reduction in deaths may be delayed. Ultimately, results following an extended follow-up period lasting until June 2020 (median length >16years) revealed no overall impact on mortality, despite screening resulting in a shift in stage at diagnosis. (Menon *et al.*, 2021) .

Based on evidence from UKCTOCS and similar trials (Table 3.1), screening for ovarian cancer is not recommended in the UK for any population- even within high-risk populations such as those with a genetic predisposition for ovarian cancer (Nash & Menon, 2020). Identifying an appropriate screening programme remains an active area of research. Research avenues include designing risk-stratification tools to identify high-risk individuals, development of longitudinal algorithms to improve

surveillance and identification and refinement of additional biomarkers–most promisingly HE4 (Human Epididymis 4) –that could be combined with existing modalities to improve test performance (Nash & Menon, 2020).

3.6.4 Acceptability and preferences towards ovarian cancer screening

As noted in section 3.6.2, public acceptability is a crucial component of any potential screening programme. Uptake of screening is non-compulsory and dependent on the preferences and perceptions of the target population. Several studies have attempted to understand public acceptance and demand for ovarian cancer screening.

3.6.4.1 Test-related factors

Existing acceptability studies have primarily focused on factors relating to test delivery and experience. A large proportion of these studies were conducted as supplementary studies alongside clinical trials investigating the efficacy of ovarian cancer screening. Analysis of compliance within ovarian screening trials suggest CA125 is a more acceptable screening modality compared to TVUS, with withdrawal being lower in participants randomised to the TVUS arm of trials (Drescher *et al.*, 2004; Jenkins *et al.*, 2015). Analysis of compliance data also revealed that increased distance from the screening facility also reduced future screening behaviour and intentions (Drescher *et al.*, 2004; Pavlik *et al.*, 1995).

Survey-based studies of trial participants revealed high levels of pain or discomfort during previous screening appointments also reduced future screening intentions, although just 3.5% of women reported experiencing moderate-to-high levels of pain during screening (Gentry-Maharaj *et al.*, 2013). Findings from studies examining the acceptability of TVUS for the purpose of ovarian cancer screening or other medical conditions generally find a weak preference for female sonographer but strong acceptability for male sonographers (Atalabi *et al.*, 2012; Cowan Bennett & Richards, 2000; Deed *et al.*, 2014). However, studies also suggest gender preference may vary based on ethnicity and religious beliefs (Russell, 2005). Preferences towards chaperones during scans also appeared to vary based on cultural differences, with a UK-based study suggesting a strong preference for no chaperones (Sharma *et al.*,

2006), whilst two studies conducted in Nigeria illustrated a stronger preference for chaperones during scans, ideally a family member or friend as opposed to additional healthcare personnel (Akintomide & Obasi, 2019; Okeji *et al.*, 2017). Finally, a survey of women attending obstetrics and gynaecology clinics found on average women were neutral to the type and familiarity of HCP performing scans, although if given an option, a familiar doctors (as opposed to an unfamiliar nurse) was preferred (Bennett *et al.*, 2018).

3.6.4.2 Sociodemographic factors influencing screening acceptability

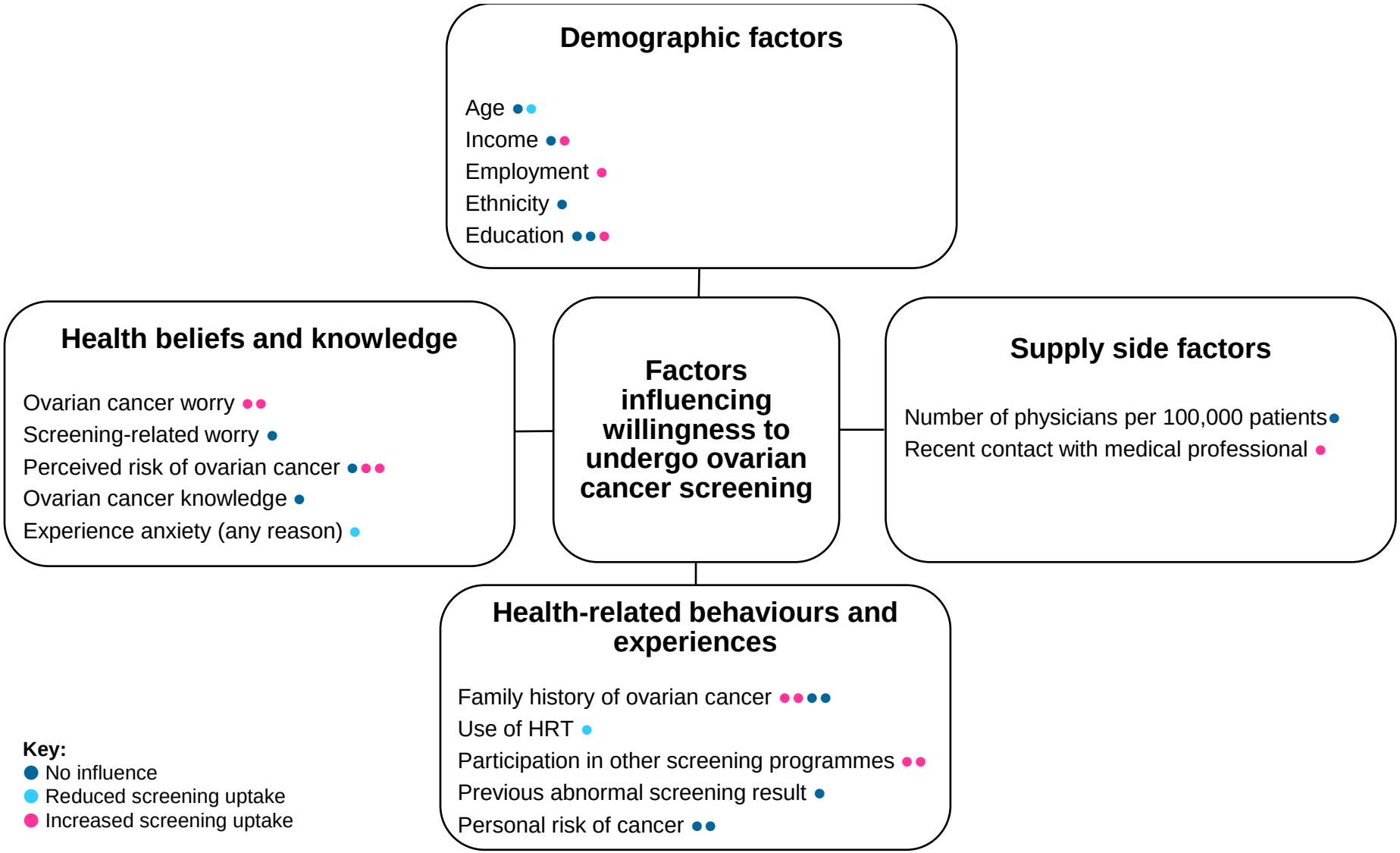
An overview of the relationship between ovarian cancer screening uptake or intended behaviour and sociodemographic characteristics is summarised visually in Figure 3.4. Studies were identified from a background literature search and not intended to be exhaustive. Current evidence appears quite limited although participation in other screening programmes, increased ovarian cancer worry and a perception of being high risk appear to increase ovarian screening uptake. A complete summary and references for included studies can be found in Appendix 3.1.

3.6.4.3 Evidence gaps in understanding of preferences

An understanding of the acceptability/preferences of ovarian cancer screening beyond a clinical trial setting is currently limited. Furthermore, existing studies have generally aimed to understand the acceptability of aspects of ovarian cancer screening in isolation and the relative importance of characteristics has been ignored. In reality, undergoing screening may involve enduring less desirable elements of screening in exchange for gains in other areas (e.g. travelling further for a test using your preferred modality or undergoing a more uncomfortable test for a more sensitive result).

Secondly, studies of acceptability have typically focused on factors relating to test experience (e.g. modality, pain, gender of HCP). In comparison, preferences and acceptability of test performance characteristics such as sensitivity and specificity remain underexplored. NSC recommendations are in part determined by the balance between benefits and harms of screening; however, the acceptable balance is currently unknown in the context of ovarian cancer.

Figure 3.4: Personal factors influencing acceptability and uptake of screening



Ovarian cancer and COVID-19 – role of preferences

Evidence relating to the impacts of Covid-19 on ovarian cancer is still emerging, however, recent publications suggest delays in diagnosis and treatment caused in part by deferred screening and diagnostic appointments due to lockdowns and prioritisation of Covid-care and subsequent backlogs in testing once restrictions eased are expected to cause a substantial increase in cancer deaths in coming years (Malagón *et al.*, 2022)

Covid-19 has disproportionately impacted vulnerable and historically medically underserved populations such as ethnic minorities and people of lower socioeconomic status (SES), highlighting and widening existing disparities in cancer care (Carethers *et al.*, 2020). During the early-stages of the pandemic, fear of Covid-19 exposure meant many postponed seeking medical attention. For others, access was limited by other Covid-19–related changes such as changes in public transportation, essential worker status, and lack of childcare, which may unequally impact already medically underserved communities. Experiences of the pandemic may have longer lasting behavioural and phycological impacts that will continue to alter testing acceptance beyond the immediate impacts of lockdown precautions and medical diversion towards Covid-related care (Carethers *et al.*, 2020; Hoehn & Zureikat, 2020).

There are potentially important implications for ovarian cancer as although there are no observed disparities in the diagnosis of ovarian cancer, studies relating to ovarian cancer survival frequently demonstrate a mortality disparity, with people from ethnic minorities and lower socioeconomic backgrounds experiencing disproportionately lower rates of survival compared to white women and those with higher socioeconomic status (Chornokur *et al.*, 2013; Hoehn & Zureikat, 2020).

Furthermore, the increased public attention placed on health services provision throughout the pandemic may also impact attitudes and behaviour towards ovarian cancer testing and symptom awareness. In particular, discourse surrounding medical testing and vaccinations including accessible discussions of covid test performance and efficacy shift in attitudes toward screening.

Understanding women's preferences towards cancer testing is essential. For instance, preferences may be used to understand who to prioritise while Covid recovery is ongoing and screening and testing backlogs continue. Understanding preferences may help to increase screening uptake in vulnerable groups and explore changes in attitudes towards testing following the pandemic.

3.7 Chapter summary

This chapter has summarised current procedures used to investigate ovarian cancer in primary care. According to current guidance, women presenting in primary care with symptoms of possible ovarian cancer who do not qualify for urgent referral must undergo two investigative tests (CA125 blood test and transvaginal ultrasound). Tests are performed sequentially, and both tests should be abnormal to prompt referral to secondary care.

Increasing early diagnosis of aggressive ovarian cancers is essential for improved outcomes for patients. Currently in the UK, around 50% of cancers are diagnosed at stage III or IV, where 5-year survival is just 27% and 13%, respectively. Improving both public and HCP knowledge of ovarian cancer symptoms is a clear priority for reducing delays in diagnosis and recent public awareness campaigns appear to be effective and are an ongoing approach to improving outcomes.

This chapter also highlighted several inefficiencies and uncertainties around current guidelines. Diagnostic guidance was last updated in 2011 and recent publications appear to demonstrate contradictions with newer NICE guidance on urgent referrals via the two-week wait pathway. Emerging evidence suggests guidance require updating in the near future.

Vague symptoms and delayed help-seeking behaviour once symptoms arise suggest screening of asymptomatic individuals may also provide a solution to improving ovarian cancer outcomes. Research to develop an appropriate ovarian screening programme have been unsuccessful but remains an area of ongoing research;

however, little is understood about the public acceptability of a potential ovarian screening.

The chapter highlighted several areas where improved understanding of patient and public preferences would be beneficial. In particular, two areas where knowledge of preferences is currently limited were identified:

1. Preferences towards diagnostic testing in primary care based on the current uncertainties in testing procedures and the likelihood that guidance will require revision. Incorporation of patient preferences should be essential for any care guidelines, particularly for care delivered within the NHS given the importance of shared decision-making. Improved understanding of preferences has also been shown to better adherence and patient satisfaction.

2. Preferences towards cancer screening. Acceptability of screening programmes and the appropriate balance of benefits and harms of screening tests are criteria of the evidence review process of the NSC. Understanding and tailoring future screening programmes to user-preferences increases the likelihood of uptake.

In particular, trade-offs between the balance of benefits and harms in both contexts is underexplored, making the area an ideal application of DCE methods to further investigate preferences

4 Existing evidence relating to the application of DCEs to cancer testing

4.1 Introduction

This chapter provides a summary and critical appraisal of the existing evidence arising from published DCEs eliciting preferences towards cancer testing. Early literature scoping revealed several existing systematic reviews synthesising DCEs in cancer. In response, the chapter begins with an overview of previously published systematic reviews. Findings informed the subsequent systematic review which refined the search methods, expanding search dates and focusing on methodological aspects of DCEs. Alongside the results presented within this chapter, an in-depth analysis of attributes included in DCE studies was conducted. Analysis identified the frequency, importance and significance of attributes used to elicit preferences towards cancer testing. This component of the review was published in a peer-review journal (Hall *et al.*, 2021). The full paper is presented in Appendix 4.1.

4.2 Overview of existing reviews

An increase in the implementation of DCEs in cancer testing in recent years has resulted in the publication of several systematic reviews aiming to summarise the existing evidence base. Reputable and widely used published guidance for conducting systematic reviews is available for a wide variety of study designs such as clinical trials (Boland *et al.*, 2017), economic evaluations (Philips *et al.*, 2004) and qualitative evidence (Noblit & Hare, 1988), however, at present there is no equivalent guidance available to researchers conducting systematic reviews of preference-based studies. A lack of standardised guidance means reviews are likely to differ greatly in terms of methods and quality, making it difficult for those seeking to use such reviews to gain an insight of the topic.

Before undertaking the systematic review, an overview of previously published systematic reviews was performed. The purpose of the scoping review was to summarise the methodological aspects of existing systematic reviews to identify any inconsistencies or gaps in the current evidence.

4.2.1 Methods

For the purposes of the overview, systematic reviews were defined as any review of the literature that systematically described the search strategy and data extraction methods and did not self-identify as a 'scoping' or 'rapid' review. Systematic reviews could consider one specific cancer site such as colon cancer or consider cancer more generally.

To identify all relevant DCEs, searches were performed across seven relevant databases; MEDLINE, Embase, PsycINFO, HMIC, Web of Science, EconLit and NHS EED using previously validated search terms relating to 'discrete choice experiments' and 'systematic review'. Searches dates were limited to 1990 to 8th December, 2020.

4.2.1.1 Quality assessment

Understanding the quality of previous reviews was a key component of the overview. Following previous overviews (He *et al.*, 2019; Moore *et al.*, 2018), the AMSTAR2 checklist (Shea *et al.*, 2017) was used to assess the quality of studies. The AMSTAR2 checklist is validated and highly cited critical assessment tool, however, it is designed to assess the quality of systematic reviews involving randomised/non-randomised trials, meaning a few adaptations were necessary. In total, the AMSTAR2 checklist consists of eleven questions. For this study, questions relating to meta-analysis were excluded and answers for questions 1, 8 and 9 were amended to better reflect preference-elicitation studies. The full amended checklist is shown in Appendix 2. Due to the adaptions, studies were awarded an overall score out of 11, with higher scores indicating increased methodological quality; scores were used to evaluate the general quality of studies and identify common areas of weakness across reviews. Studies were not excluded based on quality assessment scores.

Evidence synthesis focused on information relating to the methods utilised within previous reviews including objectives of the review, number of primary studies identified, databases searched and time periods covered.

4.2.2 Findings

Five systematic reviews were identified overall. Three of the reviews focused on a single cancer site, specifically colorectal (Ghanouni *et al.*, 2013; Marshall *et al.*, 2010; Wortley *et al.*, 2014), whereas Phillips *et al.* (2006) and Mansfield *et al.* (2016) included DCEs that elicited preferences for any type of cancer. No reviews summarising evidence relating to diagnostic preferences were identified. With the exception of Phillips *et al.* (2006), all reviews searched at least two databases with PubMed and EMBASE being most common. Search dates across all studies ranged from 1990 to 2015 (see Table 4.1 for further details).

4.2.2.1 Quality assessment

In general, all studies were of moderate to good quality, with AMSTAR2 scores ranging from 5-11 and an average score of 7 (Table 4.1). Common areas of weakness included: failure to list studies excluded at full-text (7/8); failure to register study protocol/no referral to a predetermined protocol (5/8); and unclear reporting on whether screening and data extraction was duplicated (3/8). Perhaps most critically, only four studies explicitly attempted to assess the quality of included studies using a checklist. Three of the four reviews (Bien *et al.*, 2017; Ghanouni *et al.*, 2013; Wortley *et al.*, 2014) used the ISPOR Good Practice guidance (Bridges *et al.*, 2011). A full summary of AMSTAR2 results for each review are found in Appendix 4.2.

Table 4.1: Summary of existing systematic reviews of DCEs on cancer testing

Author (Date)	Objective of review	Database(s) searched	Preference search terms*	Search period	Total no. of studies	No. of Cancer DCEs	AMSTAR score
Phillips *et al.* (2006)	To identify gaps in the literature, and determine which types of research should be conducted in the future	PubMed	MeSH terms: "patient satisfaction/economics", "patient satisfaction/statistics & numerical data", "consumer satisfaction/economics", "consumer satisfaction/statistics & numerical data", "health knowledge, attitudes, practice", "choice behaviour" Keywords: "preference(s)", "attitudes", "conjoint analysis", "contingent valuation", "stated preference", "discrete choice", "willingness to pay"	1996-2005	8	1	7
Marshall *et al.* (2010)	To provide an overview of the current state of preference measurement for colorectal cancer (CRC) screening, highlighting the implications for health policy, CRC screening program implementation, and further research	MEDLINE, EMBASE	Available upon request	1990-2009	6	4	6
Ghanouni *et al.* (2013)	To describe and evaluate key features of conjoint analysis studies of CRC screening tests	PubMed, CINAHL, Web of Knowledge, EMBASE, PsycINFO	Key words: "conjoint analysis", "discrete choice", "discrete ranking", "stated preference"	Not specified	7	5	7
Wortley *et al.* (2014)	To update previous systematic reviews on CRC screening tests and provide a methodological assessment and analysis of the key findings of included studies	MEDLINE, EMBASE, EconLit, PreMedline, Google Scholar	MeSH terms: "decision-making", "patient attitude" Key words: "conjoint", "discrete choice", "choice experiment", "choice experiments", "stated preferences", "stated preference", "willingness to pay", "contingent valuation", "choice", "preference"	up to April 2013	9	8	11
Mansfield *et al.* (2016)	To assess the types of cancer screening test attributes researchers have considered, differentiating between attributes of the screening tests themselves and attributes that capture other elements of the patient experience; and to review the use of questions to determine reported likelihood of uptake	PubMed, EconLit	"conjoint analysis" or "conjoint analyses" or "conjoint-analysis" or "conjoint-analyses" or "discrete choice" or "discrete-choice" or "discrete ranking" or "discrete rank	1990-2013	22	18	6

*Search terms relate to PubMed database for all studies except Wortley *et al.* (2014) where only EMBASE search terms were provided.

1.2.2.2. Review methods

Differences in objectives led to variations in search terms used and the type of primary studies included. One review stated that inclusion was restricted to DCE studies only (Wortley *et al.,* 2014); however, in practice it mirrored Marshall *et al.* (2010) and Ghanouni *et al.* (2013) by summarising conjoint analysis studies more broadly by including ranking or ratings scale studies. The remaining two systematic reviews aimed to synthesise evidence from studies that used any stated preference methods, including broader conjoint analysis or contingent valuation studies such as willingness to pay (Mansfield *et al.*, 2016; Phillips *et al.*, 2006).

Figure 4.1 shows a Venn diagram demonstrating the overlap of DCE studies included across reviews. Promisingly, unique studies were exclusive to the most recent systematic review (Mansfield *et al.*, 2016) suggesting differences in search dates account for several of the unique studies. However, a number of the unique studies also appeared to fall in the search dates and criteria of earlier studies where they were omitted. For example, five of the 10 studies unique to (Mansfield *et al.*, 2016) appeared fit the scope and search dates of earlier reviews.

Figure 4.1: Venn diagram showing the overlap in included studies between existing systematic reviews on cancer testing (n = number of studies unique to each segment)

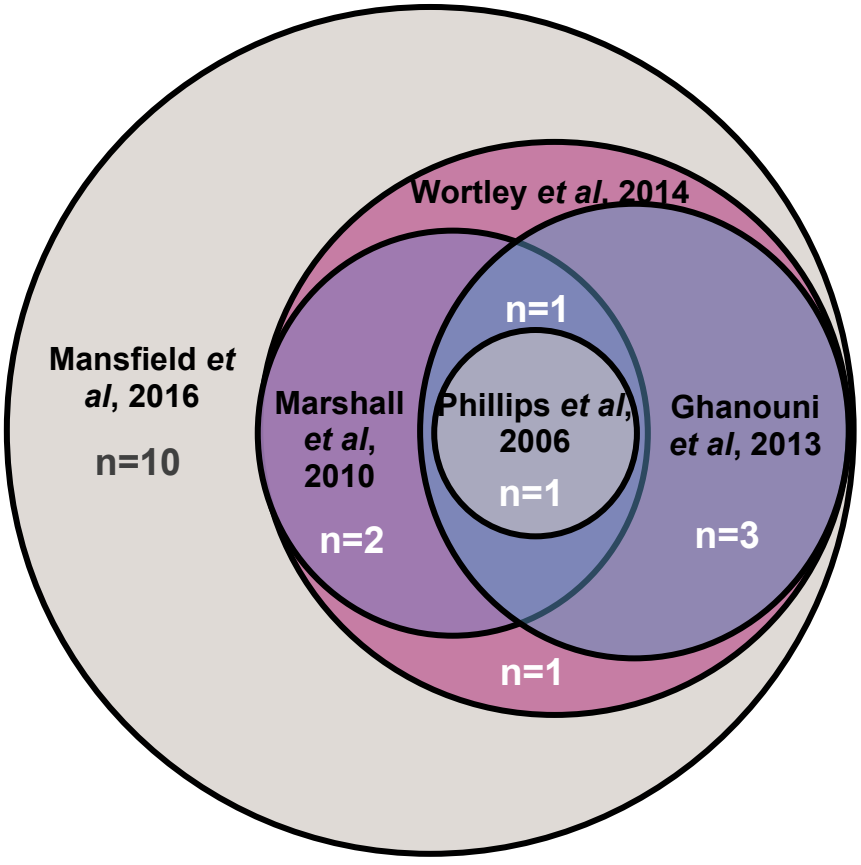


Table 4.2 demonstrates the scope of each systematic review in relation to the key elements of a DCE, as outlined by Lancsar and Louviere (2008). All reviews summarised the study population and perspective in detail and data extraction relating to sampling methods, sample size and data collection methods were also well reported. Outputs reflected the stated objectives of each review which focused on; descriptions of attributes and the selection process (Bien *et al.*, 2017; Neal *et al.*, 2018; Wortley *et al.*, 2014), interpretation of results including welfare measures such as willingness to pay (WTP) and uptake estimates (Mansfield *et al.*, 2016; Marshall *et al.*, 2010; Phillips *et al.*, 2006; Wortley *et al.*, 2014) and description and evaluation of key study features (Ghanouni *et al.*, 2013; Wortley *et al.*, 2014). More methodological aspects of DCEs such as experimental design and data analysis were reported less frequently.

4.2.3 Summary of overview findings

Previous reviews differed considerably in the scope of their search strategies (e.g. search terms, dates or cancer type), with three reviews focusing exclusively on studies relating to colorectal cancer (Ghanouni *et al.*, 2013; Marshall *et al.*, 2010; Wortley *et al.*, 2014). Furthermore, due to the date of publication, existing reviews only synthesise studies published prior to 2015.

Existing systematic reviews largely focus on the clinical implications of DCE findings and suggestions on how to implement findings in policy decisions (Ghanouni *et al.*, 2013; Mansfield *et al.*, 2016; Marshall *et al.*, 2010; Phillips *et al.*, 2006; Wortley *et al.*, 2014). In contrast, methodological issues of attribute identification, selection and construction (e.g. assignment of levels, type and complexity of attributes) and how these impact on cognitive burden have not been explored in detail.

Finally, assessment of the quality of included studies was limited across reviews. Inconsistencies in quality assessment are unlikely to be due to neglect on the authors' part and instead highlights of the lack of best-practice guidance and absence of a validated and widely accepted tool to assess the methodological quality of DCEs studies. Where methodological quality was assessed, authors typically utilised ISPOR Good Practice Guidelines which provides guidance for conducting conjoint analyses

but does not provide a definitive checklist for quality assessment (Bridges *et al.*, 2011). Ongoing work by Yepes-Nuñez *et al.* (2017) appears to form preliminary basis for standardised quality assessment in the future.

Table 4.2 Scope of systematic reviews based on key areas of DCEs highlighted by Lancsar and Louviere (2008)

	Phillips *et al.* (2006)	Marshall *et al.* (2010)	Ghanouni *et al.* (2013)	Wortley *et al.* (2014)	Mansfield *et al.* (2016)
Conceptualizing the choice process	●	●	●	●	●
Attribute selection	●	●	●	●	●
Level selection	●	●	●	●	●
Experimental design	●	●	●	●	●
Questionnaire design	●	●	●	●	●
Piloting	●	●	●	●	●
Population/study perspective	●	●	●	●	●
Sampling and sample size	●	●	●	●	●
Data collection	●	●	●	●	●
Coding of data	●	●	●	●	●
Econometric analysis	●	●	●	●	●
Validity	●	●	●	●	●
Interpretation	●	●	●	●	●
Welfare analysis	●	●	●	●	●

● Discussed in detail ● Extracted but not discussed in detail ● Not extracted or discussed

4.3 Systematic review of DCEs eliciting preferences towards cancer screening and diagnostic testing

4.3.1 Background

This review built on the evidence from existing reviews in several ways; firstly, by updating the search dates, since DCEs are an increasingly popular method and studies are continuously being published. Secondly, by ensuring a rigorous search strategy covering seven databases using the expertise of an information specialist. The scope of the review was extended to diagnostic testing alongside DCEs of cancer screening. Finally, this review focused on key areas not previously covered by systematic reviews including the link between the purpose of DCEs and the practical application of results, and the quality of included studies.

The overall aim was to perform a systematic review of the published evidence relating to the use of DCE to assess preferences towards testing for cancer, focusing specifically on the purpose of the studies and the methodological features of applications of discrete choice methods to cancer screening and diagnostic testing. The results of the review were used to gain a better understanding around preferences relating to cancer testing for different populations and to design the discrete choice experiments as described in subsequent chapters.

4.3.2 Methods

The systematic review was performed in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) Statement (Moher *et al.*, 2009) and following a predetermined protocol (Prospero protocol ID: CRD42019153834).

4.3.2.1 Inclusion and Exclusion criteria

Studies were included if they presented primary results from a DCE regarding preferences for cancer screening within an asymptomatic average-risk population or cancer diagnosis in symptomatic patients. Preferences could be elicited from any population and for any cancer site. To be included, studies were required to be peer-reviewed and published in English. Studies that used alternative preference-based methods such as ranking/rating, time trade-off including those which used a single

DCE question/profile to elicit WTP estimates were also excluded, alongside review papers or papers that performed secondary analyses on previously reported (and already identified) DCE results.

4.3.2.2 Search Strategy

To identify all relevant DCEs, a two-tier systematic search of the literature published between 1990 and 8th December, 2020 was performed. In the first stage, searches were performed in seven relevant databases; MEDLINE, Embase, PsycINFO, HMIC, Web of Science, EconLit and NHS EED using previously validated search terms relating to 'discrete choice experiments' (de Bekker-Grob *et al.*, 2012). The hits were then exported to EndNote X8 (Thomson ISI Research-Soft) and an additional search of titles and abstracts used terms relating to 'cancer' was performed. Search terms were determined in line with previous reviews and guidance from an information specialist. Further details of the search strategy can be found in Appendix 4.3. To avoid excluding relevant studies, searches were not further narrowed using terms related to 'screening' or 'diagnosis'; instead, this was done manually during the article screening stage. The references and forward citations for studies included at the full-text stage were also searched for additional studies.

4.3.2.3 Study identification

Study selection was performed by two independent reviewers and consisted of three stages; screening of titles and abstracts, retrieval and review of full-text and finally manual search of the reference lists of studies selected for inclusion at full-text. Results were compared at each stage and any discrepancies were discussed with help from a third researcher when necessary.

4.3.2.4 Data Extraction

Data extraction was performed by a single reviewer (RH) using Microsoft Excel and a proportion of the results then verified by a second reviewer (AML). Data was summarised in a narrative synthesis as a meta-analysis was not possible because of heterogeneity in the study design and methods.

Data was extracted using pre-specified extraction tables in Microsoft Excel. Data synthesis focused on key methodological components of studies, including:

- Study characteristics e.g. cancer site, date, objective,
- Sample e.g. target population, sample size calculation, response rates
- Attributes and attribute development (see appendix 4.1)
- Experimental design e.g. use of software, design approach, blocking, number of choice tasks
- Analysis e.g. econometric models, outcomes
- Use of rationality and debriefing questions

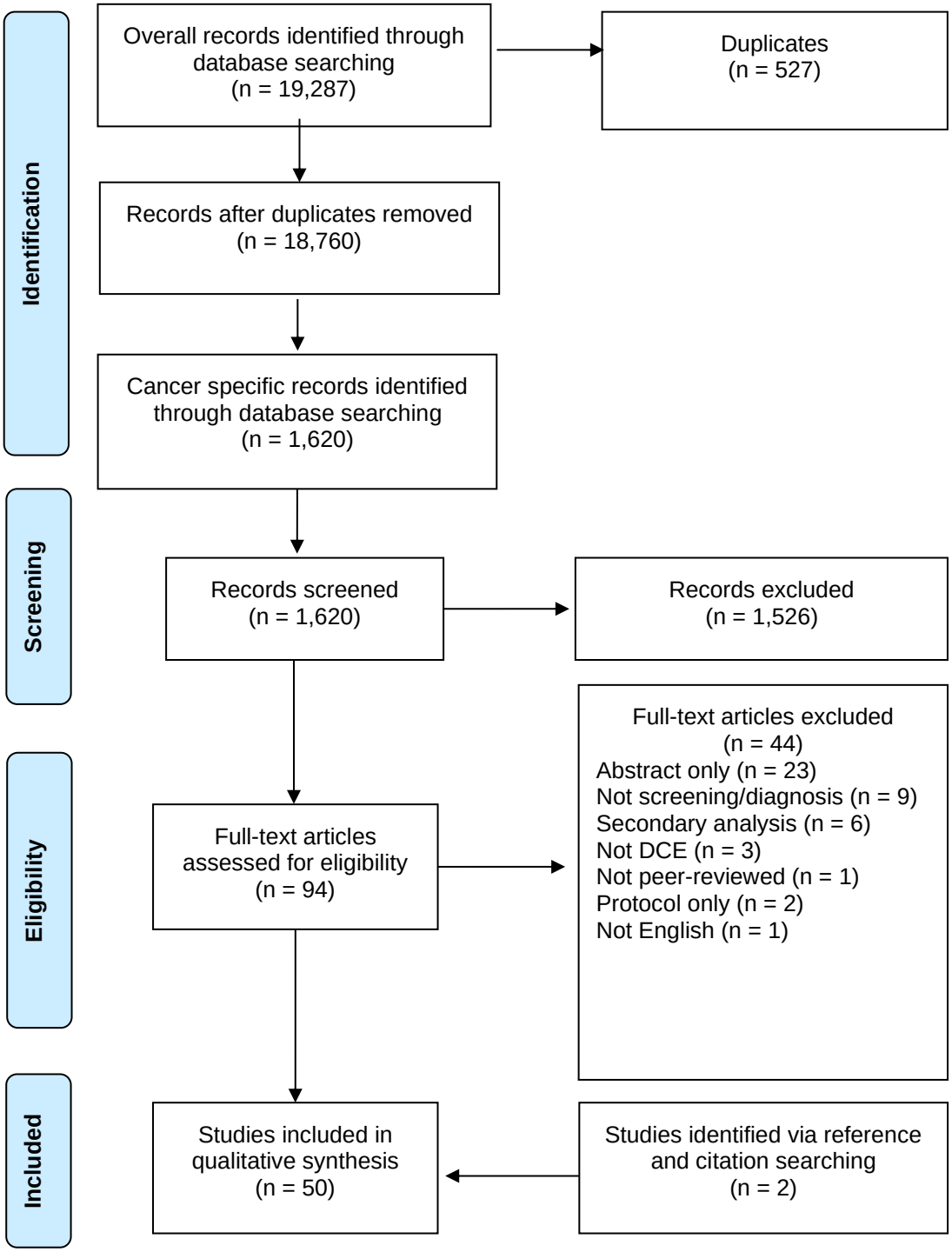
4.3.2.5 Quality assessment

Methods for assessing the quality of DCEs are an ongoing area of methodological development. As such, we followed previous reviews (Ghanouni *et al.*, 2013; Wortley *et al.*, 2014) and adapted the International Society for Pharmacoeconomics and Outcomes Research (Bridges *et al.*, 2011) checklist in order to critically appraise the studies identified within this review (Appendix 4.4). The checklist was initially designed as a guide to good research practices and reporting for the design and implementation of DCEs and consists of ten domains essential for conducting high-quality DCEs. Each domain contains three sub-questions and studies were assigned a score of 1 for each sub-question where the criteria was met and 0 if unfilled or unclear, giving a maximum score of 30. As the checklist has been adapted for the purposes of this review, scores were used to gauge a general overview of quality of reporting and to highlight common areas of weakness across studies rather than as a means of inclusion or exclusion.

4.3.3 Results

Figure 4.2 provides an overview of the results of the study selection process. Once all search terms were applied and duplicates were removed, 1,620 studies remained. After screening titles and abstracts, 1,526 were excluded leaving 94 studies to be screened at full-text. During the full-text screening stage, 44 studies were excluded. Two additional studies were found through manual searches of reference lists and forward citations, resulting in a final sample of 52 studies.

Figure 4.2: PRISMA diagram for systematic review adapted from Moher et al. (2009)



4.3.3.1 Quality assessment of reporting: adherence to best-practice guidelines

The results of the quality assessment can be found in Appendix 4.5. Scores ranged from 12 to 29 (mean=24) out of 30. On average, studies received the highest scores in criteria relating to the justification of research question and methods (question 1, mean score 2.9 out of 3), attribute and level selection methods (question 2, mean score 2.8/3.0 and the validity of results and conclusions (2.8 out of 3). Common issues related to the reporting and justification of sample size calculations (54%; 28/52), experimental design (62%; 32/52) and model estimation methods (46%; 24/52). Although in many instances it was hard to differentiate between poor practice and a lack of detail in reporting. For example, assessing the appropriateness of attribute presentation was difficult because access to survey instruments was often limited and the wording and definitions used in manuscripts often differed from what was shown to respondents.

4.3.3.2 Study characteristics

A summary of study characteristics is provided in Table 4.3 and a study-level summary can be found in Appendix 4.6.

The application of DCEs to cancer testing has increased in recent years, with 19/52 (37%) of the studies being published in the last five years. Studies were primarily clustered around five countries; Australia (13/52; 25%), USA (10/52;19%), UK (8/52;15%), Netherlands (8/52; 15%), France (7/52; 13%).

Of the 52 included studies, 49 focused on preferences towards cancer screening. Screening studies commonly considered cancer sites where there is clear existing evidence supporting the efficacy of population-based screening (cervical 11, breast 8), with colorectal cancer being the most common (26/52; 50%). The remaining studies considered prostate (Charvin *et al.*, 2020; de Bekker-Grob *et al.*, 2013b; Howard *et al.*, 2015; Pignone *et al.*, 2013), oesophageal (Peters & Siersema, 2020) and skin cancer screening (Snoswell *et al.*, 2018; Spinks *et al.*, 2016) where the potential benefits and harms of national screening programmes are debated (Light *et al.*, 2019). More than half (27/49) compared preferences for multiple screening options, most commonly when considering colorectal cancer, where the dominance of a single

modality is unclear. The remaining 22 studies explored preferences for a single screening modality: these studies typically related to screening programmes where there was an established modality with no close alternatives, such as the use of mammograms for breast cancer screening.

The remaining three studies focused on diagnostic testing. Studies elicited preferences relating to innovation and emerging technologies across three different cancer sites; colorectal (Howard *et al.*, 2011; Miles *et al.*, 2019), prostate (Ellimoottil *et al.*, 2018) and lung (Miles *et al.*, 2019).

4.3.3.3 Key components of DCEs

Conceptualising the choice process

For most studies, the choice format consisted of a series of two choice alternatives with little justification and regardless of the overall number of modalities being compared within the study or available when making decisions in real life (44/52, 85%). Studies commonly allowed participants to express dislike for test options by providing an opt-out alternative (33/52; 64%). Less common choice formats included the provision of a 'status quo' (5/52; 10%) (Hendrix *et al.*, 2020; Howard *et al.*, 2011; Salkeld *et al.*, 2000; Snoswell *et al.*, 2018; Spinks *et al.*, 2016) or indifference alternative (1/52; 2%) (Pignone *et al.*, 2012).

In total, eight studies (15%) utilised a labelled design by using alternative-specific titles for each choice profile (e.g. FOBT, colonoscopy, flexible sigmoidoscopy etc.) (Benning *et al.*, 2014a; Benning *et al.*, 2014b; Fiebig *et al.*, 2009; Hol *et al.*, 2010; Howard *et al.*, 2011; Johar *et al.*, 2013; Jonker *et al.*, 2019; Peters & Siersema, 2020; Raginel *et al.*, 2020). The remaining studies (44/52; 85%) applied an unlabelled approach by assigning neutral titles to each alternative (e.g. "Test 1", "Screening option A", "Program 1" etc.), however, in many instances the use of detailed/specific levels meant that interventions were often identifiable, particularly where a "process" attribute, describing the nature of the test, was included.

Table 4.3: Summary characteristics of DCE studies

Characteristic		n (%)
Year of publication	2000-2005	4 (8%)
	2006-2010	9 (17%)
	2011-2015	16 (31%)
	2016- present	23 (44%)
Country*	Australia	13 (25%)
	Canada	2 (4%)
	France	7 (13%)
	Netherlands	8 (15%)
	UK	8 (15%)
	USA	10 (19%)
	Other	8 (15%)
Cancer site and intervention(s)*	**Breast**	**9 (17%)**
	Mammogram	7 (13%)
	Other	2 (4%)
	Cervix	**11 (21%)**
	Standard care/Pap-smear	8 (15%)
	Liquid-based cytology	3 (6%)
	Human papillomavirus (HPV) testing	2 (4%)
	Visual inspection	1 (2%)
	Self-sampling	3 (6%)
	Other	2 (4%)
	Colon	**28 (54%)**
	Faecal occult blood test (FOBT)	22 (42%)
	Faecal Immunochemical Test (FIT)	5 (10%)
	Colonoscopy	11 (21%)
	CT colonography	11 (21%)
	Sigmoidoscopy	10 (19%)
	Barium enema	3 (6%)
	Blood test	5 (10%)
	Other	3 (6%)
	Oesophagus	1 (2%)
	Upper endoscopy	1 (2%)
	Nasal endoscopy	1 (2%)
	Pill on a string	1 (2%)
	Breath test	1 (2%)
	Blood test	1 (2%)
	Prostate	**5 (`0%)**
	Prostate-specific antigen (PSA) test	4 (8%)
	Other	1 (2%)
	Skin	**2 (4%)**
	Teledermoscopy	2 (4%)
	Self-examination	2 (4%)
	Examination by professional	2 (4%)
	Lung	1 (2%)
	Whole-body MRI	1 (2%)
	Conventional staging	1 (2%)
Population*	General public- mixed experience	34 (65%)
	General public- Screening naïve only	1 (2%)
	Patients/ previously screened individuals only	12 (23%)
	Healthcare providers	10 (19%)
Sample size*	< 50	2 (4%)
	50-100	7 (13%)
	101-200	11 (21%)
	201-500	22 (42%)
	501-1000	9 (17%)
	>1000	8 (15%)
Response rate (%)*	<20	5 (10%)
	20-50	13 (25%)
	51-80	17 (33%)
	>80	8 (15%)
	Not stated	12 (23%)
Survey administration method*	Self-completed- online	22 (42%)
	Self-completed- postal	21 (40%)
	Interview-led/ in-person	15 (29%)
	Not specified	1 (2%)
*Some studies fall into multiple categories so n>52		

Attribute and level selection

The number of attributes per choice task ranged from 2 to 13, with a median of five. Most studies stated the attribute selection method; however, the level of detail on how these methods were applied was often limited, particularly where qualitative methods were utilised. When selecting attributes, 72% (37/52) of studies took a multi-method approach, most commonly combining literature reviews (40/52; 77%) with an additional qualitative and/or quantitative source such as expert opinion (17/52; 33%). Over half the studies (29/52) reported consulting the target population throughout the attribute selection process, typically using qualitative methods such as interviews (16/52; 31%) and focus groups (10/52; 19%). Four studies (8%) did not report any sources of attribute selection; instead, attributes were chosen according to the research question and the assumptions of authors.

The assignment of attribute levels was generally described in less detail than attribute selection, with eight studies not describing the process at all or relying on the assumptions of authors. The number of levels ranged from two to 12, averaging three per attribute. Where described, literature reviews (30/52; 58%) were the most common source for obtaining attribute levels, with existing policy (8/52; 15%), qualitative methods (5/49; 8%), trial data (6/52; 12%) and expert opinion (4/52; 8%) also cited.

Experimental design

Almost all studies (92%, 48/52) utilised a fractional factorial design to reduce the number of possible combinations of attributes and levels to a manageable size. Main effects only designs appeared to be most popular (21/52; 40%) compared to designs allowing for selected or all two-way interactions to be analysed (14/52; 27%); however, fifteen (29%) studies failed to clearly specify which approach had been utilised, although presented results suggest most of these studies also applied a main-effects approach (Table 4.4). The decision of whether to apply blocking during the experimental design (i.e. spread the choice tasks across multiple questionnaires) was generally well-reported and typically motivated by piloting results or methodological guidelines. Overall, 63% (33/52) of studies applied a blocked design to manage the cognitive burden for respondents by minimising the total number of choice tasks per

respondent. Alternatively, sixteen (37%) studies opted for an unblocked design, with these studies requiring respondents to complete 16 choice tasks on average- 4 more than the average study applying blocking.

The use of software to aid the experimental design development process appeared to be popular amongst authors with the use of specialist software such as Ngene (9/52; 17%) and Sawtooth (9/52; 17%) being common. Even so, SAS (10/52; 19%) was most frequently used overall. Thirteen studies failed to acknowledge any software they had used.

Reporting on design properties used to assess the quality of the experimental design varied in detail, with over a quarter of studies (14/52) providing no clear details whatsoever. Of the remaining studies, maximising the d-efficiency was the most commonly reported objective during selection of the optimal experimental design within a study (27/52; 52%). Orthogonality (9/52; 17%), level balance (9/52; 17%), attribute overlap (8/52; 15%) and the restriction of improbable combinations (5/52; 10%) were also used and reported by the authors of several studies but to a much lesser extent.

Piloting

Reporting on piloting was mixed but generally descriptions of the purpose, process and outcomes of piloting across studies were limited in detail. Overall, 39 of 52 (75%) studies reported performing some form of piloting and 34 of these studies specified applying an iterative, adaptive approach (Table 4.5). Authors reported testing three primary areas during piloting; attribute coverage (30/52; 58%), understand and complexity of the choice task (34/52; 65%) and length and timing of the final questionnaire (29/52; 56%). Thirty-one (60%) studies described the population and sample size of the piloting study, however just 16 (31%) studies described the method of administration which varied between interviews (9/52; 17%), focus groups (2/52; 4%) and self-completion of an early version of the final questionnaire (6/52; 12%).

Table 4.4: Summary of experimental design features from included studies

Characteristic		n (%)
Design type	Full factorial	4 (8%)
	Fractional factorial	48 (92%)
Design plan	Main effects only	21 (40%)
	Main effects and two-way interactions	14 (27%)
	Not applicable	2 (4%)
	Not reported	15 (29%)
Blocking	Yes	33 (63%)
	No	16 (31%)
	Not reported	3 (6%)
Design software	Ngene	9 (17%)
	SAS	10 (19%)
	Sawtooth	9 (17%)
	SPEED	2 (4%)
	Other	2 (4%)
	Not reported	21 (40%)

Table 4.5: Summary of pilot studies carried out within included studies prior to final data collection

Characteristic		n (%)
Was piloting carried out?	Yes	39 (75%)
	Not reported	12 (25%)
Was piloting iterative?	Yes	34 (65%)
	Not reported	18 (35%)
What did piloting cover?	Attribute coverage	30 (58%)
	Understanding and complexity	34 (65%)
	Length and timing	29 (56%)
Was piloting population described?	Yes	31 (60%)
	Not reported	21 (40%)
Method of administration	Interviews	9 (17%)
	Focus groups	2 (4%)
	Postal/online questionnaire	6 (12%)
	Not reported	36 (69%)

Sampling and data collection process

Preferences of the general public (35/52; 67%) were most commonly collected, with the preferences of HCPs (10/52; 19%) and patients (11/52; 22%) considered less frequently. Five (10%) studies collected the preferences of multiple populations within a single study (Arana *et al.*, 2006; Boone *et al.*, 2013; Fiebig *et al.*, 2009; Marshall *et al.*, 2009; Plumb *et al.*, 2014).

Sample size calculations were described in less than half of studies (24/52; 46%) and final sample size varied vastly from 35-2067, with a median of 316 participants per study. For the most part, data was collected using a self-completed survey instrument administered via post (20/52; 38%) and/or online (22/52; 42%), just thirteen (25%) studies opting to collect data in a face-to-face format. Response rates were described inconsistently across studies and varied widely from 5-91% where reported. Self-completed questionnaires predictably reporting lower rates than researcher-administrated studies on average.

Econometric analysis

Reported methods used to analyse discrete choice data were diverse. Multinomial logit (13/52; 25%) and mixed logit (14/52; 27%) were used most frequently. Probit models appeared to be popular in earlier studies with a shift towards logit-based models in more recent years. In more recent studies, there also appears to be a move towards more complex methods such as latent class or hierarchical Bayes models that capture heterogeneity in preferences and/or scale across respondents.

Incorporating sociodemographic information

All but one study (Salkeld *et al.*, 2000) collected and reported sociodemographic characteristics of respondents. For 38% (20/52) of studies, utilisation of sociodemographic data was limited to narrative descriptions only. Remaining studies incorporated respondent characteristics into analyses by examining associations with uptake (10/52; 19%), preferred test modality (6/52; 12%) and attribute utility using

subgroup analysis (8/52; 15%), interaction terms (9/52; 17%) and/or latent class membership (4/52; 8%). The inclusion and influence (both significance and direction) of sociodemographic characteristics varied across studies and are visually summarised in Figure 4.3. Evidence was generally mixed but in general significant interactions between preferences and demographic characteristics appeared limited.

Reliability/validity of responses

Table 4.6 demonstrates the range of validity tests included within studies. External validity relates to the ability of a DCE to produce reliable inferences about the true behaviour of individuals beyond the confines of the hypothetical experiment. None of the studies within this review carried out external validity checks. Almost half (25/52; 48%) of the studies reported the inclusion of at least one internal validity check. Most commonly this was a test to check the axiom of non-satiation by including a choice task where one alternative within a choice set was irrefutably preferable to the other alternatives (dominant alternative) (18/52; 35%). Alternative or complimentary validity measures included across studies were test-retest reliability to check the stability of choice by repeating a choice set twice within a questionnaire to see if the respondent answers were consistent (5/52; 10%), identification of attribute domination by identifying non-traders (i.e. decisions based on a single attribute) (5/52; 10%), transitivity (2/52; 4%) and flat-lining (2/52; 4%).

Respondents who failed validity checks were generally interpreted to be irrational in their choice behaviour. Exclusion of irrational responders is debated; many studies opted to include all respondents by default (6/52; 12%) or following sensitivity analysis (7/52; 13%). Alternatively, six studies (12%) excluded failing responders automatically and a further three studies (6%) following sensitivity analyses. Two studies (4%) received no irrational responses and two did not report the outcomes or consequences of validity check questions.

Figure 4.3:Summary of interactions between respondent characteristics and preferences towards cancer testing including in DCEs

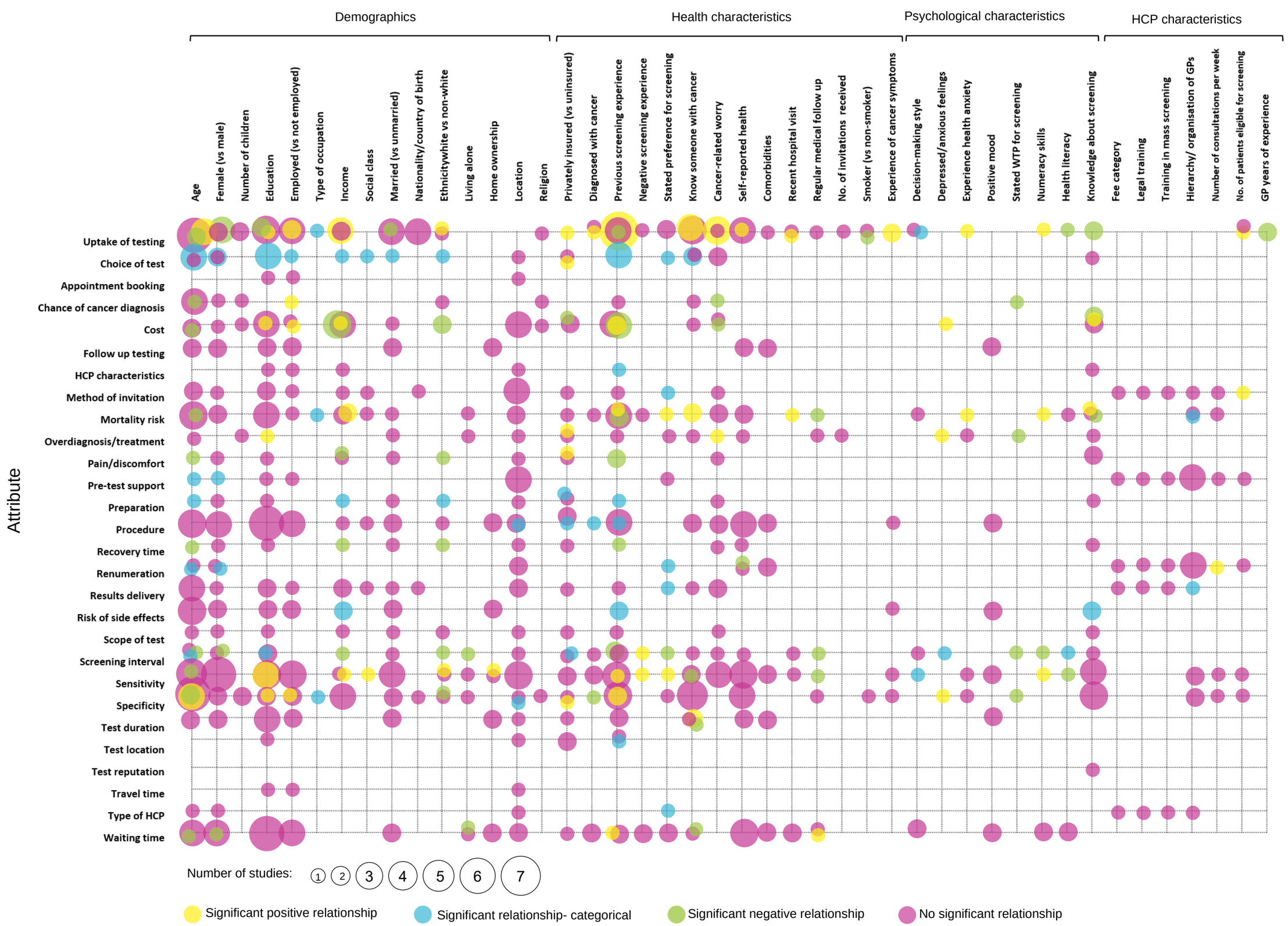


Table 4.6: Summary of rationality check questions included within studies

Characteristic		n (%)
Type of validity check	External validity	0 (0%)
	Test-retest stability	5 (10%)
	Monotonicity (dominant profile)	18 (35%)
	Transitivity	2 (4%)
	Attribute domination (non-traders)	6 (12%)
	Flatlining	1 (2%)
	Other	2 (4%)
	None/ not reported	27 (52%)
What happened to failing respondents?	Reported only	5 (10%)
	Sensitivity analysis-kept	7 (14%)
	Sensitivity analysis- excluded	3 (6%)
	Excluded automatically	5 (10%)
	Unclear/not reported	2 (4%)
	NA- not included/ no failures	29 (56%)

Qualifying questions

Eighteen studies (35%) reported the inclusion of one or more questions used to qualify or validate responses. Qualifying questions included alternative preference elicitation methods (e.g. ranking or rating tasks, stated WTP), debriefing questions (e.g. self-reported task difficulty, choice certainty or satisfaction) and exercises to evaluate respondents' ability to engage with tasks (e.g. health literacy, numeracy, comprehension questions)

The utilisation and reporting of qualifying question results was limited. Five studies incorporated results into the econometric modelling of DCE responses, for example to understand scale heterogeneity or interactions with attributes (de Bekker-Grob *et al.*, 2020; Peters & Siersema, 2020; Ryan & Wordsworth, 2000; Salkeld *et al.*, 2003; Vass *et al.*, 2018a). Seven studies used results to aid the interpretation of DCE results (Boone *et al.*, 2013; Brenner *et al.*, 2014; Kistler *et al.*, 2015; Osborne *et al.*, 2018; Pignone *et al.*, 2012; Pignone *et al.*, 2014; Pignone *et al.*, 2013). For remaining studies it was unclear if/how results were utilised (Charvin *et al.*, 2020; de Bekker-Grob *et al.*, 2013b; Ghanouni *et al.*, 2014; Mansfield *et al.*, 2018; Miles *et al.*, 2019; van Dam *et al.*, 2010)

4.3.3.4 Study type

Figure 4.4 shows the link between the stated purpose (or research question), measurable outcomes and the application results for each study. The purpose of mapping the application of results was to determine if and how findings from studies are being used, given the high volume of published studies. The application of DCE results were mapped according to details about the broader context of the studies given by authors within the paper (particularly within the introduction and discussion sections) as well as tracking forward citations and ongoing work by authors. This is an imperfect method but gives a general idea of the application of results. A small number of studies had a clear practical application such as a HTA submission (Kitchener *et al.*, 2016), other studies were used to inform further research on preference elicitation by authors, but for the majority of studies, it was unclear if and how results had been used in any practical way.

Patterns in study design and the types of outcomes measured were distinguishable by study-type rather than according to cancer site. Studies broadly sought to address policy-centric or methodological research questions and could be grouped in to eleven common study-types.

Policy

The objectives of these studies directly relate to policy issues surrounding the design and implementation of current and future screening programmes and can be described by seven categories:

i. Increasing uptake for existing technologies (n=3)

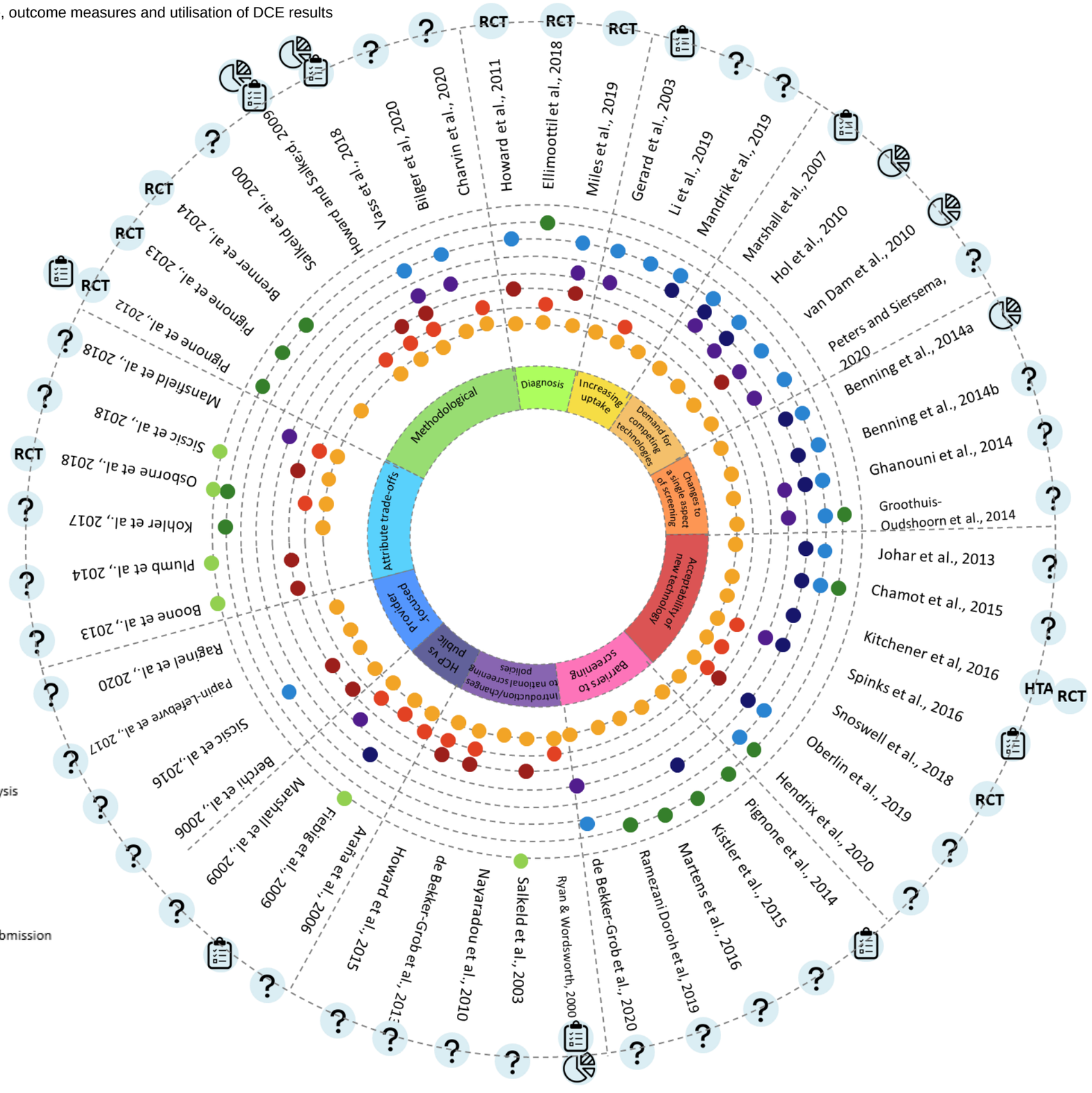
Three studies aimed to understand preferences to improve uptake of existing breast (Gerard *et al.*, 2003; Mandrik *et al.*, 2019) or cervical (Li *et al.*, 2019) screening programmes. Attributes primarily centred around experiential factors relating to screening whereas the inclusion of outcomes was limited to as single attribute in one study (Li *et al.*, 2019). The number of attributes within this group was also higher than average, with Gerard *et al.* (2003) and Mandrik *et al.* (2019) including ten attributes each. Alongside, standard regression coefficients, the primary outcome of these

studies was uptake rates and/or changes to uptake rates according to screening modality.

ii. Demand for competing technologies (n=4)

In some instances, several screening technologies exist and have all been consistently found to be cost-effective compared to no screening with no clear dominating modality (Ran *et al.*, 2019). However, tests differ greatly in terms of patient burden and potential harms and benefits. In response, four studies estimated the share of demand for multiple, competing screening modalities alongside overall uptake rates. Three studies focused on colorectal cancer (Hol *et al.*, 2010; Marshall *et al.*, 2007; van Dam *et al.*, 2010) whereas Peters and Siersema (2020) considered alternative screening tests for oesophageal cancer. Hol *et al.* (2010) used a specific, labelled design, designing test attributes in accordance with each of the screening modalities considered and providing extensive background information on each test. The remaining studies estimated demand for each technology by mapping predetermined attributes to each modality to provide a profile of test characteristics. Attributes heavily focused on differences between tests captured via test-specific characteristics and outcomes, giving very little attention to service delivery or costs. Common outcomes included demand for screening, elasticity of demand and demand for each available screening test. Results varied across tests; Hol *et al.* (2010) and Peters and Siersema (2020) found respondents valued test process most highly with less invasive tests being preferred, whereas Marshall *et al.* (2007) and van Dam *et al.* (2010) found outcomes such as effectiveness and accuracy to be most important, resulting in respondents preferring more invasive but accurate tests such as colonoscopy.

Figure 4.4 Visual summarisation of the purpose, outcome measures and utilisation of DCE results



iii. Changes to a single aspect of screening (n=4)

Four studies examined the importance of an improvement to a single aspect of colorectal tests, traditionally considered a barrier to screening. All DCEs estimated overall demand for screening and share of demand according to test modality. Ghanouni *et al.* (2013) examined preferences relating to hypothetical improvements to currently intense preparation procedures performed prior to undergoing a CT colonography, using a generic, unlabelled design. The remaining studies focused on the impact of reduced invasiveness of either newly-emerging screening tests (Benning *et al.*, 2014a; Groothuis-Oudshoorn *et al.*, 2014) or follow-up testing (Benning *et al.*, 2014b). These studies used a test-specific design with three profiles representing each screening technology compared within each question. Due to the hypothetical nature of the technologies examined and the specific research questions, attribute selection was less rigorous compared to other studies within this review- relying more on author's assumptions with no utilisation of qualitative research.

iv. Acceptability of new technology (n=7)

Seven studies examined how technological advances, particularly the introduction of new screening modalities affected demand (Chamot *et al.*, 2015; Hendrix *et al.*, 2020; Johar *et al.*, 2013; Kitchener *et al.*, 2016; Oberlin *et al.*, 2019; Snoswell *et al.*, 2018; Spinks *et al.*, 2016). Four of the studies (Chamot *et al.*, 2015; Johar *et al.*, 2013; Kitchener *et al.*, 2016; Oberlin *et al.*, 2019) focused on cervical screening programs. Alternative research aims included the impact of the potential introduction of teledermoscopy for melanoma screening (Snoswell *et al.*, 2018; Spinks *et al.*, 2016) and the use of AI in breast screening (Hendrix *et al.*, 2020). The primary aim of these studies was to estimate demand for a proposed or newly introduced test compared to existing test modality, meaning share of demand or elasticity of demand were key outcomes alongside standard estimates of attribute utility. Despite the emphasis on uptake and demand for screening, only two studies of this kind to estimate overall demand for screening (Oberlin *et al.*, 2019; Spinks *et al.*, 2016). The objectives of these studies focused on differences between tests, therefore all used a test-specific or mapped design. Kitchener *et al.* (2016) elicited preferences of women who previously opted out of screening. Remaining studies focused on retesting amongst

previously screened patients. Studies typically included a range test-specific and structural attributes and interestingly, did not consider outcomes at all.

v. Barriers to screening (n=5)

Five studies focused on preferences for screening in vulnerable or low uptake populations such as low income or uninsured (Martens *et al.*, 2016; Pignone *et al.*, 2014; Ramezani Doroh *et al.*, 2019), elderly (Kistler *et al.*, 2015) or previous non-attenders (de Bekker-Grob *et al.*, 2020) groups, where underlying needs may differ systematically and observed screening rates are traditionally lower than the national average. The aim of the studies was to understand what drives demand in under-screened population groups to improve screening adherence. All studies examined preferences surrounding colorectal cancer screening and typically examined preferences towards multiple, widely available screening modalities. All DCEs allowed for preference heterogeneity by utilising hierarchical Bayes (Kistler *et al.*, 2015; Martens *et al.*, 2016; Pignone *et al.*, 2014), latent class (de Bekker-Grob *et al.*, 2020) or segregation (Ramezani Doroh *et al.*, 2019) approaches to model estimation. Alongside average utility scores for each attribute level, a common outcome within these studies was attribute importance scores. Despite the emphasis on improving uptake rates, only de Bekker-Grob *et al.* (2020) attempted to estimate demand for screening by conducting sensitivity analysis examining the impact of test and personal characteristics on uptake.

vi. Introduction of and changes to national screening policies (n=5)

Five DCEs examined general public preferences with the purpose of designing national screening policies, specifically examining marginal rates of substitution between the benefits and harms of screening. The publication of these studies tended to coincide with the introduction of new mass screening programs (Salkeld *et al.*, 2003) or changes to existing policy (Nayaradou *et al.*, 2010; Ryan & Wordsworth, 2000) or the consideration of potential national screening programs with ambiguous efficacy (de Bekker-Grob *et al.*, 2013a; de Bekker-Grob *et al.*, 2013b; Howard *et al.*, 2015).

Studies did not attempt to draw conclusions on the most acceptable test; therefore, they used generic survey designs avoiding test-specific attributes and instead selected

levels based on the accepted screening modality. Selected attributes were wide-ranging; however, these studies included the highest number of attributes per study. This group also had a greater focus on the associated risks of screening than other DCEs, with each study including at least two attributes incorporating risk. Despite the emphasis on the importance of participation, uptake was not estimated within these studies. Instead, the primary outcome was trade-offs between attributes using WTP or WTA calculations. Results varied across studies but outcome attributes, particularly reduction in cancer-related mortality, were consistently ranked more important than structural attributes.

vii. HCPs vs Public (n=3)

Three papers compared the preferences of doctors and the public to evaluate discrepancies that may impact screening behaviours (Arana *et al.*, 2006; Fiebig *et al.*, 2009; Marshall *et al.*, 2009). Arana *et al.* (2006) and Marshall *et al.* (2009) both used a single questionnaire across population groups but differed slightly in research questions: Arana *et al.* (2006) asked doctors about their own screening behaviour to assess the impact of medical experience whereas Marshall *et al.* (2009) required HCPs to answer of behalf of their patients (i.e. predict patient preferences). Alternatively, Fiebig *et al.* (2009) asked HCPs about their own preferences for patient care using different surveys for the different populations. The use of different surveys limited comparability across attributes but was more representative of real practice where doctors consider issues such as patient age, screening history and remuneration when making screening recommendations. These studies generally found that the order of importance of attributes was consistent between doctors and their patients; however, the relative importance consistently differed across studies. Perhaps due to information and experience asymmetry, responses of HCPs tended to be less variable than the general public (Fiebig *et al.*, 2009) and predictions of uptake and willingness to pay were much more reflective of real practice compared to patient responses (Marshall *et al.*, 2009).

viii. Provider-focused (n=4)

Four studies focused specifically on the preferences of HCPs towards mass screening programmes, primarily colorectal cancer (Berchi *et al.*, 2006; Papin-Lefebvre *et al.*,

2017; Raginel *et al.*, 2020; Sicsic *et al.*, 2016). All four studies were set in France and typically coincided with changes in national screening policies such as the initial introduction of colorectal screening in 2008 or change in the standard screening modality from FOBT to FIT in 2015 (Leuraud *et al.*, 2013; Pellat *et al.*, 2018). DCEs aimed to understand how idiosyncrasies of the French healthcare system may hinder promotion of screening by doctors directly impacting participation rates for their patients. For example, Berchi *et al.* (2006) and Sicsic *et al.* (2016) examined the optimal level of remuneration to increase willingness to engage with screening programs under the fee-for-service system that pays for curative medical treatments but can hinder screening uptake if preventative services are not compensated. Alternatively, Papin-Lefebvre *et al.* (2017) focused on preferences relating to legal responsibility for incorrect results or late detection of cancer. Finally, Raginel *et al.* (2020) focused on the promotion of cervical screening in the context of social inequalities. Studies typically included attributes capturing structural characteristics of screening programs rather than outcomes or features of specific tests and only considered screening modalities approved for mass screening.

ix. Attribute trade-offs (n=6)

The objective of these studies was to estimate the trade-offs between attributes of screening tests (Boone *et al.*, 2013; Kohler *et al.*, 2017; Mansfield *et al.*, 2018; Osborne *et al.*, 2018; Plumb *et al.*, 2014; Sicsic *et al.*, 2018). Osborne *et al.* (2018) and Mansfield *et al.* (2018) specified levels according to existing screening modalities. All remaining studies applied a generic design using non-specific attributes and levels that could be applied across current and future modalities. The aim was not to predict screening behaviour or preferences for competing tests therefore, with the exception of Sicsic *et al.* (2018), studies did not include an opt-out option or attempt to estimate uptake. Instead, common outcomes included the MRS between attributes using WTP or WTA methods with false-positive results as the numerator. Alternatively, Kohler *et al.* (2017) assessed the relative importance of attributes by calculating mean attribute importance scores.

x. Diagnosis studies (n=3)

Three studies focused on diagnostic testing for cancer (Ellimoottil *et al.*, 2018; Howard *et al.*, 2011; Miles *et al.*, 2019). All studies elicited preferences of patients towards existing tests compared to recent or emerging innovations in diagnostic imaging. All were embedded within a wider clinical trial comparing technologies. Ellimoottil *et al.* (2018) was the only study to provide an opt-out alternative. Remaining studies justified the omission on the basis that participants had previously undergone the diagnostics under investigation within the DCE and were unlikely to decline testing when experiencing cancer symptoms. All studies measured trade-offs with WTA (Howard *et al.*, 2011; Miles *et al.*, 2019) or WTP (Ellimoottil *et al.*, 2018). Additional outcomes included the probability of choosing the existing test or new test under different scenarios according to changes to attribute levels or personal characteristics (Howard *et al.*, 2011; Miles *et al.*, 2019).

Methodological studies

A second group of studies used cancer testing as a case study to demonstrate underlying practical issues relating to the design and implementation of DCEs within health economics. Research questions focused on specific methodological issues and results primarily focused on implications and recommendations for future practice. Interpreting how results may impact cancer testing was less common across these studies. Within the methodological studies three broad categories were identified:

i. Comparison of preference elicitation techniques (n=3)

Three studies used cancer screening to compare the outcomes of DCEs with alternative conjoint analysis methods including ranking and rating tasks (Brenner *et al.*, 2014; Pignone *et al.*, 2012; Pignone *et al.*, 2013). This group focused the least on consequences for cancer screening policies; underreporting on issues such as attribute selection and experimental design, and inconsistently including results from DCE tasks.

ii. Underlying theoretical issues (n=5)

Five studies focused on issues relating to the theories underpinning DCEs using cancer screening as an exemplar (Bilger *et al.*, 2020; Charvin *et al.*, 2020; Howard & Salkeld, 2009; Salkeld *et al.*, 2000; Vass *et al.*, 2018a). Salkeld *et al.* (2000) used the case of colorectal cancer screening to examine whether people prefer a service they have previously experienced compared to hypothetical new policies- a question which incorporates several economic and psychological principles within decision theory such as status quo bias, hypothetical bias and regret aversion. Results confirmed the existence of a 'veil of experience' but authors did not interpret results in terms of consequences for screening. Howard and Salkeld (2009) used the example of colorectal cancer to assess the influence of attribute framing on preferences and specifically WTP estimates. Results demonstrated significant differences in model estimates based on negative and positive framing of the sensitivity and specificity of competing tests. Due to comprehensive descriptions of the development process, experimental design and outcomes, this study provides more insights into the implications for colorectal screening than other methodological papers. Vass *et al.* (2018a) investigated the impact of presentation of risky attributes on preferences in the context of breast screening, finding no differences in model estimates when risk was presented visually or numerically. Finally, two studies investigated the impact of increased information on decision processes relating to prostate screening (Bilger *et al.*, 2020; Charvin *et al.*, 2020) with contrasting results. Charvin *et al.* (2020) found increased information on the benefits and harms of screening significantly increased the intention to abstain screening given the current ineffectiveness of test strategies. Alternatively, findings from Bilger *et al.* (2020) suggested differences in the information relating to breast and cervical screening had no impact on uptake or preferences.

4.3.4 Discussion

4.3.4.1 Main findings

This review demonstrates that the number of DCEs for cancer testing continues to grow. Almost all studies within this review focused on preferences towards cancer screening. The publication of DCEs eliciting preferences towards diagnostic testing is comparatively limited; however, the three studies included within this review

demonstrate a willingness to engage with choice tasks and form preferences in relation to this topic. DCEs relating to testing for colorectal cancer were particularly popular and represented 52% (27/52) of studies. Oppositely, preferences relating to ovarian cancer is a clearly underexplored area, with no studies identified.

Studies covered a broad range of research questions, particularly focusing on policy-based issues such as barriers to uptake of existing screening programmes and acceptability of newly developed or emerging screening technologies. Preference-based evidence emerging from the DCEs provides valuable insights given the current status of screening in the UK. The effectiveness of screening programmes in maximising health gains and reducing mortality relies on the willingness of individuals to participate. In recent years screening rates have become stagnant, inconsistently meeting participation targets (Hirst *et al.*, 2018; Shahidi & Cheung, 2016) and even declining in some instances (NHS Digital, 2018). Preference-based studies mean test procedures can be refined and barriers or misconceptions of screening can be identified and addressed through public health campaigns to increase uptake and improve patient satisfaction. Simultaneously public pressure for extending cancer screening to other tumour sites, particularly prostate cancer is enduring despite evidence that any benefits are outweighed by harms at present. In this instance, understanding preferences provides understanding of minimum thresholds new tests must fulfil to be acceptable to the public.

Despite the emphasis on policy-based research and continued growth in published literature relating to cancer testing, it was often unclear how the results of DCEs can and have been applied in practice. The conclusions of studies frequently recommend the most appropriate screening modality or approaches to increase participation; however due to the relative nature of results, inconsistencies in research practice/quality and heterogeneity in attribute inclusion, these conclusions frequently differ from study to study and are unable to be effectively pooled. This inability to externally validate or even effectively compare results across DCEs impacts the reliability of results and likely limits the practical application of findings.

Amongst the more pragmatic studies, Kitchener *et al.* (2016) developed a DCE to assess the acceptability of self-testing cervical screening kits and timed appointments among women who did not initially respond to a routine screening invitation. The DCE was submitted as part of HTA application although it is unclear what weighting this held for any decision-making as both interventions were found to be effective at increasing screening uptake and cost-effective at £7,593 and £8,438 per QALY based on traditional cost-utility analysis using trial-based data. Additionally, several other studies including all three diagnostic studies were set in the broader context of a randomised trial that aimed to measure the feasibility and acceptability of new testing modalities (Ellimoottil *et al.*, 2018; Howard *et al.*, 2011; Miles *et al.*, 2019; Osborne *et al.*, 2018; Snoswell *et al.*, 2018).

However, for most studies (32/40, 62%), it was unclear if and how results had been utilised to inform clinical practice or policy making decisions. The most common observable purposes were for the design of a subsequent DCE (9/40) or secondary analysis of the data with the purpose of exploring methodological issues associated with DCEs (6/10).

Areas of limitation

To reflect the growth in the use of DCEs in healthcare in recent years, several formative guidelines have been published to offer methodological advice for conducting and reporting DCEs covering topics such as survey development, experimental design, model specification and interpretation of results (Bridges *et al.*, 2011; Lancsar & Louviere, 2008; Coast *et al.* (2012). This section highlights a few examples of mismatches between “best practice” and methods used within the DCE included in this review. Although it should be noted that fourteen (27%) of the studies were published prior to the publication of the checklist used to assess the quality of studies in this study (Bridges *et al.*, 2011). It is also likely that several of the studies published in subsequent years were also designed and conducted prior to publication or dissemination and widescale adoption of the guidance as best practice. Encouragingly, results of the critical appraisal suggest the quality of studies continues to improve over time as dissemination of guidance documents increases. Additionally,

in many instances it is difficult to distinguish between inadequate reporting and suboptimal research design.

Guidance acknowledges that attribute selection requires a balance between the factors relevant to respondents and the specific policy or research question of interest (Bridges *et al.*, 2011). In general, descriptions of attribute development were brief, making the ability to assess the rigorousness of the selection process difficult and limiting replicability in future research. This was particularly true for studies which included qualitative elements, with descriptions of how evidence was collected and utilised underreported in most instances, an issue that has been discussed in depth by Coast *et al.* (2012). The perceived importance of attributes is entirely dependent on the other attributes included (or excluded) within the DCE. This means attribute selection must strike a balance between avoiding omitted variable bias by excluding key attributes and increasing complexity of choice tasks due to too additional attributes (Bridges *et al.*, 2011). To limit omitted variable bias, the importance of consulting the target population during the attribute selection stage is widely acknowledged. However, only 55% (29/52) of the studies within this review directly incorporated the views of the target population during the attribute selection process.

To further reduce the risk of excluding potentially relevant attributes, guidelines stress the importance of an iterative face-to-face piloting (Lancsar & Louviere, 2008). Lancsar and Louviere (2008) recommend pretesting to check the choice of attributes during attribute selection and understanding, complexity and timing of the final questionnaire to identify common areas of misunderstanding or non-compliance early in the development process to ensure validity of the final questionnaire. Although three quarters (39/52) of studies reported conducting some form of piloting, descriptions of purpose, methods and outcomes of piloting were under-reported.

The level of detail relating to the reporting of the more quantitative elements of DCEs (i.e. sample size calculations, experimental design and econometric analysis) varied considerably across studies making judgements of research quality difficult. There remains no gold standard for these aspects of DCEs, meaning it is important for researchers to describe, evaluate, and document how the particular design meets the

goals of the study (Bridges *et al.*, 2011). In general, it appears that the experimental design and econometric modelling of responses continue to become increasingly sophisticated to maximise the information that can be obtained from choice experiments and incorporate more complex issues such as preference or scale heterogeneity. However, increased complexity does not appear to be matched with in-depth descriptions of the underlying methods and features, which were often insufficient according to the leading guidance available. Shortcomings in reporting may be due to word count restrictions of journals but also appear to be exacerbated by the increasing use of software packages such as Sawtooth where advanced methods are made more accessible but lead to a potential 'black box' effect.

4.3.4.2 Ongoing issues/ areas of debate

This review includes examples of high-quality DCE studies applied to the elicitation of preferences for cancer screening and diagnostic testing. However, previous sections highlighted that DCEs generally appear to be somewhat self-referential in nature (i.e. largely used to inform other DCEs). To be useful to decision-making bodies, DCEs must provide reliable, reproducible, comparable and generalisable results to the relevant population and/or healthcare system of interest. In a commentary on the use of healthcare-focused DCEs to inform quantitative benefit-risk assessment in regulatory decision-making, Vass and Payne (2017) highlight a number of key areas where more supportive evidence on the readiness of DCE evidence is required, including the role of training materials and defining the appropriate study population.

The high volume of methodological papers included within this review indicate steps to consolidate many unresolved debates in the best practice in the application of DCEs to health-related questions. The results of this current review identify several additional areas of methodological debate where additional evidence may improve the uptake and application of DCE findings in policy decisions. Many of these areas have been long debated in the broader DCE literature but investigation within the healthcare setting remains largely unaddressed.

Validation of DCE results

Maximising the validity of DCE responses is a fundamental requirement for the incorporation into decision-making contexts. Validity checks, both internal and external, provide a way of assessing the quality and accuracy of the experimental data collected in DCEs regarding the ability to capture the true preferences of respondents and applicability to real world decision-making. External validity requires observation the actions respondents in a real world setting and is therefore notoriously hard to measure since an inability to observe true behaviour is typically the motivation for the application of stated preference methods such as DCEs in the first place (Lancsar & Swait, 2014). In the absence of, or complementary to external validity measures, internal validity tests provide an accessible way of checking the logic, consistency and underlying assumptions of choice behaviour within discrete choice experiments.

Twenty-five studies in this review included at least one additional choice set to test the rationality/validity of respondent's choices. Qualifying or debriefing questions provide an alternative method to evaluating the quality and validity of results. Sixteen (31%) studies reported the use of debriefing questions. The interpretation of both validity check and debriefing questions varied greatly across studies and in many instances it was unclear how/if responses were utilised. Currently there is no widely accepted or standardised set of questions which are recommended to assess the quality of responses and many questions regarding the application of rationality check and debriefing questions for the purpose of assessing the validity of findings remain (Pearce *et al.*, 2020). For instance, how many and which questions to include? What is level of failures is acceptable at both an individual-level and sample-level? What should be done about respondents who appear to 'fail' rationality checks? How does the inclusion of validity check questions impact responses to choice tasks?

The presence of indifference

Guidance on the use of an opt-out option is mixed; Lanscar and Louviere (2008) discourage the use of forced-choice, particularly when considering interventions that rely on voluntary participation (e.g. screening), as results can only be interpreted under the implicit assumption that all respondents choose to participate even when both options are unappealing and would not be chosen in practice - a particularly relevant

issue when considering uptake. Bridges *et al.* (2011) however, acknowledge that the appropriateness of an opt-out is entirely dependent on the research objectives and is not necessary and often not suitable when estimating underlying preference structures due to the censoring of data (since the selection of the opt-out alternative provides no information on trade-offs between attributes).

Despite these differences in guidance, the use (or disuse) of opt-out alternatives was generally well-justified within the studies within this review. Furthermore, there is a growing evidence-base surrounding unresolved debates such as the consequences of including or excluding no choice alternatives (Determann *et al.*, 2019; Veldwijk *et al.*, 2014) and guidance on how such responses should be incorporated into the analysis of responses (Campbell & Erdem, 2019; Ryan & Skåtun, 2004).

Conversely, despite being explicitly listed as a criterion in the ISPOR Conjoint Analysis guidelines (Bridges *et al.*, 2011), the inclusion of "indifference" or "no preference" alternatives has been less considered within the healthcare literature to date. An indifference alternative allows respondents to express neutrality between test profiles within each choice task. Absence of an indifference alternative may force respondents to artificially assign a preference to an alternative which could potentially lead to stochastic decision-making, reducing the overall quality and reliability of data (Cantillo *et al.*, 2010) Despite this, just one study in this review included a "no preference" alternative, and it is unclear how such responses were analysed (Pignone *et al.*, 2012). A key obstacle in the inclusion of indifference alternative is the current lack of evidence on how such responses should be modelled during econometric analysis. Previous studies investigating indifference in the broader DCE literature have used a variety of methods no leading/ clear analytical approach has emerged (Cantillo *et al.*, 2010; Dekker, 2014) meaning studies remain largely theoretical in nature and use in applied studies remains uncommon.

Attribute non-attendance

Attribute nonattendance (ANA) is widely acknowledged to have detrimental impact of bias in parameter estimates and can be particularly misleading when attempting to

make comparisons on the marginal rates of substitution between attributes (e.g. WTP measures). Under standard theories of behavioural choice (such as Random Utility Theory), discrete choice models assume all attributes are considered by individuals and are traded off in a compensatory manner. Non-attendance presents a challenge to this assumption whereby one or more attributes within a choice alternative is ignored by a subset of respondents (i.e. it has no influence on the choice made).

ANA can be measured using a variety of approaches; self-reported non-attendance through debriefing questions, observed through methods such as eye or mouse tracking or inferred through econometric analysis methods.

In total, six (12%) studies included a check for non-attendance, sometimes referred to as 'non-trading'. However, no studies attempted to adjust model estimates to account for these respondents beyond excluding them from analysis. Attempts to measure and address attribute non-attendance (ANA) is an area of neglect within the studies within this review. Furthermore, despite gathering large amounts of attention in disciplines such as transport economics debates remain about how to best measure attribute non-attendance and how to analyse non-trading responses (Alemu *et al.*, 2013; Heidenreich *et al.*, 2018; Lew & Whitehead, 2020).

Risk communication

The balance of risks and benefits are central to decision-making relating to testing as a result risk-based attributes feature heavily in the studies within this review (Appendix 4.1). The incorporation of risk increases the cognitive burden of choices, particularly when applied to already complex medical concepts such as test sensitivity and specificity. If risk information is not well presented or well understood, it may be detrimental to the validity of results. A few studies within the review focused on how variations risk presentation may affect results (Howard & Salkeld, 2009; Vass & Payne, 2017). Despite the growth in methodological papers addressing challenges around risk communication challenges still remain.

Challenges specifically relating to risk communication within DCEs have also been highlighted in (Harrison *et al.*, 2014) and include framing effects, optimal visual aids and relative versus absolute risks. Although, there is currently no gold standard method for risk communication, outside of the DCE literature a number of emerging guidelines and best practices continue to evolve (Fagerlin *et al.*, 2011; Naik *et al.*, 2012). Much of this guidance has not been applied within healthcare DCEs to date. Improved understanding of risk communication applied to risky attributes in necessary to limit bias and demonstrate the robustness of results of DCEs.

4.3.4.3 Limitations with review

This review was conducted using a rigorous search strategy, performed and reported with full transparency. However, a few limitations remain. First, the search was limited to published peer-reviewed studies meaning some unpublished studies that add to the knowledge base may have been excluded. Second, analysis was limited to what was reported by authors of the original studies; where supporting documents such as the survey instrument and further contextual information were unavailable this may have led to an unduly critical assessment of a study. Finally, quality assessment was performed using a methodological checklist because of the unavailability of an established risk-of-bias assessment tool to evaluate discrete choice studies (Bridges *et al.*, 2011). This enabled discussions of the quality and limitations of included studies in a broad sense, specifically in relation to the quality of reporting within DCEs, but the methodological quality cannot be fully inferred meaning it was not possible to calculate a measure of the risk of bias or make decisions on the inclusion/exclusion of studies based on quality.

4.4 Chapter summary

This chapter summarised and critically appraised existing evidence relating to the application of DCEs to testing for cancer in both asymptomatic (screening) and asymptomatic (diagnosis) settings. The chapter began by performing an overview of existing systematic reviews on the topic. In total, five systematic reviews were identified, all of which focused on cancer screening. The overview identified several gaps in the existing literature such as evidence relating to methodological aspects of DCEs.

Next, a systematic review of DCEs eliciting preferences towards cancer screening and diagnostic testing was conducted. The systematic review built on existing reviews by expanding the search terms, dates and databases.

Studies relating to cancer screening were common and appeared increasingly popular over time. The exploration of preferences towards diagnostic testing were comparatively unexplored, with only three studies identified. Studies typically focused on policy-based questions such as increasing uptake, assessing the acceptability of potential screening programmes and improving access for underserved populations. Despite the continued growth in the application of DCEs to cancer testing, the application of findings to policymaking appears limited at present.

The inability to externally validate DCE findings means refining DCE methods to ensure quality and minimise bias in estimates is crucial to improving the utilisation of results by decision-making context. This chapter highlighted several areas where more evidence may help to improve the methodological quality and consistency of future DCEs applied to cancer testing. These areas include the use of alternatives to allow individuals to express indifference between test options, clearer reporting of methods and results within studies and refining the communication of risk-based attributes. Findings from this chapter were used to refine the methodological research questions addressed within the DCEs described in future chapters. Furthermore, in-depth analysis of attribute selection, importance and significance was a fundamental step for attribute selection for the DCEs within this thesis (Appendix 4.1).

Literature update: September 2022

Searches were rerun on 10th September 2022 to account for the time between completion of the systematic review and submission of the thesis. In total, 162 new publications were identified to be screened. No new systematic review fitting the criteria of the scoping review were identified, however, four newly published primary DCE studies were identified. All studies related to cancer screening; no additional studies eliciting preferences towards diagnostic testing for cancer were identified. Similar to the majority of studies included as part of the full systematic review, new studies all focused on policy-based questions, in particular the acceptability of potential new programmes such as endoscopic screening for gastrointestinal cancer (Liu *et al.*, 2022) , lung cancer (Zhao *et al.*, 2021) and multi-cancer screening using a blood test (Gelhorn *et al.*, 2022) . Alternatively, (colorectal) aimed to compare preferences towards an existing colorectal screening between HCPs and patients (Heidenreich *et al.*, 2022).

Three of the four studies included a combination of service delivery, test-specific, cost and outcome characteristics. For (multi test and colorectal) test accuracy was the most important attribute overall, while respondents in (endoscopy) prioritised the pain associated with a test above all else. Interestingly, Zhao *et al.* (2021) focused only on service delivery attributes with screening interval being the most important determinant of choice of lung screening test overall.

Data for these studies were all collected during the Covid-19 pandemic, however no studies discussed the potential impacts on the pandemic on preferences. It is likely that preferences for screening may have changed since the pandemic due to a refreshed focus on physical health and new exposure to the intricacies of screening and diagnostic testing as discussed in Chapter 3. Future DCEs in this area may be useful to explore changes in preferences and attitudes towards screening. For instance, by repeating pre-pandemic experiments in matched populations.

5 Preferences towards diagnostic testing: attribute development

5.1 Introduction

This chapter describes the process of selecting attributes and levels to be included in a DCE aiming to explore preferences towards diagnostic testing for ovarian cancer in primary care. Attribute development in this chapter utilised a multi-method approach by combining a range of methods, including literature reviews, quantitative questionnaires, and qualitative workshops. The chapter takes a sequential approach describing each of the methods and outcomes in turn before summarising all the results at the end and describing the final outcomes. Next, a summary section describes the process of consolidating evidence from the different stages of attribute selection and provides a description of the selected attributes and level assignment. The chapter finishes by discussing challenges relating to attribute development, in particular, the complexities of combining multiple methods during attribute selection.

5.2 Background

Attribute selection is an early step in the development of any DCE. During a DCE, a combined set of attributes is used to describe the alternative choices available to respondents. To generate high quality data, alternatives must reflect the choice context as realistically as possible to respondent (Bridges *et al.*, 2011).

To manage the complexity and cognitive burden of decisions, DCEs typically limit the number of attributes to between four and nine (Soekhai *et al.*, 2019). However, decisions relating to medical testing are complex and multifaceted. It is therefore important to identify and include the attributes that are most relevant to decisions for the maximum number respondents (Lancsar & Louviere, 2008). Respondents may make their own assumptions about any motivating factors that are omitted from the DCE task, reducing the quality and validity of results.

Methods for developing attributes within DCEs are varied. Current guidance places emphasis on the importance of target population involvement using qualitative

methods such as focus groups or interviews (Bridges *et al.*, 2011; Kløjgaard *et al.*, 2012).

This chapter builds on the increasing number of published studies by adopting a multi-method approach to attribute development (De Brún *et al.*, 2018; Helter & Boehler, 2016; Obadha *et al.*, 2019; Webb *et al.*, 2021), A combination of quantitative and qualitative methods were utilised and combined to select attributes in the context of the primary research question: what are women's preferences towards diagnostic testing for ovarian cancer? Figure 5.1 summarise the methods used during attribute aims and purpose of each of the methods utilised during attribute development. The remainder of this chapter describes each stage of the development process chronologically. Attributes and levels were then further refined during the pilot study described in Chapter 6.

A summary of attribute inclusions and exclusions at each stage of the development are shown in Figure 5.2.

Figure 5.1: Overview of the attribute development process

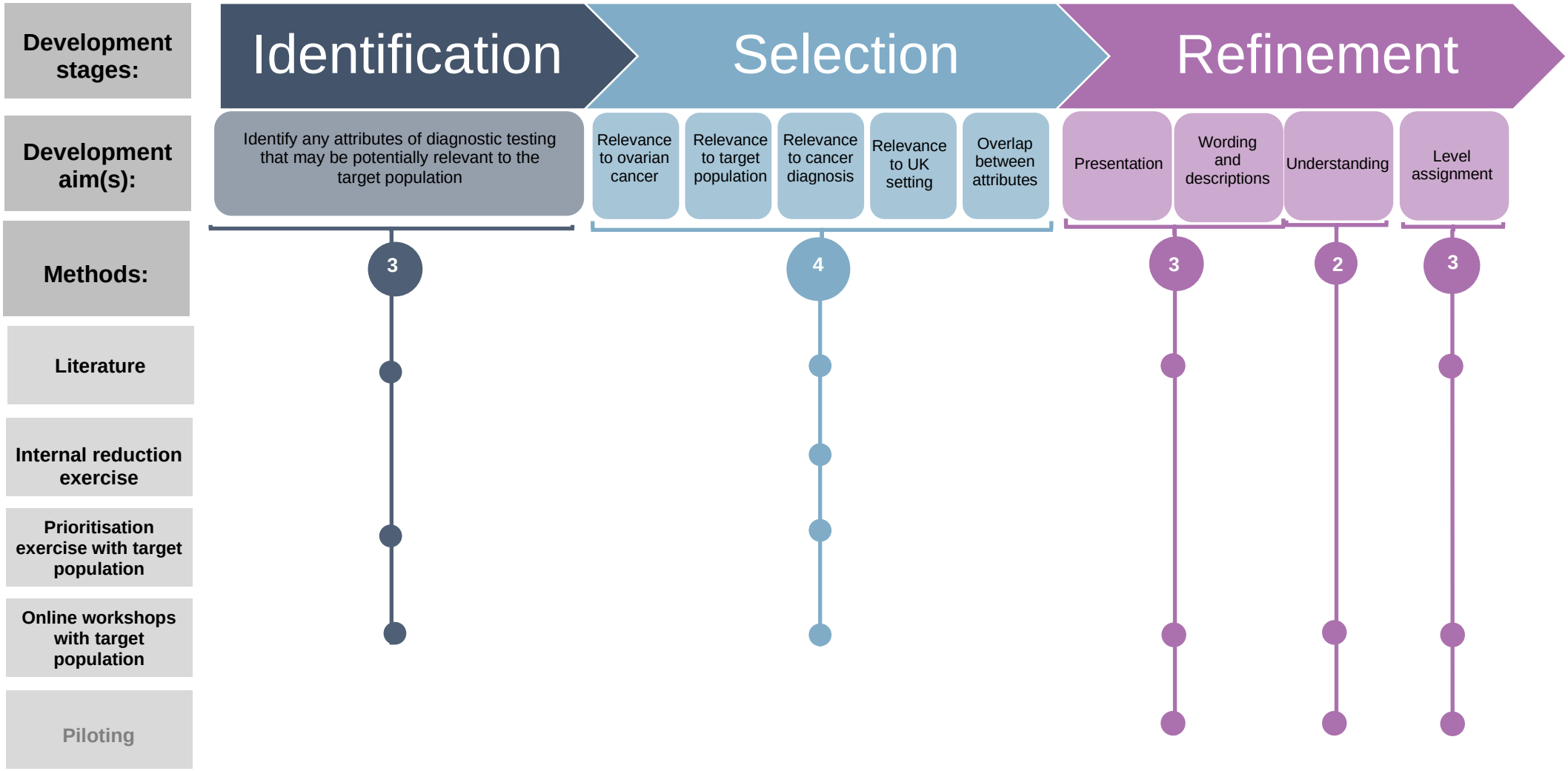
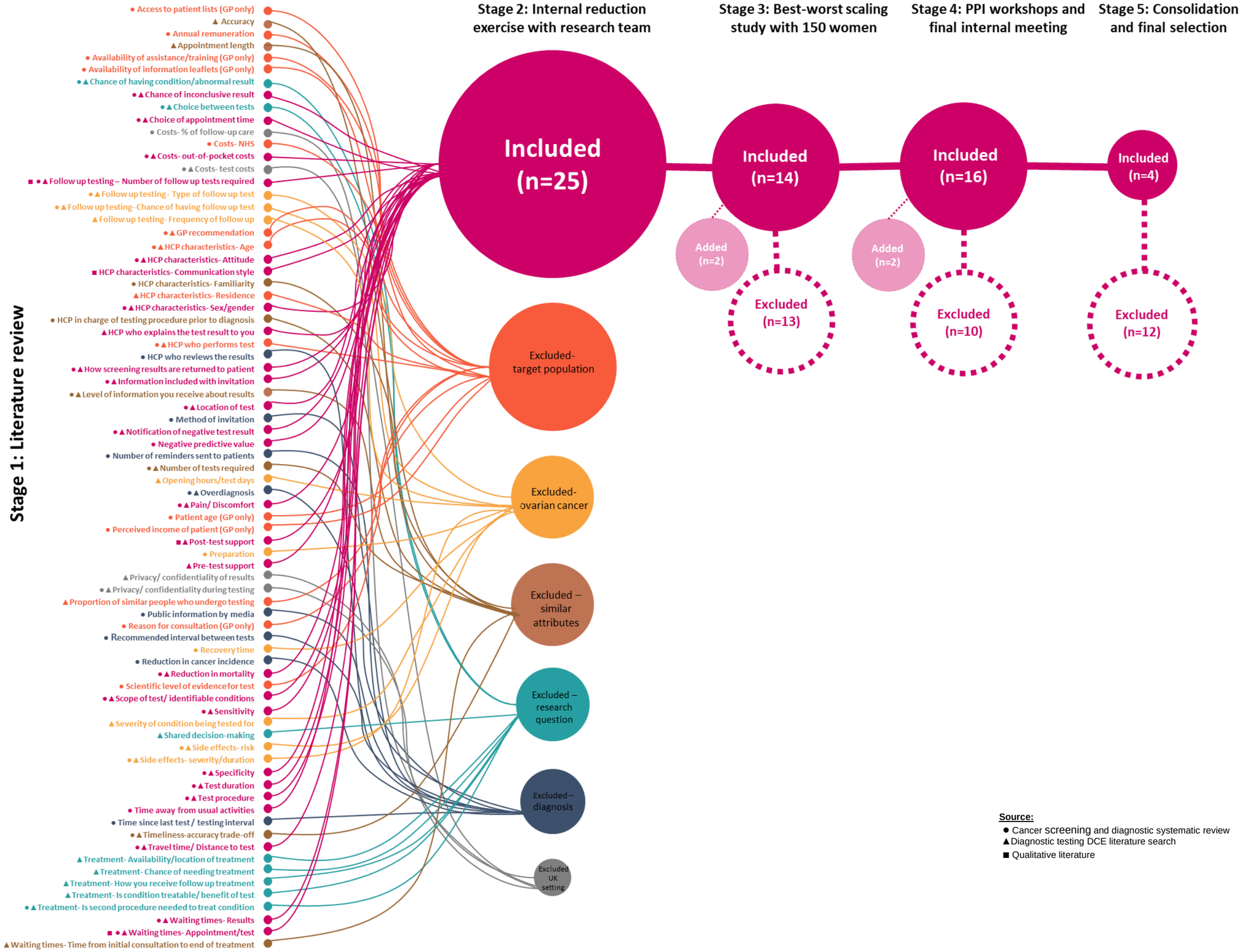


Figure 5.2: Summary of inclusions and exclusions during each stage of attribute development



Source:

● Cancer screening and diagnostic systematic review
▲Diagnostic testing DCE literature search
■ Qualitative literature

5.3 Literature reviews

5.3.1 Methods

The primary source of attribute identification was the systematic review of previous cancer screening and diagnostic DCEs presented in Chapter 4. To compensate for the heavy emphasis on cancer screening within the review, a complimentary literature search of discrete choice experiments towards diagnostic testing for any medical condition was conducted and any additional attributes were extracted. The literature search utilised the existing Endnote database used during Chapter 4 and applied search terms relating to diagnosis ("diagnose", "diagnostic", "diagnosis" and "test") to identify relevant studies. Searches took place in March 2019.

The final source of attribute identification was qualitative studies relating to the experiences of women diagnosed with ovarian cancer, discussed in Chapter 3 (Fitch *et al.*, 2002; Jelicic *et al.*, 2019).

Existing literature was also used to identify initial wording and descriptions of potential attributes to be refined throughout the development process.

5.3.2 Results

In total, 50 additional DCEs on diagnostic testing were identified. Studies covered a range of disease areas, most commonly; antenatal testing (n=16), genetic testing (n=16) and testing relating to sexual health (n=13). Analysis of included attributes was conducted and results are found in Appendix 5.1.

The attributes identified from the literature are shown in Figure 5.2. Seventy-seven attributes were identified in total, 17 of which were exclusive to the additional search of diagnostic DCEs and qualitative literature.

Summary of findings from the literature:

- Seventy-seven attributes were identified from multiple literature searches covering DCEs on cancer screening and diagnostic testing, DCEs of diagnostic testing in healthcare and qualitative studies on experiences of women diagnosed with ovarian cancer.

5.4 Internal reduction exercise

5.4.1 Methods

To condense the extensive list of potential attributes identified from the literature, the first step was to carry out an internal reduction exercise. This consisted of three stages:

i. Each member of the PhD team independently evaluated each attribute using a template in excel. This involved stating whether they believed attributes should be included or excluded according to a set of pre-specified criteria
ii. Reponses were merged and areas of agreement or disagreement were highlighted using a colour coded spreadsheet
iii. Results were discussed as a group in multiple rounds until consensus was reached

To minimise speculative judgements, attributes were excluded according to the specific criteria listed below:

- Relevance to the research question
- Relevance to tests available for ovarian cancer
- Relevance to diagnosis (given most attributes were identified from cancer screening literature)
- Relevance to a UK setting
- Relevance to population
- Overlap between attributes

Exclusions were made on the full agreement of the team; attributes were retained where a consensus could not be reached. The internal reduction process was not intended to be final. Attributes could be reintroduced at later stage if deemed important after engagement with the target audience.

5.4.2 Results

After two rounds of responses and discussion, fifty-two attributes were excluded upon the full agreement of the research team. Reasons for exclusion at this stage are indicated in Figure 5.2 by colour. Following this stage, twenty-five attributes remained.

Summary of findings from the internal reduction exercise:

- Twenty-five attributes remained following the internal attribute reduction exercise
- Fifty-two attributes were excluded based on five pre-determined exclusion criteria

5.5.1 Methods

Object case best-worst scaling (BWS) was used to determine the prioritisation of the remaining twenty-five attributes to the target population following internal reduction exercise. Included attributes and definitions used in the BWS survey are shown in Table 5.1.

In object case (Case 1) BWS, items are defined at the "object" level (e.g policies, products, medical treatments or attributes of policies, products or medical treatments) but the "levels" or range of each item is not specified to respondents (Figure 5.3)(Louviere *et al.*, 2013). The object case is the broadest approach to BWS and is often used as an alternative to ranking or rating tasks using a Likert scale. This approach is useful where there are many items of interest and the aim is to gain an understanding of the overall order of importance. Object case BWS provides a full ranking of attributes, however, due to the simplicity of the approach outcomes are somewhat limited compared to the other BWS approaches and DCEs (Mühlbacher *et al.*, 2016).

Figure 5.3: Example of a BWS object case (case 1) choice task

Most important	Characteristics of the test	Least important
	Mode of administration	
X	Risk of side effects	
	Test duration	X
	Out-of-pocket costs	

5.5.1.1 Experimental design

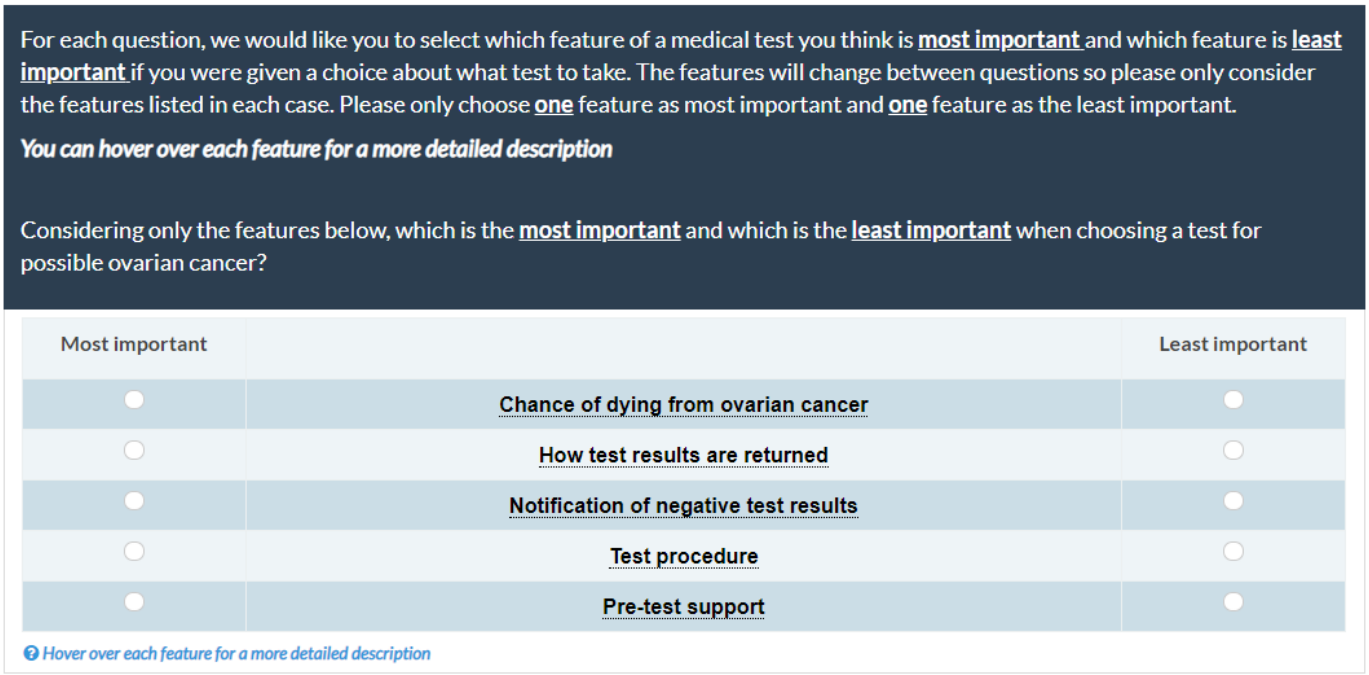
A balanced incomplete block design (BIBD) for the remaining attributes (n=25) was constructed using SAS 9.4. A BIBD is used when it is not possible to include all attributes in every block (choice task), instead a subset of the attributes of interest are shown in each block (i.e. an incomplete block design). To be balanced, the design was constructed such that each attribute occurs an equal number of times and attributes occur together an equal number of times across choice tasks.

Eight potential designs were identified ranging from 4-16 attributes per choice set and between 25-50 choice sets in total. A final design consisting of 30 choice tasks with 5 attributes each was selected, based on a trade-off between the complexity of the survey for respondents and the efficiency of the design (final d-efficiency of 83.3%). Each attribute occurred six times across the choice tasks and every pair of attributes occurs together once (i.e. concurrence of one). Comparing attributes with each other more frequently would increase the internal validity of the survey but at the expense of increased length and repetitiveness. The position of the attributes within choice tasks was optimised such that attributes were listed in 1^{st}-5^{th} position as equally as possible and the order of choice tasks was randomised between participants. Each participant completed all thirty choice tasks. An example choice task is shown in Figure 5.4.

Table 5.1: Summary of the twenty-five attributes included in the BWS study

	Attribute and definition
1	**Test sensitivity** *Chance that the test will miss cancer in a patient who actually does have the disease*
2	**Chance of dying from ovarian cancer** *How much having the test decreases the chance of dying from ovarian cancer*
3	**Choice of appointment time** *Whether a person can choose an appointment time or if the appointment time is assigned by the healthcare provider*
4	**Who explains the results** *Type of healthcare provider who explains the test results e.g. Nurse, doctor etc.*
5	**Pain and discomfort** *The level of pain and/or discomfort experienced during and after the test*
6	**Notification of negative test results** *Whether you are contacted if your results are normal*
7	**Chance of diagnosing another condition** *If symptoms are not caused by cancer, the chance the test can identify other reason for the symptoms*
8	**Pre-test support** *Level of support received before having the test describing what will happen during the test and what the results might show*
9	**Test procedure** *What having the test will involve. For ovarian cancer this could be a blood test or an transvaginal ultrasound (internal ultrasound of the reproductive organs)*
10	**Staff attitude** *How the healthcare provider treats you while conducting the test*
11	**Post-test support** *Level of support received after getting the results of the test relating to the meaning of your results and what will happen next*
12	**Time away from usual activities** *The total amount of time spent having the test instead of doing your usual daily activities*
13	**Test specificity** *Chance of unnecessary further invasive testing*
14	**Travel time** *The total amount of time spent travelling to and from the test*
15	**Time to notification of results** *The length of time it takes to hear the results after having the test*
16	**Openness of healthcare providers** *How open healthcare providers are with their thoughts about the cause of your symptoms and the tests they recommend*
17	**Number of follow up tests** *How many additional tests are needed to confirm a diagnosis*
18	**Chance of an inconclusive result** *The chance the results are unclear and the test would need to be repeated after a waiting period*
19	**Out of pocket costs** *How much it will personally cost a person to have the test e.g. Travel costs, childcare costs, time off work etc. The cost of the test is covered by the NHS*
20	**Gender of healthcare provider** *Gender of the staff member giving you the test*
21	**How test results are returned** *e.g. in person, phone, letter*
22	**Test location** *Where the test takes place*
23	**Test duration** *The length of time spent having the test*
24	**Information included with the invitation** *The level and type of information received about the test*
25	**Waiting time for the test** *How long a person has to wait to have the test after being referred by their GP*

Figure 5.4: Example of a choice task from the BWS study



5.5.1.2 Questionnaire development

The questionnaire was developed and collected using LimeSurvey. A full version of the questionnaire is provided in Appendix 5.2. The survey consisted of four stages:

i. Sociodemographic questions (e.g. age, ethnicity, education, employment status)
ii. Best-worst questions (including a warm-up task)- Respondents completed 32 choice tasks; a warm up task, thirty choice tasks from the experimental design plus an additional attention check.
iii. Follow up questions relating to the best-worst questions (e.g. task difficulty, the importance of included attributes and if there were any additional attributes, they considered important)
iv. Background questions relating to health history and behaviours (e.g. behaviours such as smoking, conditions such as endometriosis etc.), health and testing history, risk attitude

Question framing and answer categories were copied directly or adapted from established national surveys such as the census where possible. Early piloting with a subset of women (n=5) suggested the survey would take 30 minutes to complete. Given the length of the survey and the online administration method, three attention checks were embedded into the survey to ensure adequate attention throughout the survey. Attention checks were designed following the instructional manipulation format and were designed in line with LimeSurvey 'fair attention check' guidelines (e.g. "Select 'very important' to indicate you are paying attention") (Hauser & Schwarz, 2015). Respondents who failed all three attention checks were removed from the analysis.

5.5.1.3 Participants

The target sample size was 150 women and was estimated alongside the experimental design generation in SAS. Participants were recruited via an online platform explicitly targeting researchers conducting social and economic science experiments (Prolific). Prolific offers higher transparency in terms of subject pool and screening compared to other online platforms; results have been shown to be of comparable or better quality than university research lists and has been used widely in within hundreds of published studies across disciplines (Palan & Schitter, 2018; Peer *et al.*, 2017). Participation was limited to women over the age of forty years (no upper limit) living in England or Wales, no other limitations were applied. Electronic consent was obtained at the outset of the survey and participants received a payment of £3.50 (£7 per hour) for their time.

5.5.1.4 Analysis

Online survey data was downloaded to SPSS Statistics 26. Analysis of data was conducted in R using the Support.BWS (Aizaki, 2014) and mlogit (Croissant, 2012) packages.

Importance scores for each attribute were calculated using the counting approach; whereby the number of times an item was picked as 'most important' is subtracted from the number of times it was chosen as 'least important'. Each item appeared 6 times across all choice tasks, meaning best-worst scores could range from -6 to +6 at

the individual level. Individual scores were aggregated to calculate an overall mean score for the population. Scores were then standardised to a scale from -1 to +1[7]. Confidence intervals for all estimates were defined as ±1.96 times the standard error (SE).

Importance scores represent the relative importance of each item across the women in the sample. A score tending toward +1 indicates an attribute is highly influential to decision-making around testing whereas attributes with scores tending towards -1 demonstrate less salience. This simplistic approach provides an easily calculable and interpretable results and scores have been demonstrated to lead to similar results as coefficients from logit models (Finn & Louviere, 1992; Flynn *et al.*, 2007; Marley *et al.*, 2012).

A multinomial logit (MNL) model was also estimated and results between the two approaches were compared for completeness.

5.5.2 Results

Survey responses were collected in March 2020. In total, 159 responses were collected; four submissions were incomplete, two were removed due to failing all attention checks and a further three responses were removed due to incorrect completion of the best-worst section of the survey (e.g. ignoring the 'least important' column across all choice tasks), resulting in 150 responses being included in the final analysis. The average response time for the questionnaire was 29 mins 51 seconds. Respondents varied substantially in how difficult they found the best-worst portion of the questionnaire with 42% (63/150) finding it easy or very easy but 38% (57/150) finding it difficult or very difficult.

5.5.2.1 Individual characteristics

Descriptive statistics of the sample are presented in Appendix 5.3. The age of respondents ranged from 40 to 87 years old with a mean age of 51.4 years (SD=9.1).

[7]Scores were standardised using the following equation: best-worst score/(number of times item appeared x total sample size)

The majority of participants were white (120/150; 80%), married or in a relationship (97/150; 65%) and employed on a part-time or full-time basis (79/150; 52%).

The majority of women perceived their risk of cancer as low or average (116/150; 77%) and anxiety relating to ovarian cancer was generally moderate-low amongst respondents (90/150; 60%). Forty (27%) women reported being previously tested for ovarian cancer, with CA125 blood test being the most common test. Overall, fifty women (33.4%) reported undergoing a TVUS for any reason. Twenty-five respondents reported knowing someone who had previously been diagnosed with ovarian cancer, seventeen of whom were immediate blood-relatives.

Crucially, when asked, 127 out of 150 women (88.7%) stated they wished for a great deal/a lot of involvement in decisions relation to their own care but only 34/150 (22.7%) currently feel this was achieved, with 17/150 (11.3%) respondents reporting feeling currently unable to be involved in medical decisions at all.

5.5.2.2 Best-worst results

Counting results are presented in Figure 5.4, in which differences in relative importance are represented spatially by the distance between each circle. Scores were bound between -1 and 1. Scores tending toward the extremes of the scale imply homogeneity across respondents and consistency between responses across questions on an individual level. Importance scores ranged from -0.224 to 0.380, suggesting high levels of heterogeneity in preferences regarding test characteristics across respondents (Appendix 5.4).

Overall, "*Chance of dying from ovarian cancer*" (0.380, 95% CI [0.26–0.49]) was the most important attribute to women when considering ovarian cancer testing, followed by "*Test sensitivity*" (0.308, C I[0.21–0.40]). Conversely, "*Time away from usual activities*" (-0.244, 95% CI[-0.33–(-0.15)]) and "*Gender of healthcare provider*" (-0.243, 95% CI[-0.35–(-0.14)]) were considered least important to women when facing diagnostic testing and were statistically indistinguishable from each other. The most important attributes to respondents were clear and distinct: however, spatial

visualisation of results (Figure 5.5) demonstrated clustering towards the centre and bottom of the scale. Clusters are distinct to other attributes but preferences between attributes within clusters is less distinguishable.

MNL results are shown in Appendix 5.5. The order of importance remained consistent across the two analysis methods and the estimates were highly correlated; (Appendix 5.6) therefore subsequent subgroup analyses utilised the counting methods to facilitate ease of interpretation.

5.5.2.3 Self-reported importance of attributes

Following the BWS exercise, respondents were asked to indicate any attributes that they would always consider important if faced with a decision between competing tests for ovarian cancer and which factors, if any, they would never consider important. Figure 5.6, shows the responses for each attribute. Results generally mirrored “most” and “least” scores from the BWS section of the survey. “Chance of dying from ovarian cancer” was considered to be always important by 80.6% (121/150) of respondents, with just one individual responding with “never important”. Except for “time away from usual activities”, at least 50% of respondents found all attributes important to some degree, demonstrating the relative nature of BWS results.

Figure 5.5: Best-worst scaling results. The distance between attributes is a spatial representation of the difference in relative importance between attributes on the latent importance scale.

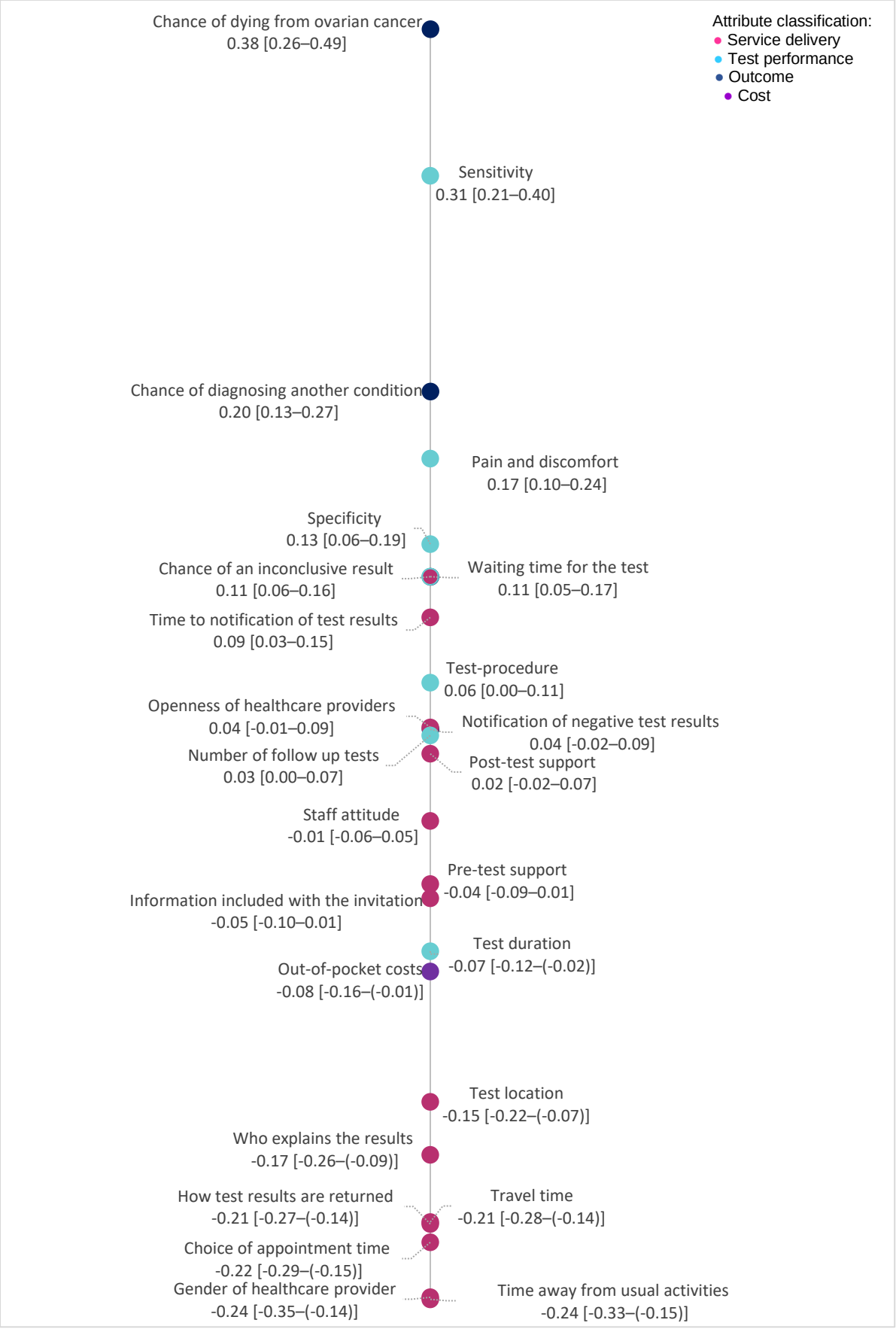
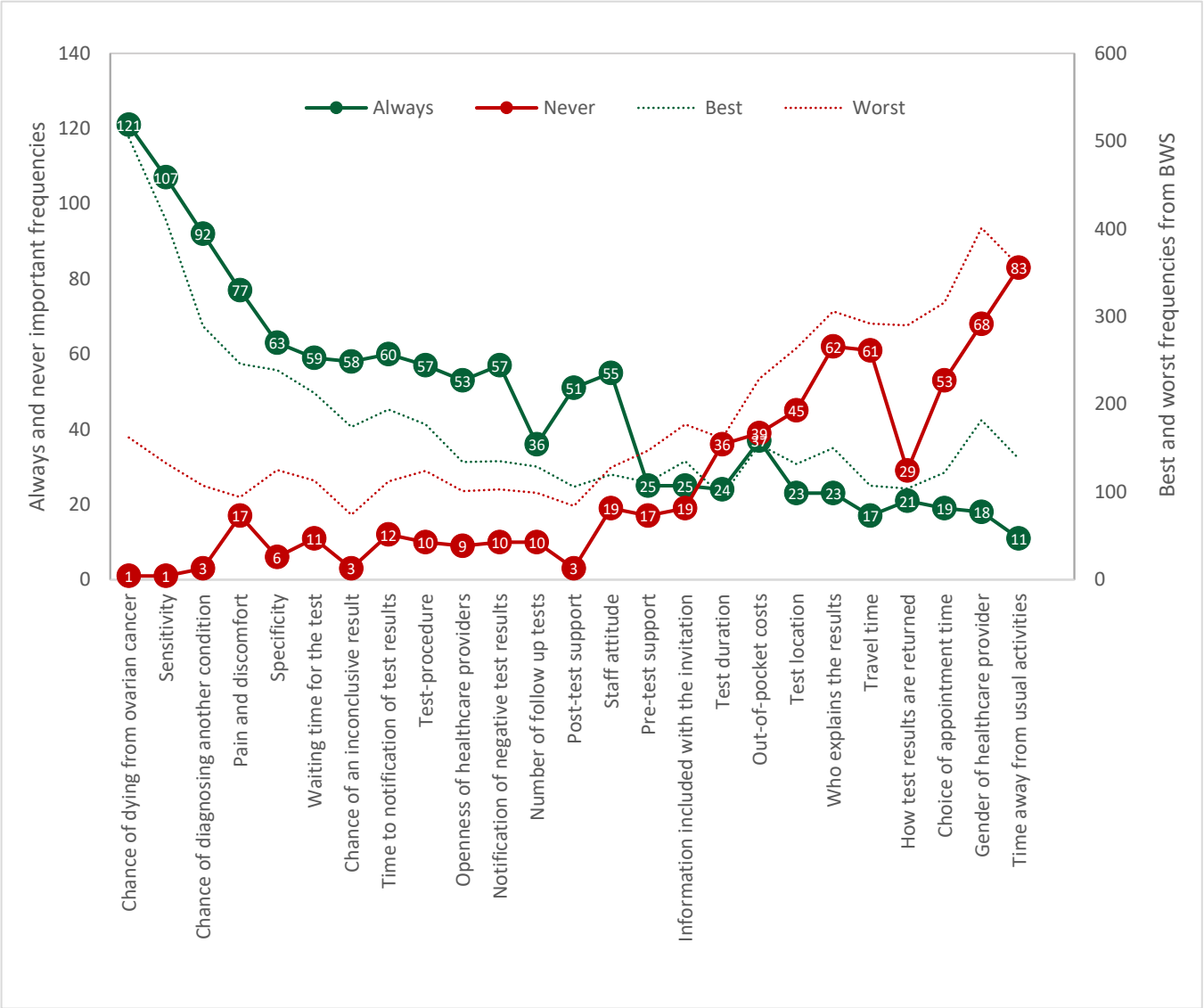


Figure 5.6: Number of respondents that indicated each attribute was "always important" or "never important" to them when considering ovarian cancer testing



5.5.2.4 Additional attributes

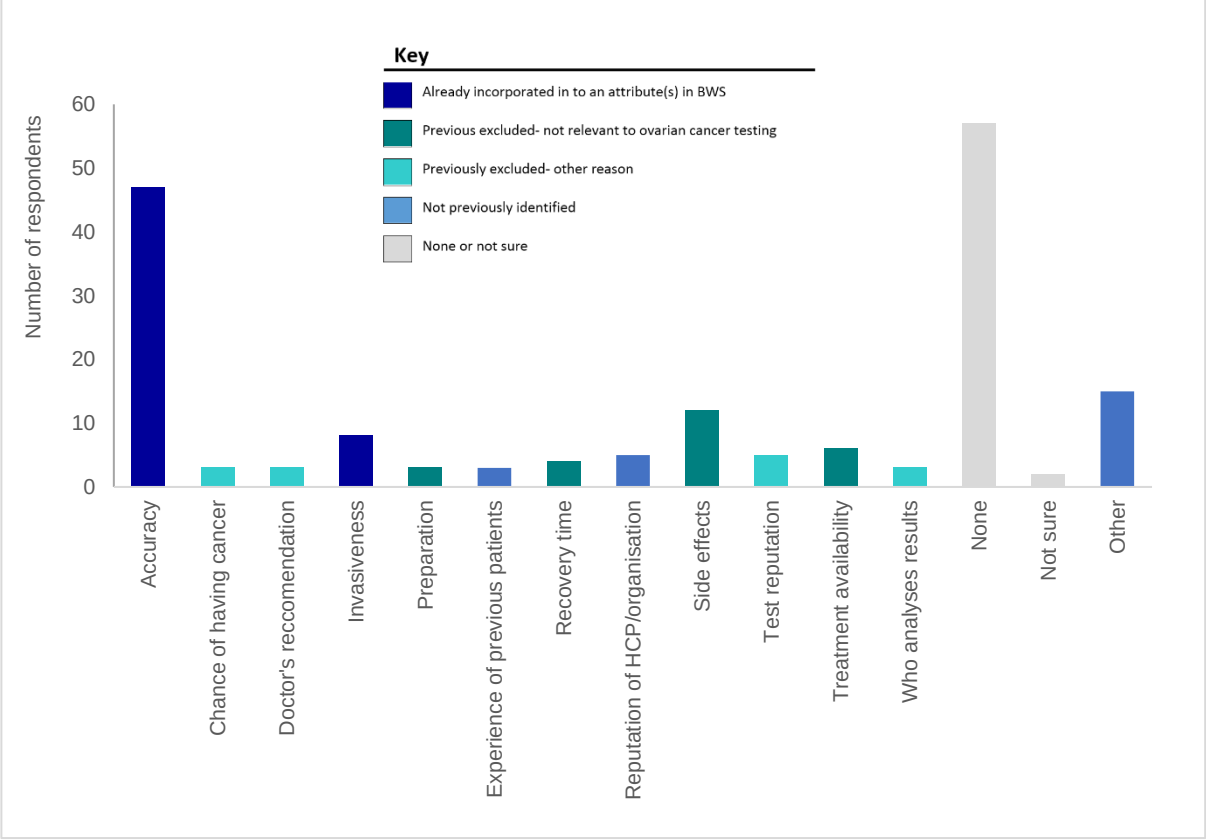
As part of the questionnaire, women were given the opportunity to suggest additional factors which they would consider important if they were experiencing symptoms and making decisions in real life. Answers were collected in a free-text format. Responses were content analysed within Microsoft Excel and similar responses were grouped together. Results are shown in Figure 5.7; 38% of respondents indicated that the twenty-five attributes we included were sufficient for decision-making. Remaining respondents suggested additional attributes covering thirteen categories.

Interestingly, responses relating to accuracy were most common, despite the inclusion and high importance of, both test sensitivity and specificity. Responses used a number of terms reflective of accuracy including “reliable”, “safest”, “best” or “false-positive” results.

Almost all of the remaining suggestions had been previously identified and excluded during the development of the survey due to lack of lack of relevance to tests available for diagnosing ovarian cancer (e.g. preparation, recovery time, side effects). On the other hand, the exclusion of items such as “doctor’s recommendation” and “test reputation” were more discretionary, based on researcher assumptions and highlighted a need to verify this decision through qualitative investigation.

Finally, there were several additional attributes that had not previously been identified within the literature, including the reputation of the HCP or facility performing the test and the experiences of previously tested women.

Figure 5.7: Additional attributes identified from the free text question inviting respondents to add any additional attributes they consider important



5.5.3 Discussion of BWS results

The results of the BWS study highlight the importance and current neglect of incorporating the preferences of patients into medical decisions, particularly those around medical testing. When asked, almost 90% (134/150) of the women in this study stated they wished to be heavily involved in medical decision-making, however, at present less than a quarter of women felt they are able to do so. Best-worst scaling provides a straightforward method for capturing the priorities of women that is preferable to ranking/rating tasks due to its ability to measure the relative importance of a large number of attributes whilst also limiting complexity and cognitive burden for respondents by requiring the comparison of just five attributes at a time.

BWS results were particularly useful in identifying attributes at either extreme of the scale, clearly demonstrating several attributes that were most important and least important to women when considering testing for ovarian cancer. Overall, "*chance of dying from ovarian cancer*", "*sensitivity*" and "*chance of diagnosing another condition*" were the most important attributes across all estimation methods and subgroup analyses. Oppositely, "time away from usual activities" was always scored as the least important factor, with "*choice of appointment tim*e" and *"who explains the results*" also consistently featuring amongst the lowest scoring attributes. Perhaps due to the low concurrence between attributes and high levels of heterogeneity, discrimination between mid-range attributes was less clear.

The prioritisation of outcome and test-specific attributes and the relative lack of importance placed on service delivery attributes mirrors findings from the wider literature examining preferences towards cancer screening and diagnosis (De Bekker-Grob *et al.*, 2013c; Howard *et al.*, 2011; Mansfield *et al.*, 2018; Miles *et al.*, 2019), previously highlighted within (Hall *et al.*, 2021).

5.5.3.1 Limitations

Differences between tested and untested women also serve to highlight the limitations of the object case (case 1) BWS approach. Using the object case approach meant respondents were given limited context about the potential range (levels) of attributes.

Feedback during the piloting stage revealed respondents tended to make assumptions about likely levels during choice tasks, particularly for more familiar concepts such as “travel time”. This may be problematic in cases where there are significant gaps between the predicted levels and reality, for example testing-naive women may have anticipated higher levels of pain and discomfort during tests than actually experienced by the majority of women. Perhaps for this reason, case 2 (“profile case”) BWS, which includes levels, have been predominately used in health research to date (Cheung *et al.*, 2016). However, the inclusion of levels increases the cognitive burden for respondents and was not deemed appropriate or necessary due to the large number of attributes and choice tasks within this study and given the ultimate aim was prioritisation.

The length of the survey and the high number of attributes considered throughout the survey are another potential limitation of this study and reflected by the high proportion of respondents that reported the task as difficult or very difficult. However, drop-out rates were low (3%; 4/150). Additionally, there were no statistical differences between responses according to reported task difficulty.

Despite these shortcomings, BWS provides a straightforward method for capturing the priorities of women. The method remained preferable to ranking/rating tasks for the purpose of this chapter due to the ability to measure the relative importance of a large number of attributes whilst also limiting complexity and cognitive burden for respondents.

Summary of findings from the BWS prioritisation exercise:

BWS scores demonstrate the importance or unimportance of an attribute, relative to the other attributes within the study and do not necessarily signal absolute importance. An issue that highlights the importance of combining BWS findings with other methods to ensure a rigorous selection process to avoid excluding any potentially important attributes. Uncertainty around the importance of mid-scale items combined with heterogeneity across respondents means BWS results were used for attribute exclusion rather than identification of attributes to include.

Excluded attributes consisted of those with an importance score of less than zero. This meant the exclusion of 11 attributes, leaving fourteen for consideration during the subsequent stages.

Findings also highlight potential misunderstanding/misinterpretation of some attributes, especially relating to test performance/accuracy

5.6 Online workshops and questionnaire with women over 40

5.6.1 Methods

BWS findings were supported by qualitative engagement with the target audience through online workshops and a supplementary follow up questionnaire. Workshops were broader in topic—asking women to identify and discuss features of medical testing that were most important to them. This was an introduction to thinking about this issue, since women would have different knowledge and experience of ovarian cancer and testing for ovarian cancer. This approach was adopted to avoid overwhelming participants particularly due to the time constraints and online format of sessions. Findings from the workshops were then used to generate a follow up questionnaire which asked participants to reconsider their priorities specifically in relation to ovarian cancer testing.

5.6.1.1 Workshops

Three workshops, each with three participants were held in May and June 2020. Smaller groups were preferred to maximise collaboration and manage discussions during the sessions to give all participants the opportunity to contribute. The workshops were designed to be as interactive as possible with researchers acting as facilitators rather than leaders of the sessions. Workshops were facilitated using Conceptboard (conceptboard.com), an online platform that allows participants to collaborate by adding to and editing communal work pages, known as boards, in real-time. Conceptboard was chosen because it has integrated video and audio calling, meaning participants would not have to load multiple programmes to take part in the workshop.

Boards with background information and tasks were designed and piloted prior to the workshops (Appendix 5.8). Participants were sent the link to the board and instructions on how to access conference calls before the session. Participants were unfamiliar with Conceptboard prior to the session so the smaller groups also allowed participants to have more help whilst practising using the tools needed during each session. Online workshops are considered more demanding than in-person for many, therefore sessions were limited to a maximum of 1.5 hours.

The workshops consisted of three main tasks:

1. Women were asked to think about a time when they or someone close to them had been offered a medical test. They were then asked to identify up to five factors or features of the test they wanted or would have wanted to know about before deciding whether to accept the test. The answers were added to the board as post-it notes so everyone could view them and each woman could explain their answers.

2. The attributes from the best-worst study (n=14) were introduced to the women. The women discussed their understanding and agreement with the attribute wording and definitions. They then compared their responses from task 1 with the existing attributes and grouped similar responses together.

3. Voting- women were shown a board with all the unique attributes discussed throughout the workshop (BWS plus additional attributes identified by the women). Each woman was assigned 5 colour-coded icons and asked to place these on the five attributes they felt were most important overall. Women were then shown the five attributes that were most important during the BWS and discussed this. Finally, women were asked if there were any attributes they thought were unimportant to them or any additional attributes they still felt were missing.

Recruitment

Women were invited to take part via social media (Facebook) and word-of-mouth through personal contacts. Interested individuals were contacted to discuss the project and were sent a plain language summary of the study before confirming participation.

5.6.1.2 Follow up questionnaire

All workshop participants were invited to complete a questionnaire on the importance and acceptability of remaining attributes. The purpose of the questionnaire was to consolidate the findings from the workshops in the context of ovarian cancer testing. Attributes were revised based on workshop findings and preliminary levels were assigned based on available literature relating to ovarian cancer testing and insights and expertise of the PhD team. These levels would be formalised following attribute finalisation.

Questionnaire description

An invitation to complete the online questionnaire was sent via email. Participants were now familiar with the themes of study allowing more detail and contextual information to be introduced to ensure responses were of maximum relevance to ovarian cancer testing. Information about ovarian cancer and tests available were provided in an interactive presentation hosted on Prezi and embedded in to the survey (https://prezi.com/p/6llukfnhqmgv/?present=1) .

Respondents were introduced to the revised attributes and newly assigned levels and asked to complete two rating tasks:

1. The importance of each attribute on a 5-point Likert scale (Not at all important—Very important)
2. The acceptability of each of the levels assigned to each attribute on a 5-point Likert scale (Highly unacceptable—Highly acceptable).

This two-dimensional approach allowed the attributes of most relevance to be identified. For instance, an attribute may be considered highly important but if all levels were acceptable this attribute may be considered less relevant for inclusion. A copy of the questionnaire is provided in Appendix 5.9.

5.6.2 Results

A visual summary of the results from all three workshops is shown in Figure 5.8. Full results can be found in Appendix 5.10.

5.6.2.1 Understanding of existing attributes

Most attributes were well understood by participants, although as expected, some difficulties understanding and interpreting attributes relating to accuracy were expressed across all three workshops.

During the first workshop, women struggled to distinguish between test reputation, sensitivity and specificity attributes. Terms like “false positive” were viewed as “too medical”, instead participants chose to focus on “test reputation and evidence” which was interpreted as a combination of how established the test was (i.e. how long it had been used for) and accuracy (i.e. receiving a correct result). Subsequent groups were more familiar with these terms and even had direct experience of such results; however, incorrect results were viewed more as mistakes that needed to be reported and corrected by test manufacturers rather than an inherent risk associated with medical tests.

These findings highlighted the need to re-work and simplify the definitions for sensitivity and specificity attributes. To provide more clarity, “test reputation or

evidence” was reworded into two attributes with wording based on the literature reviews and workshop discussions: “Length of use” and “Acceptability of the test to GPs”.

5.6.2.2 New attributes

Seventeen additional attributes were introduced across the three workshops. Many of the newly introduced attributes had been previously ruled out during the BWS or internal reduction stage (e.g. preparation, procedure, side effects, test duration, test location).

New attributes and discussions throughout the workshops heavily focused on communication and information-sharing throughout the testing process (e.g. information about the possible outcomes, information provided before the test, reason for the test, and alternative options to the test). Women also discussed issues about feeling dismissed or unheard by doctors (e.g. knowing their own bodies/symptoms and the need to self-advocate). These themes were reflective of the importance placed on communication within the existing qualitative literature (Fitch *et al.*, 2002). In response, these issues were combined with the “openness of healthcare providers” attribute to create a new attribute; “communication skills of HCPs”.

Access to treatment was also a common new theme (e.g. “quick treatment if needed”, “access to right specialists”, “future options”); although important issues they were deemed beyond the scope of the research question which specifically focuses on diagnostic testing.

5.6.2.3 Most important attributes

In all workshops, participants stated that they found all the attributes important and could not pick any to eliminate. Despite the large number of “new” attributes, when selecting the five most important attributes, 82% of votes were assigned to the existing attributes identified during the earlier stages of attribute development. There was substantial variation in the most important attributes, mid-range attributes during the BWS were frequently identified as among the most important (e.g. post-test support,

waiting time, inconclusive results) and every pre-existing attribute was chosen as important by at least one participant across the three workshops.

All attributes that received at least one vote for “most important” during the workshop were maintained at this stage provided they had not been excluded based on relevance earlier in the selection process. This meant sixteen potential attributes remained and further refinement was needed.

Figure 5.8: Visual summary of results from the three workshops

5.6.2.4 Follow up questionnaire

Seven of the nine women from the workshops responded to the questionnaire and completed all questions. Results are summarised graphically in Figure 5.9. The attributes are listed in order of importance from left to right, meaning "*communication skills of HCPs*" was considered most important and "*length of use*" was least important to respondents.

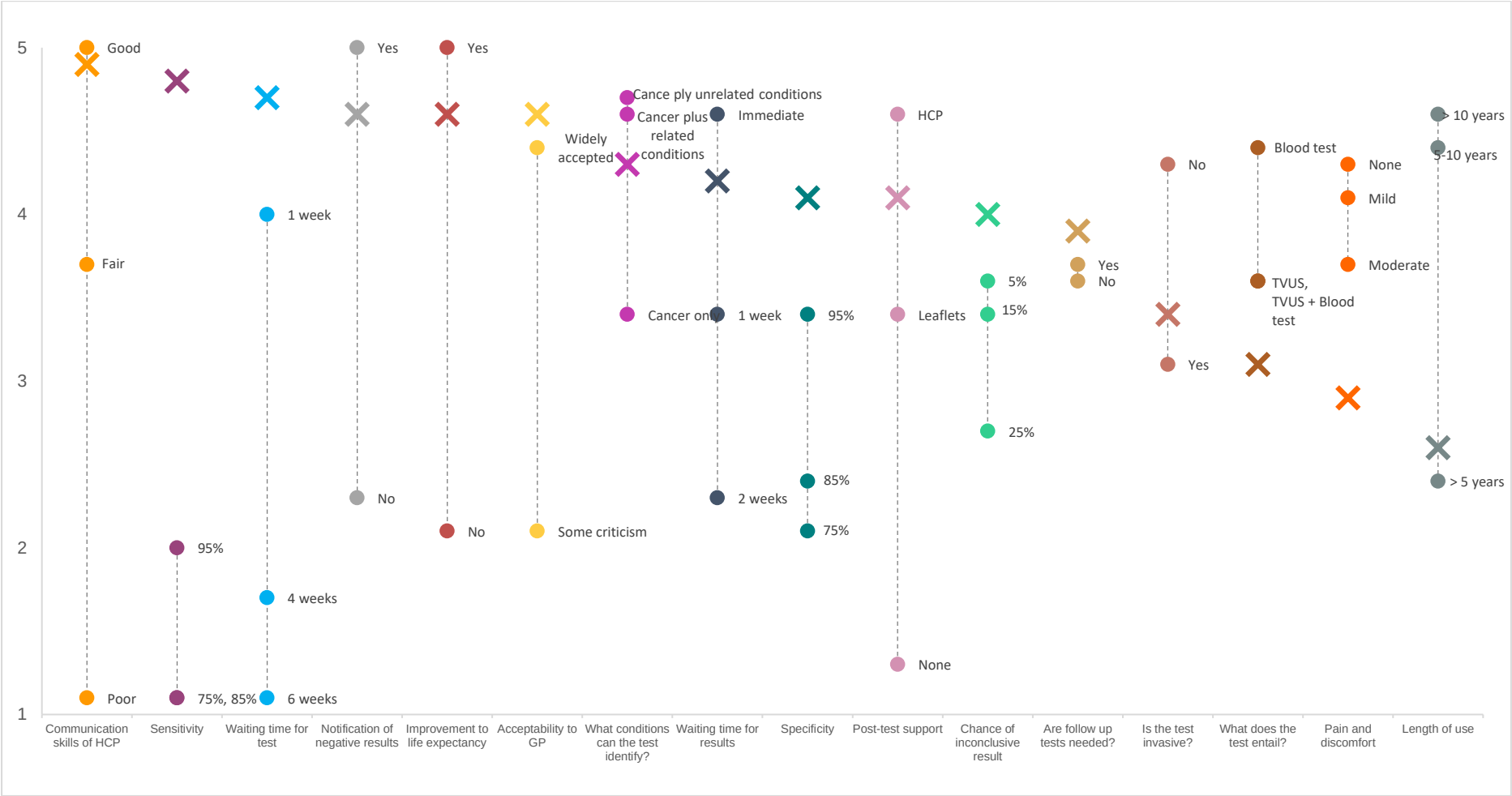
Higher levels of attribute importance were typically accompanied by larger ranges of acceptability between the highest and lowest levels, as demonstrated by "*communication skills of HCPs*", "waiting time for the test" and *"improvement to life expectancy*". Oppositely, attributes which were rated less important typically experienced little variation in the acceptability of attribute levels (e.g. "*is a follow up test needed", "what does the test entail?"* and *"pain and discomfort*") demonstrating evidence of greater apathy towards changes in these attributes.

5.6.3 Discussion and reflection of workshops and questionnaire with target audience

The workshops provided invaluable insights into the importance and understanding of potential attributes. There were disparities between findings from the BWS and workshops. Most noticeable was the importance relating to communication with HCPs. These differences highlighting the importance of engaging with the target population during attribute selection and the use of a multi-method approach.

Workshops were initially intended to be in person but were moved online in light of the COVID-19 pandemic. The use of smaller groups and interactive task-led worked well for the online setting by allowing everyone to share their views whilst limiting the length of the sessions as recommended when using virtual methods (Rupert *et al.*, 2017). However, there were also some challenges. Feedback from participants was largely positive; but there were some technical difficulties within each session, mainly when joining the session and during familiarisation tasks. Overall, the main limitation of the online setting was decreased scope for discussion and debate between participants, introducing the need for a follow up questionnaire to consolidate findings.

Figure 5.9: Summary of results from the workshops follow up questionnaire



✕ Mean importance of attribute (1= not important at all, 5= highly important)

● Mean acceptability of attribute levels (1= highly unacceptable, 5= highly acceptable)

Summary of findings from the online workshops and questionnaire for attribute selection:

- Workshops helped to gauge understanding of existing attributes and identify and additional attributes.
- All 14 attributes remaining from the BWS were voted most important during the workshops meaning further exclusions could not be made
- Two further attributes were added meaning 16 were considered during the follow up questionnaire where attributes were ranked in terms of importance and acceptance
- The follow up questionnaire introduced potential attribute levels to the target audience for the first time
- No exclusions were made based on workshop findings alone

5.7 Consolidating evidence—attribute finalisation

5.7.1 Methods

Final attribute selection was determined through discussion amongst the research team. Evidence from all previous stages was consolidated and considered alongside any additional emerging evidence published since attribute development stage began. Selection was based on specific criteria meaning the top attributes from each stage were not automatically included.

To be included attributes must:

i. Be anticipated to be highly important to the target audience (BWS and PPI evidence)
ii. Be clearly explained and understood by the target audience (BWS and PPI evidence)
iii. Be able to be assigned realistic levels that differ distinctly across available test modalities and/or healthcare settings
iv. Be modifiable or implementable within the current healthcare system if found to be important (Existing literature)

Following guidance and previous studies, the target number of attributes was limited to between 4-6 to control the cognitive burden of the final DCE.

5.7.2 Results

Four attributes were selected for inclusion:

i. Test sensitivity
ii. Identifiable conditions
iii. Time to completion of testing (combined two previous attributes: waiting time for test and waiting time for results)
iv. Communication skills of HCPs

Descriptions of these attributes can be found in Table 5.2. Reasons for the exclusion of other attributes considered during this phase are shown in Appendix 5.11. The primary reason was due to evidence of low importance relative to the final attributes. Notably, two seemingly important attributes "*Improvement in life expectancy"* and "*specificity*" were excluded. This was based on newly published evidence that demonstrated these attributes were unlikely to vary substantially between available ovarian cancer tests when used in a diagnostic context for symptomatic patients (inclusion had previously been based on best available evidence which considered the screening setting) (Funston *et al.*, 2020a).

5.8 Level assignment

5.8.1 Methods

The levels of attributes were reviewed upon final selection of attributes to incorporate best-available evidence from patient information forms, policies and latest peer-reviewed published evidence.

5.8.2 Results

Levels assigned to each attribute are shown Table 5.2. Attributes were selected based on a combination of available literature surrounding tests for ovarian cancer and the expertise of the PhD supervision team.

Sensitivity: Existing evidence on the sensitivity of CA125 and TVUS in primary care were limited at the time level selection took place. Levels were adapted from a study estimating the rate of false negative results of CA125 blood tests in primary care (Funston *et al.*, 2020a). Funston *et al.* (2020a) analysed routinely collected data from 50,780 women who had undergone CA125 testing, results indicated a sensitivity rate of 77% (95% CI: 73-81%). However, results also indicated sensitivity was age dependent, with younger patients (<50 years) subject to a lower sensitivity rate of 63% (95% CI: 51-73%) . It was assumed TVUS would achieve more sensitive results, so coverage of levels was extended to accommodate this assumption (Doubeni *et al.*, 2016). The implications of this assumption are explored in Chapter 7.

Identifiable conditions: Compared to CA125 testing, TVUS benefits from the ability to identify other conditions which may the underlying cause of symptoms given the low prevalence of ovarian cancer (e.g. ovarian cysts or fibroids) (Farghaly, 2014; Guerriero *et al.*, 1997; Levens *et al.*, 2009). Recent studies have demonstrated raised CA125 levels may also be an indication of range of other conditions, many of which are unrelated to symptoms of ovarian cancer (e.g. non-ovarian cancers, liver cirrhosis, lung disease) (Funston *et al.*, 2020a; Tahmasebi *et al.*, 2018). However, given the non-specific nature of CA125 testing, identification of alternative conditions requires additional testing. Incidental findings from arising from tests vary in terms of consequences and the benefits and harms to patients, therefore it was considered important to consider public preferences for tests where such results may arise (Ells & Thombs, 2014; Luu *et al.*, 2021; Smith-Bindman, 2018).

Time to completion of testing focused on the time to completion of primary care testing where cancer could either be ruled out or when a patient enters secondary care for confirmatory testing. Levels aimed to covered waiting times for a range of test strategies: CA125 only, ultrasound only and CA125 plus ultrasound (sequentially or concurrently). Levels were selected based on NHS target waiting times and actual waiting times (pre-pandemic) (NHS England, 2018). Waiting times for CA125-based testing were assumed to be 2 weeks since this is a routinely performed blood test (NHS 2022).

Communication of skills of HCPs was assigned qualitative levels. Level descriptions were kept general given the diversity of what was considered “good” or “bad” communication highlighted during the qualitative stages of attribute development.

Levels were further refined following piloting, as described in Chapter 6.

Table 5.2: Summary of the final selection of attributes and levels

Attribute	**Wording**	**Levels**	**Ranking**	
			BWS	**PPI**
Sensitivity	If 100 people ***with ovarian cancer*** had this test, the test would:	**L1:** correctly identify 65 with cancer but miss 35 cases **L2**: correctly identify 75 people with cancer but miss 25 cases **L3:** correctly identify 85 people with cancer but miss 15 cases **L4:** correctly identify 95 people with cancer but miss 5 cases	2	2
Identifiable conditions	**What conditions can the test identify?** *Definition: Most women will not have ovarian cancer but the test may be able to identify other conditions. Some conditions may be the cause of the symptoms but other conditions might be completely unrelated (not caused by the symptoms that you went to the GP about)*	**L1:** ovarian cancer only **L2:** ovarian cancer plus alternative conditions related to your symptoms (e.g. ovarian cysts) **L3:** Ovarian cancer plus unrelated conditions	3	7
Time to completion of testing	**Time to completion of testing** *Definition: The time from the test being ordered to receiving the results. Includes the time spent waiting for the test and time waiting for the results*	**L1**: 2 weeks **L2**: 5 weeks **L3:** 8 weeks	6 (test), 8 (results)	3 (test), 8 (results)
Communication of HCPs	**Communication skills of the healthcare providers** *Definition: Ability of staff to listen and explain things clearly throughout the testing process*	**L1**: Good **L2:** Fair **L3:** Poor	10	1

5.9 Chapter summary

The inclusion of attributes that accurately capture the most important and relevant aspects of the research question to the target population is fundamental to the quality and validity of findings. This chapter utilised a rigorous multi-method approach and placed increased emphasis on transparency of reporting. It is important to note results of this section were not final and were subject to change during the piloting stage prior to final data collection.

Attribute identification through the analysis of attributes used in previous DCEs highlighted the volume and diversity of attributes relating to cancer testing.

Small group workshops allowed in-depth exploration of issues most important to the target population. A key finding highlighted across all three workshops but less apparent in other stages of development, was the heavy emphasis placed on the doctor-patient relationship and the importance of communication. Workshops were also crucial for refining the wording and definitions used to describe attributes. Overall, the addition of this qualitative stage helped to gain a deeper understanding of the complexities around medical testing and the scale of attributes that patients consider important. However, findings were somewhat at odds with the ultimate goal of reducing attributes into a manageable number for a choice experiment. Coast and Horrocks (2007) have previously highlighted similar findings describing the tension between the typical purpose of qualitative work (i.e. obtaining a deep understanding of phenomena) and DCEs which reduce phenomena to a limited number of key concepts. In this instance the challenge was exacerbated by the online format of the workshops which limited discussion and debate during the sessions.

Quantitative elements of attribute development, primarily the BWS study but also the follow up questionnaire with workshop participates provided a more a pragmatic approach to reduce attributes whilst still ensuring attributes remain relevant to the target population. Best-worst scaling had the added benefit of allowing input from a larger sample of the target population increasing the representativeness and generalisability of findings. Comparisons of BWS and DCE findings reveal similar patterns in preferences between the two methods further increasing the motivation for

utilising best-worst scaling during attribute selection (Potoglou *et al.*, 2011). However, BWS scores imply the importance or unimportance of an item, relative to the other items within the study and do not necessarily signal absolute importance or relevance overall. This shortcoming highlights the importance of combining with other methods to ensure a rigorous selection process- through literature and qualitative work to avoid missing important attributes.

Overall, each phase of attribute development extended the cumulative evidence with complementary but also contrasting viewpoints. To date, guidance on attribute development has strongly emphasised the importance and authority of qualitative work during the development of attributes (Bridges *et al.*, 2011; Coast *et al.*, 2012; Lancsar & Louviere, 2008). However, the results of this chapter demonstrate that reliance on a single method during attribute selection may lead to inconclusive or suboptimal evidence on the which attributes to include. Instead, this chapter demonstrates the value of a multi-method approach and in particular alternative approaches such as prioritisation surveys.

Combining contrasting evidence from multiple sources was a key challenge throughout attribute development. Despite the increasing number of published frameworks for attribute selection, advice on reconciling contrasting viewpoints and final selection of attributes from a shortlist remains limited. This crucial step was the most subjective component of the selection process, relying on discussion amongst the research team. In this instance, specific criteria were outlined during this phase to provide a systematic and transparent approach and limit the subjectivity of judgements. However, the reconciliation of competing evidence appears to an area of future research within attribute development.

6 Preferences towards diagnostic testing: Development and testing of stated choice survey

6.1 Introduction

The previous chapter described the attribute development process for a DCE investigating preferences towards investigative testing for ovarian cancer in primary care. This chapter builds on these results by describing the accompanying stages necessary to develop a DCE study. DCE development is a multi-stage procedure requiring the use of multiple techniques. The chapter draws on best practice guidance to ensure a rigorous approach and there is an emphasis on transparent reporting to allow the quality of the study design to be fully assessed by readers.

The remainder of the chapter describes the stages of development, including refining the research question and choice context, generation of an appropriate experimental design using specialist software, survey design and programming. The later stages of the chapter focus on the pilot study used to refine the survey instrument prior to final data selection. A two-stage iterative approach was utilised combining qualitative and quantitative methods. Analysis of pilot responses resulted in several changes to the study and survey instruments. The chapter ends by providing an overview of the final research questions, experimental design and survey instrument.

6.2 Aims

The aim of this chapter is to describe the design and pilot testing of an online survey with an integrated DCE study aiming to measure women's preferences towards diagnostic testing for ovarian cancer.

The objectives were:

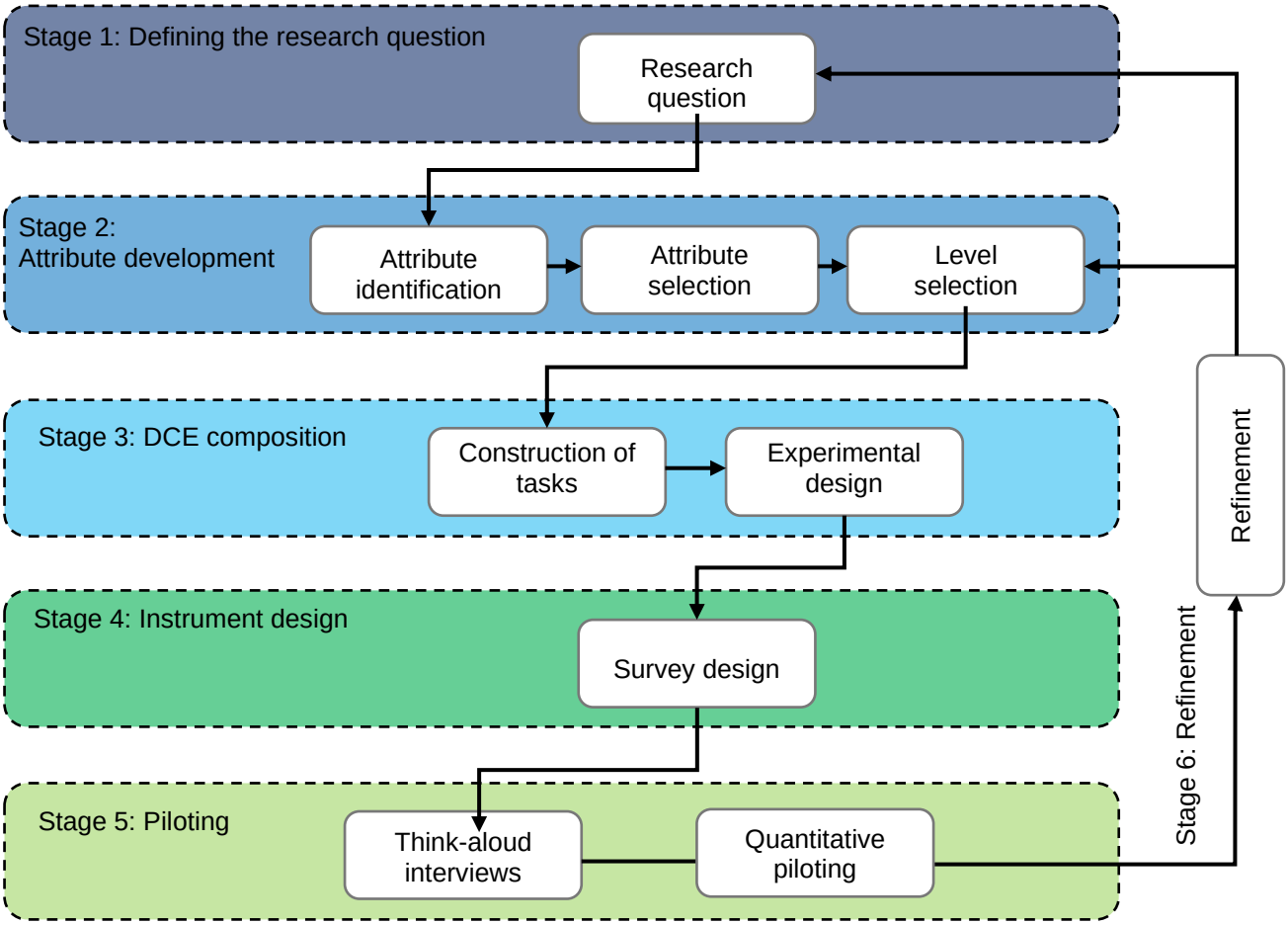
- To develop a discrete choice experiment instrument using leading guidance (Bridges *et al.*, 2011; Lancsar & Louviere, 2008)
- To develop an online questionnaire to collect demographic data to enable subgroup analyses of DCE data allowing relationships between respondent characteristics and choice behaviour in the final study

- To test the feasibility and acceptability of the online DCE survey instrument to the target population through pilot testing
- To refine and finalise the survey instrument through an iterative process based on piloting feedback

6.3 DCE Development

The preference elicitation tool was developed and tested using a multi-method approach. The development process is summarised in Figure 6.1 and followed good research practice guidance (Bridges *et al.*, 2011).

Figure 6.1: Summary of the discrete choice development process. Figure adapted from Bridges et al. (2011)



6.3.1 Stage 1: Defining the research question

The primary research aim was to understand preferences towards tests used to investigate possible ovarian cancer in primary care settings. The purpose was to understand how members of the public make trade-offs between attributes of tests. For example, how much longer are people prepared to wait for a test that provides increased accuracy?

Best-worst scaling results in the previous chapter highlighted the importance of the attribute 'risk of cancer' when making testing decisions. In response, a secondary aim was to understand how preferences vary as the risk of ovarian cancer increases. Banks *et al.* (2014) previously demonstrated willingness to undergo investigative testing was high even when the risk of cancer was low, using a vignette study. This thesis aims to extend this finding by exploring how preferences for the type of test vary based on the underlying risk of cancer. For example, it is hypothesised those with a lower risk of cancer would place greater importance on a test that can identify other conditions, since it is less likely their symptoms are caused by cancer. Addressing this question provides useful insights into delivery of care for low risk but not no risk patients, who are common within primary care.

6.3.2 Stage 2: Attribute development

Four attributes were selected for inclusion in the DCE. Attribute descriptions and associated levels for the pilot version of the study are shown in Table 6.1 (full details on attribute development can be found in Chapter 5). The combination of included attributes allows many questions relating to ovarian cancer to be explored. For example, the use of quick, less accurate triage testing vs longer waiting time for a more accurate test. Or non-specific testing versus a test capable of identifying multiple conditions.

The use of generic attributes allows uncertainties relating to aspects of testing (particularly test performance) to be accommodated. While the incorporation of existing evidence when specifying levels means scenario analysis can be used to

explore changes in demand based on test profiles with differing test performance assumptions.

Table 6.1: Attributes and levels used in the pilot study

Attribute	Wording	Levels
Sensitivity	If 100 people ***with ovarian cancer*** had this test, the test would:	**L1:** correctly identify 65 with cancer but miss 35 cases **L2**: correctly identify 75 people with cancer but miss 25 cases **L3:** correctly identify 85 people with cancer but miss 15 cases **L4:** correctly identify 95 people with cancer but miss 5 cases
Identifiable conditions	**Identifiable conditions** *Definition: Most women will not have ovarian cancer but the test may be able to identify other conditions. Some conditions may be the cause of the symptoms but other conditions might be completely unrelated (not caused by the symptoms that you went to the GP about)*	**L1:** ovarian cancer only **L2:** ovarian cancer plus alternative conditions related to your symptoms (e.g. ovarian cysts) **L3:** Ovarian cancer plus unrelated conditions
Time to completion of testing	**Time to completion of testing** *Definition: The time from the test being ordered to receiving the results. Includes the time spent waiting for the test and time waiting for the results*	**L1**: 2 weeks **L2**: 5 weeks **L3:** 8 weeks
Communication	**Communication skills of the healthcare providers** *Definition: Ability of staff to listen and explain things clearly throughout the testing process*	**L1**: Good **L2:** Fair **L3:** Poor

6.3.3 Stage 3: DCE composition

6.3.3.1 Construction of tasks

Choice tasks should aim to create a decision context that closely simulates the real-life choices faced by respondents. Currently, patient-led diagnostic testing decisions are limited: therefore within this study, task construction aimed to balance respondent burden with realism should greater patient input be an option in the future (e.g. feasible testing options). Numerous factors should be considered during the construction of choice tasks; the use of full or partial profiles, an assessment of the appropriate number of alternatives per task, and the inclusion of opt-out, status-quo and indifference options.

Full or partial profiles: Within each choice set, alternatives may be presented with all attributes or with only a sub-set of the attributes (partial profile). The use of full-profiles is generally considered best practice (Bridges *et al.*, 2011) provided the complexity of profiles is manageable to respondents. It was anticipated that limiting the number of attributes to four would control for complexity; therefore, a full profile format including all attributes in each choice task was selected and tested during the piloting stage.

Number of test alternatives: Choice tasks involved the comparison of two unlabelled test alternatives: 'Test A' and 'Test B'. A paired format was selected on the basis of limiting complexity and current practice in healthcare research (Soekhai *et al.*, 2019). Furthermore, given the choice context, the inclusion of additional choices would not reflect current clinical practice and may be counterintuitive or overwhelming to respondents.

Opt-out alternative: As with any healthcare intervention the final decision to undergo diagnostic testing is at the discretion of patients. Previous research suggests that willingness to be tested for cancer is high (Banks *et al.*, 2014), however, in certain instances, available tests may be deemed too burdensome or poor quality to respondents and some patients may wish to avoid testing altogether (e.g. due to age or pre-existing conditions). Ultimately, given the voluntary nature of testing it was considered necessary to include a "neither test" alternative to reflect the clinical context of the experiment. The wording used to describe the alternative throughout the survey was: "Neither: I would not have either test". No information on attribute levels for the opt-out alternative were provided. Inclusion of a "neither" option also helps to understand demand for testing under different conditions such as alternative modalities or symptom severity (Campbell & Erdem, 2019).

Indifference alternative: Efficient experimental aim to maximise the information that can be gained from each choice tasks by pairing alternatives with similar levels of predicted utility (i.e. utility balance) (Huber & Zwerina, 1996). Inherently, utility balance also makes distinguishing between alternatives more difficult or sometimes impossible. Therefore, an additional alternative was included to allow respondents to

express indifference between alternatives. An indifference alternative avoids forcing respondents to arbitrarily chose between alternatives which they value equally reducing stochastic decision-making and therefore improving the quality of information that can be obtained from responses (Bahamonde-Birke *et al.*, 2017).

Alternative versions: To investigate how the underlying level of cancer risk affects preferences for testing, two alternative sets of DCE questions were created (V1: 3% cancer risk) and (V2: 1% cancer risk). Both versions followed the same experimental design. The risk level was described in percentage and frequency format plus a description of symptoms that pertain to the selected risk.

To provide additional context, respondents were asked to imagine they were experiencing a set of symptoms which corresponded to the level of cancer risk within each survey. Symptoms were identified from (Hamilton *et al.*, 2009) based on positive predictive values. The symptoms were:

- 1% risk of cancer: Persistent abdominal pain and a loss of appetite
- 3% risk of cancer: Persistent bloating and a build-up of fluid or gas in the abdomen

6.3.3.2 Experimental design

The experimental design determines the combinations of attribute levels used to construct alternatives and assigns them to choice tasks. The goal when generating an experimental design is to create a set of choice tasks which maximise the statistical information which can be obtained from respondents and used to precisely estimate model parameters (Johnson *et al.*, 2013).

Based on findings from the systematic review in Chapter 4, a main-effects only design was considered appropriate, with previous authors finding no evidence of interaction effects between similar attributes in the context of cancer screening.

A full factorial design would result in $2^1 \times 3^1 \times 4^2 = 96$ choice tasks and 4560 paired combinations. This was considered too large; instead, a fractional factorial

experimental design was subsequently chosen. A d-efficient experimental design was identified using Ngene 1.2 (Choice Metrics). This is a model-specific approach which optimises efficiency by minimising the uncertainty around parameter estimates (i.e. minimise the standard errors associated with each model parameter) (Rose & Bliemer, 2009). Standard errors are estimated and minimised during the generation process based on the asymptotic variance-covariance matrix of the experiment combined with information about the probable parameter values ("priors") (For further information, see (Rose & Bliemer, 2009). For the purposes of piloting, an experimental design was selected assuming categorical attribute levels and using very small priors to indicate direction and allow dominated pairings to be avoided.

The optimal number of choice tasks was determined by normalising the d-error for the most efficient designs (after 10,000 evaluations) and detecting where the point of stabilisation. This approach was first suggested by Rose and Bliemer (2013) and utilises the normalisation formula:

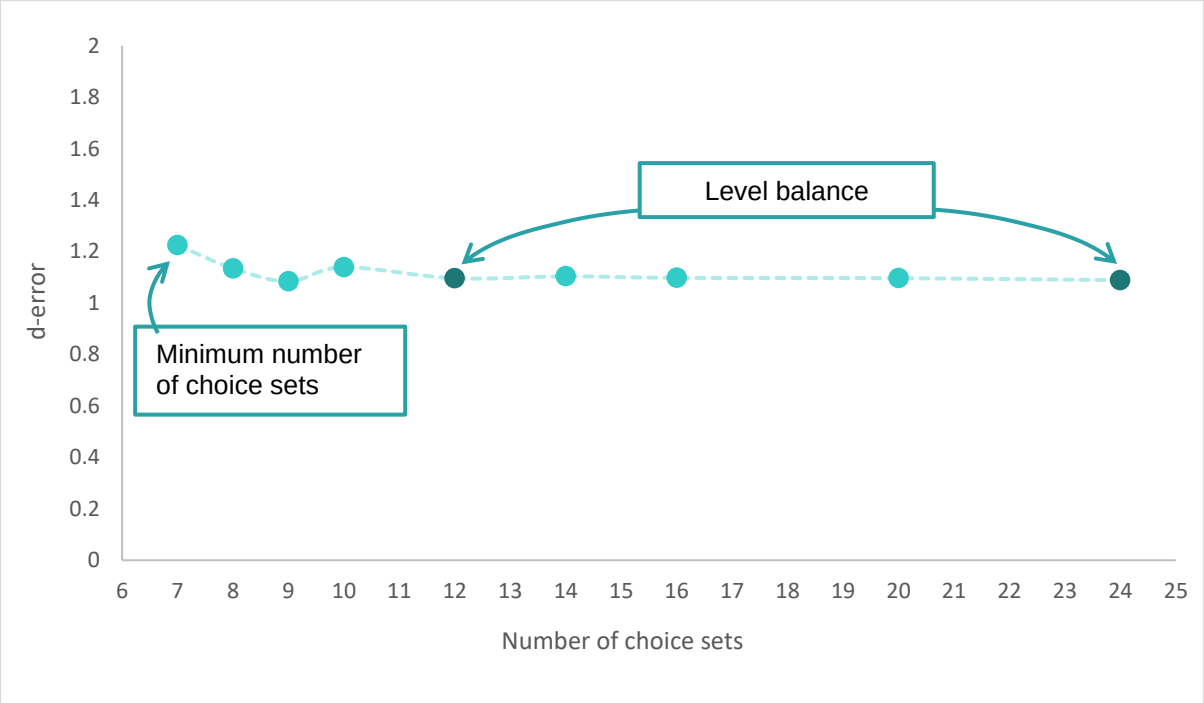
$$S(J-1) > K$$

where S = choice tasks,
J = alternatives
K = number of parameters.

A design with twelve tasks was chosen for the pilot since there no longer appeared to be an increase in information gained per additional choice task beyond this point (based on the most efficient design after 10,000 evaluations) (Figure 2). A design with 12 tasks also allowed level balance to be achieved (i.e. each attribute level appears equally across the experiment). Evidence about the appropriate number of choice tasks per respondent is mixed. In health-based applications researchers typically opt for between 9-16, with a median of 12 (Soekhai *et al.*, 2019). Based on previous studies, blocking was not considered necessary—an assumption that was later tested during piloting.

Following piloting, the experimental design was modified to a Bayesian efficient approach incorporating pilot estimates as priors to improve the statistical efficiency of the design and determine the final sample size (Bliemer *et al.*, 2008). The suitability of each generated design was evaluated by examining post-generation output. Including level valance and level overlap. The correlation between parameters was checked to determine whether a main-effects design was appropriate or whether interaction effects should be accounted for within the final experimental design. An upper value on 0.70 was applied when examining correlation matrices (Bliemer *et al.*, 2008).

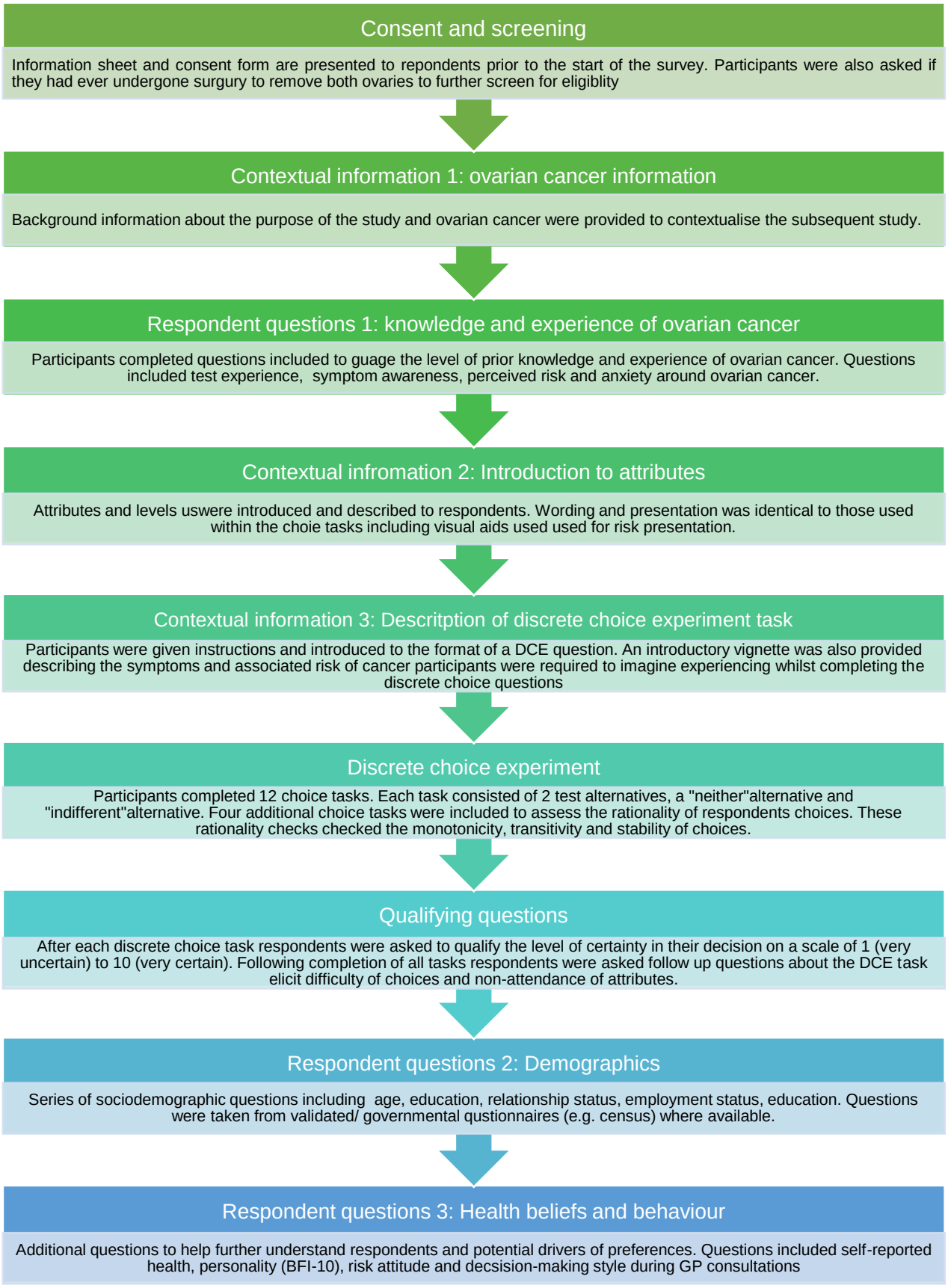
Figure 6.2: Normalised d-error of designs with different numbers of choice tasks



6.3.4 Stage 4: Instrument design

A multi-stage online survey was created and hosted on Limesurvey (limesurvey.org). An overview of the survey is shown in Figure 6.3.

Figure 6.3: Overview of the survey instrument containing the DCE used to elicit preferences



6.3.4.1 Contextual information

The level of information provided to respondents requires balance. Respondents should be given enough information to feel motivated and able to meaningfully complete the tasks. Whereas, too much information can cause information overload or lead to strategic responses (Bridges *et al.*, 2011).

The survey provided contextual information on ovarian cancer and the reasons for the study. Attributes and levels were introduced prior to the start of choice tasks. Attribute development indicated that the ideal communication style differs between people, therefore an additional open-text question was added to understand what participants considered 'good' communication in the context of healthcare.

Respondents were provided with instructions on how to complete choice tasks, and a vignette explaining the hypothetical context respondents should consider when making their decisions prior to the start of the DCE tasks (Figure 6.4). The pilot study also began with a practice question consisting of a choice task with a dominant alternative, designed to introduce respondents to the style of question and to check for non-satiation (i.e. more is always preferable) (Figure 6.5).

Figure 6.4: Introductory vignette presented to respondents before and during the DCE section of the survey to provide context for their decisions

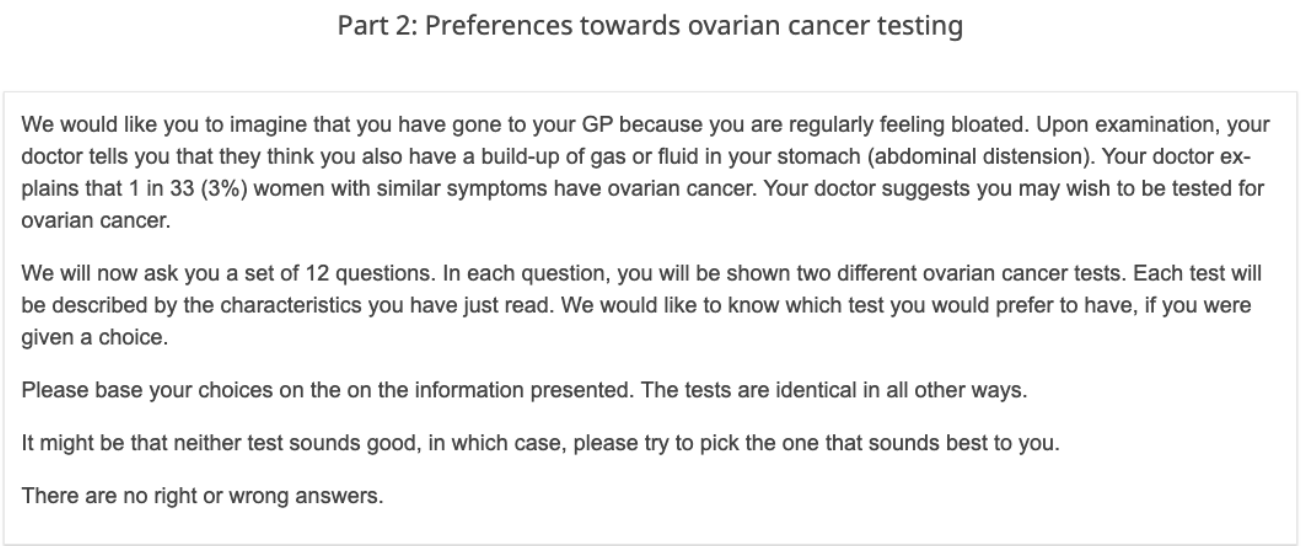
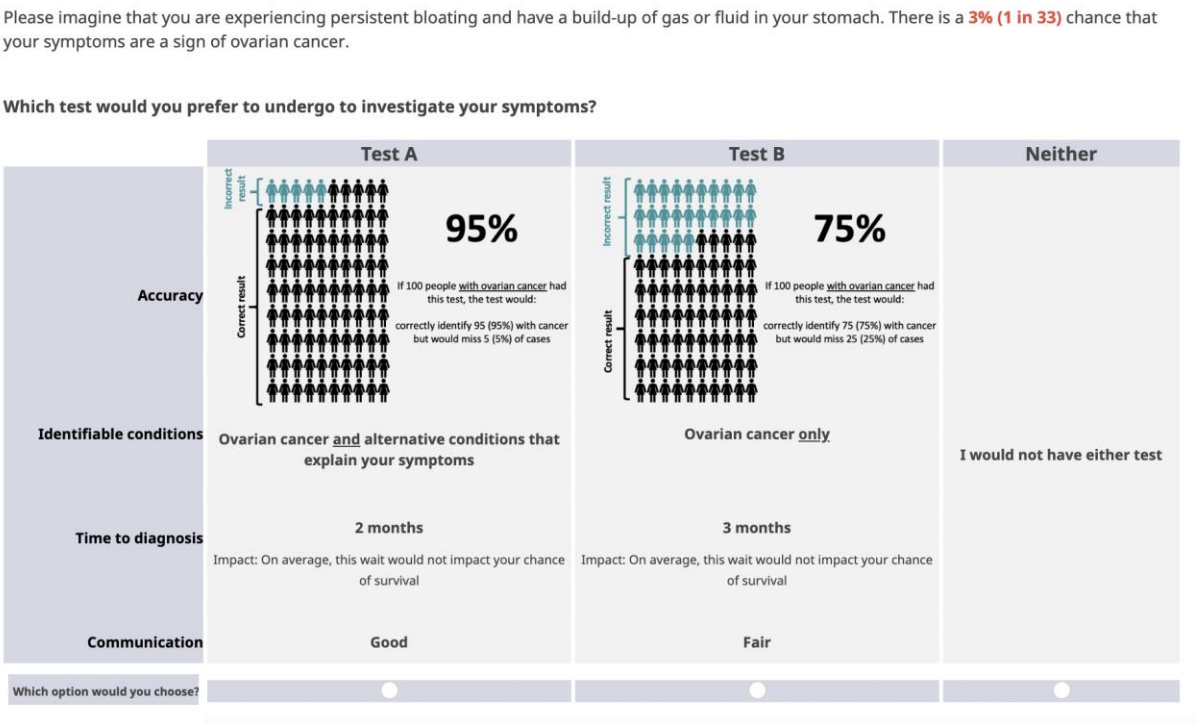


Figure 6.5: Dominant choice task used as a warm-up question during piloting



6.3.4.2 Respondent questions

Respondents were required to complete additional questions alongside the DCE tasks. These questions related to ovarian cancer knowledge, health behaviours and beliefs and general sociodemographic characteristics. Questions adopted wording from established surveys (e.g. census, ONS survey) or previous DCE surveys where available. These questions aimed to gain a better understanding of the population and enabled investigation of preference heterogeneity. These questions were seperated into two sections appearing before and after the discrete choice portion of the survey.

6.3.4.3 Discrete choice tasks

Respondents were randomised and completed one of the two versions of DCE tasks (i.e. 1% or 3% cancer risk). The experimental design for each survey was the same.

6.3.4.4 Rationality checks

The use of rationality checks to assess the internal validity and reliability of responses is recommended (Johnson *et al.*, 2013; Johnson *et al.*, 2019). In addition to the 12 choice tasks included within the experiment design, respondents also completed four

extra questions to evaluate different dimensions of rationality. These validity checks are derived from utility theory axioms of choice which assume decision behaviour should be monotonic, transitive and stable (Plott, 1993). Descriptions of alternative rationality check questions are provided in Appendix 6.1. In total, five aspects of rationality were assessed:

- Monotonicity: An additional choice task with a dominant alternative (objectively better) was included to test to check non-satiation assumption (i.e. more is always preferred) (requires 1 additional choice task)
- Transitivity: Transitive responses imply that if a respondent prefers option A to B and B to C they must also prefer A to C (requires 2 additional choice tasks)
- Stability of preferences: An early choice task was repeated later in the questionnaire to test the stability in choice responses (requires 1 additional choice task)
- Non-trading behaviour: A simple test to investigate non-compensatory behaviour by examining responses for evidence of decisions made based on a single attribute (no additional choice tasks required)
- Flat-lining: This is a straightforward inspection of responses to check whether any participant has repeated chosen a choice task in the same position throughout all question (e.g. selected “Test A” for all tasks) (no additional choice tasks required).

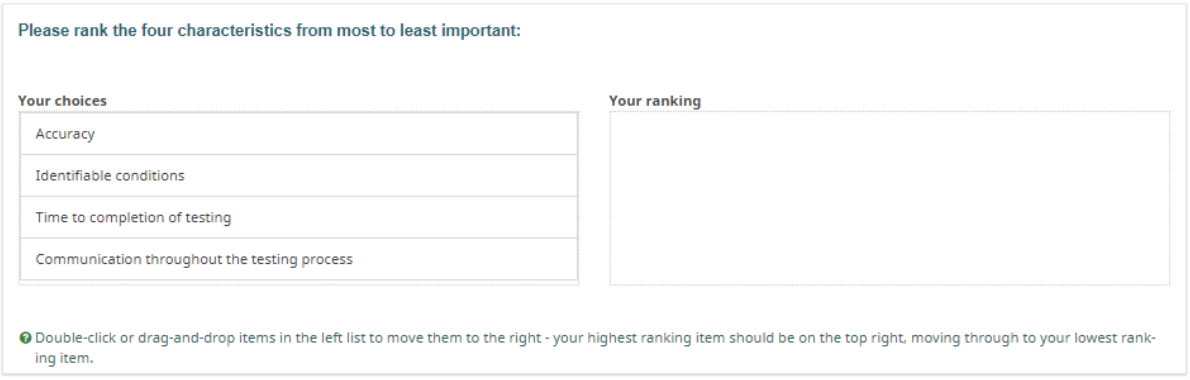
6.3.4.5 Qualifying questions

DCE tasks were followed up with qualifying questions to further understand the responses of participants and the decision-making process. Immediately following each DCE choice task respondents were asked to state confidence in their decision on a scale of 1 (‘Not confident at all’) to 10 (‘Extremely confident’) (Figure 6.5). Following completion of all choice tasks, respondents were asked how difficult they found the DCE questions. Respondents were also asked to rank the attributes from highest to lowest in terms of importance. Responses were later compared with the DCE for consistency (Figure 6.6). The placement of the ranking task was randomised to before and after the choice tasks to control for position bias.

Finally, similar to the BWS survey in Chapter 5, two attention check questions were placed throughout the survey to screen for inattention. An example question was:

"On average, how often do you visit your GP every year?
Please enter 'yes' to show that you are paying attention".

Figure 6.6: Ranking exercise included in the questionnaire



6.3.5 Stage 5: Piloting

The pilot aimed to determine: i) the relevance and understanding of attributes and associated levels to the target audience; ii) understanding of the DCE task and ability and willingness to trade between attributes; iii) the decision process associated with the completion of choice tasks; iv) the acceptability and feasibility of online data collection; v) relevance and acceptability of the survey instrument including accompanying demographic questions.

To fulfil these aims, a two-stage iterative approach was adopted, combining qualitative and quantitative evidence from the target population. Results were used to refine the final survey instrument and embedded DCE design.

6.3.5.1 Qualitative testing: Think-aloud interviews

Methods

Five interviews were conducted using the think-aloud method (also known as cognitive interviews). The purpose of the interviews was to provide in-depth feedback on the DCE choice task, exploring the acceptability and clarity of the survey instrument (including wording and attribute framing attributes) and interface (Willis, 2004). The think-aloud method has been successfully used in DCEs during piloting in many studies to assess the internal validity, acceptability and cognitive processes associated with the completion of choice tasks (Cheraghi-Sohi *et al.*, 2007; Ryan *et al.*, 2009; Whitty *et al.*, 2014). The aim was to assess the survey itself and therefore survey responses from interviews were not analysed.

Recruitment

The sample population was people with ovaries over the age of 40. Respondents were recruited via social media (e.g. Nextdoor, Facebook), advertisement posted is shown in Appendix 6.2. Participants received a £10 gift voucher to thank them for their time.

Interview process

Interviews took place online using Microsoft Teams or Zoom, according to the preferences of the participant. The online setting was adopted to align with national and university-level Covid-19 guidance. Prior to the interview, participants were sent an information sheet (Appendix 6.3) and asked to complete an electronic consent form (Appendix 6.4) Participants were then emailed instructions, including a link to the survey, which went live at the start of the interview and were asked to share their screen whilst completing tasks.

The interview began with a short warm up task to introduce the participant to the process of thinking aloud. For the main task, participants were asked to complete the discrete choice section of the survey, verbalising thoughts about what they were reading, seeing, thinking, doing and feeling throughout. Tasks were completed with minimal interruptions from the interviewer. Where necessary neutral prompt questions such as “what are you thinking now?” or “why did you do that?” were used to encourage the participant to continue think-aloud or to probe for more detail. Before ending the interview, participants were given the opportunity to share any further

observations about the tasks and asked a few follow up questions. In total interviews were designed to last for approximately 45 minutes, with a maximum length of 90 minutes. Two facilitators were present during each interview; one acted as the lead during the sessions (RH) and one observed and took notes (AML). A full interview schedule is provided in Appendix 6.5.

Analysis

Data from think-aloud interviews were formed from two distinct perspectives: firstly, from the participant verbalising their thoughts and feelings whilst navigating the interface and completing the questionnaire but also from direct observation by the researchers of what the subject was doing. Interviews were digitally recorded and transcribed verbatim before being analysed thematically according to the purposes of piloting. Thematic analysis refers to the analysis of qualitative data through the identification of common codes or "themes" (Castleberry & Nolen, 2018).

Think-aloud results

Five interviews were conducted with women living in South West England, aged between 42-73 years old. Interviews lasted between 33 and 57 minutes.

Interviewees were randomly assigned a survey version once the interview was scheduled:

- 1% cancer risk (p1, p2, p5)
- 3% cancer risk (p3, p4)

Think aloud themes

i. Interpretation and understanding of attributes and attribute levels

Feedback during the interviews identified a number of areas of misunderstanding or misinterpretation of attributes.

Time to completion of testing was frequently misunderstood and was interpreted as either the time to receiving a final diagnosis or time to start of treatment by all

participants. Participants also had trouble contextualising the attribute and associated levels. For instance, participants expressed difficulty in establishing a time scale that seemed too long without prior knowledge of average waiting times:

"I'm obviously not an expert to know how long this test actually takes to do" (p1)

A lack of information on the implications of longer waiting times was also highlighted as a point of difficulty when interpreting the attribute:

"I'm not familiar with how progressive ovarian cancer can be, I don't know if that's something you want me to factor in, you know? How critical eight weeks could be in terms of, what the outcome might be or how time sensitive it might be" (p4)

"I mean, I'm guessing. Just from the knowledge I do have that it's not, it wouldn't be that time sensitive" (p5)

Participants had problems distinguishing between related and unrelated conditions when considering the ***identifiable conditions*** attribute and also felt the attribute lacked contextual information relating to the type and severity of potential alternative conditions:

"That one's got cancer plus the conditions. Additional related conditions, unrelated? Just noticed that, unrelated conditions?" (participant 1, q9- noticing that there are two different levels)"

"test A is picking up alternative conditions and test B is picking up additional conditional unrelated ones. I don't, I don't really differentiate there" p4

The attribute appeared to be well-considered during early questions but as tasks progressed identifiable conditions were viewed as less important by some participants:

"It would come down to the reliability and the alternative conditions is again another, is a bonus" p5

"They both pick up other conditions but ovarian cancer would be the primary concern." P4

Accuracy appeared to be well understood by most participants. During the early questions one participant wrongly interpreted the attribute as referring to specificity (or false-positive results). Encouragingly, their understanding appeared to change as questions progressed:

"So, I'm homing in straight away on wrongly being told because that's the case I know with breast cancer there are a lot of false positives aren't there" p2 (q3)

"…a couple of rows there of people being told they haven't got cancer" Q7

"…you've got half of the people being told that they've got cancer, half told they haven't when they have. Wrongly told they've not got it." Q11

Accuracy was the only attribute to include a risk element and accompanying visual aid. Responses to the visual aid were mixed, with some participants commenting that it was helpful whilst others appeared to find it more confusing:

"My confidence is waning rather because they just look so peculiar these, these pictures really" p2

"The colours were good. That was really helpful" p5

Overall, it appeared that visual aids were somewhat distracting to respondents regardless of how helpful they were perceived to be.

Communication throughout the testing process appeared to be well understood and the most relatable to participants early on, with several participants sharing their personal experiences with different styles and quality of communication within their medical care and using this as a baseline of what is expected/acceptable. The attribute

offered an accessible way for participants to familiarise themselves with the task and was focused on more heavily during early questions by all participants. However, similarly to ***identifiable conditions,*** as choice tasks progressed, for some participants, improvements in ***communication*** were viewed as a bonus rather than a crucial deciding factor:

"Communication is poor against fair, that, there isn't a lot of, that wouldn't really affect my decision" p4

For others ***communication*** appeared to remain highly important throughout:

"It depends on who communicates and how well I know them" p3

ii. Hypothetical nature of the task

Realism of scenarios was an issue for some participants, particularly at the start of the questions. Some participants felt the test profiles *"seem a bit too good to be true"* (p2) in comparison to their personal experiences with the medical system:

"It seems so unbelievable that I still can't feel very confident" p2

Whilst other participants expressed disbelief at the attribute levels or combination of attributes, particularly in relation to the communication attribute:

"How can they end up with poor communication? That just doesn't. If they are in this business, how can they have poor communication?" p1

"Well, how would I know how accurate the result was if the communication was poor?" p3

Other participants had less trouble immersing themselves in the scenarios:

"I was thinking about it as I would really feel in those shoes, rather than just as a sort of number" p4

And even participants that struggled with the hypothetical nature towards the beginning of survey expressed increased realism as the survey progressed and they began to feel more familiar with the question format:

"When I first started out it felt, it felt completely theoretical so I felt quite confident because it felt like a made up exercise but as the questions went on you actually started to feel a bit like you were actually having to consider these questions…so my confidence became less because it felt, I know it's not real but it felt a bit more real if that makes sense" p2

iii. Number of choice tasks: adaptive preferences throughout the choice task

Since the total number of choice tasks was on the higher end of current practice within healthcare, the length of the survey was important to explore during the interviews.

Two participants expressed frustration or fatigue during the later stages:

"Not another one" p3

"I was ready to finish. I was a bit fed up of seeing all the charts really" p2

The remaining participants did not have any problems with the length when asked. One participant even viewed the number of questions as a positive aspect, allowing more time to gain familiarity and think more deeply about their true priorities:

"I think it was actually better for me because it did make me think about it more, so it, actually reason it more than I did when I first started." P1

Increasing familiarity and evolving preferences described above were also observed in other participants as questions progressed:

"…it wasn't until later on that I really thought that actually speed might be of the essence and I wasn't at the beginning considering speed" p2

"I started off wanting good communication…but when you then start to look at it, I did know that I changed over time because I changed my thought pattern" p1

Participants also demonstrated greater familiarity with the question format and attributes during the later stages of the questions. For instance, one participant did not acknowledge the opt-out alternative until question 9:

"Now funny enough, I just realised the other box on the end there that says "not choose either test"" p5

"Ah. That's not something I've actually been thinking about properly really" (p2, regarding the relationship between waiting times and survival, q14)

However, participants also displayed signs of heuristic behaviour as tasks progressed, particularly, non-attendance:

"I think I might be ruling out the other conditions" p2 q15

Although this may be due to irrelevance rather than simplification since most participants said task difficulty was manageable when asked although some difficulty with accuracy and visuals was expressed p2

"Oh blimey, this is complicated looking at these isn't it?" p2

"It was confusing looking at so many" p2

iv. Trade-off behaviour

Participants appeared to be daunted when first presented with the first-choice set. Though, once acclimatised, the purpose and process of considering the choice sets

appeared to be well understood and respondents verbalised comparing and equating attributes across the test options well:

"So it looks like we're trying to find balance here between the accuracy and the communication and the time of testing" p1

"…test A for the accuracy and communication and the fact that is picks up other conditions at the same time, even though you have to, completion will take twice as long as test B" p4

As choice tasks progressed, despite acknowledging each attribute when presented with a new choice task, some respondents appeared to narrow down the attributes they considered during trade-off decisions suggesting potential issues of non-attendance of simplifying heuristics:

*"Communication I'm ruling out because I don't, that just, I'm looking at the length of time"*p2

"When you've got variations to play with, it's not so easy. Should I get it done quickly, or shall I have an accurate one?" p3

In particular, one participant began to focus solely on accuracy, suggesting non-attendance could be a potential issue and something to be conscious of during quantitative piloting:

"Accuracy is just the most important thing. It seems a false economy to have a quicker answer if it's not the right one" p4

v. Opt-out and indifference alternatives

Despite expressing difficulty, participants were always able to decide between the two test options and did not use the indifference alternative at any point. Similarly, the opt-out alternative was not selected by any participants during the pilot interviews and the

consequences of selecting the opt-out appeared clear (i.e. you would not receive any test rather than receiving an existing 'status quo' option):

"I don't see why you wouldn't have any test at all" p4

"I didn't really consider the box on the end that said neither because I think, I looked and thought, well actually you wouldn't, because at the end of the day you want a test at some point" p1

vi. Choice certainty

There was some ambiguity around the wording of the choice certainty question which accompanied each choice task asking, "How confident are you about your answer?" with one participant asking for further clarification and another interpreting the question to be asking how happy they were about the test they had chosen. The question also did not appear to be generally well-considered with responses not seeming to match the level of deliberation observed during each question within the interview, with one participant even stating:

"I just went for seven or eight or nine" p3

vii. Difficulty

The perceived difficulty of the choice tasks was mixed across participants, some participants reported finding making the choices easy whilst others verbalised difficulty during the early stages for choice tasks with greater utility balance but also noted that this reflected the real process of decision-making:

"It was very difficult to try and do it quickly and try and weigh things up, to compare and contrast…but I guess maybe that's a bit naturalistic because maybe people don't have a lot of time when you're in the surgery" p2

6.3.5.2 Changes following qualitative piloting

Based on feedback during the think-aloud interviews several changes were made to the survey prior to quantitative piloting.

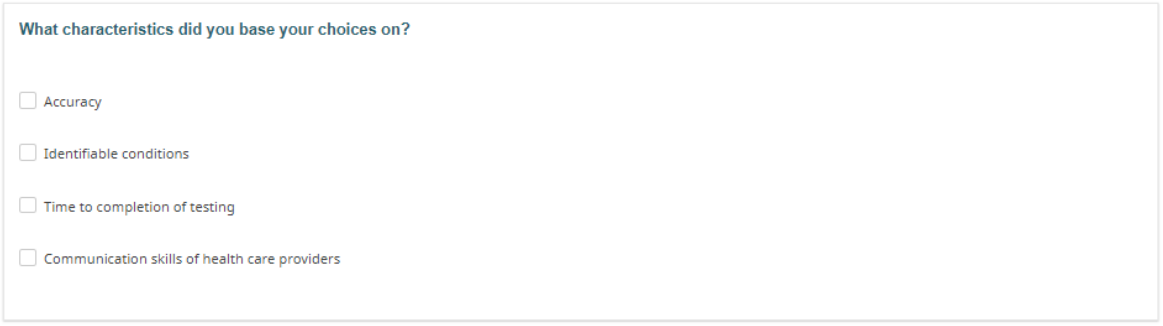
i. Attribute wording and levels

The levels associated with ***identifiable conditions*** were reduced to two: “cancer only”, and “cancer plus alternative conditions related to your symptoms”. Additional information was also added to the description of the attribute at the beginning of the survey to provide examples of what conditions may be identified e.g. fibroids.

Communication throughout the testing process was retained as it appeared very important to some participants and provided an accessible entrance to the survey for participants to familiarise themselves at the start of the survey. Aware that non-attendance may be an issue for some participants- a stated non-attendance question was added to be able to account for this during modelling (Figure 6.7).

Time to completion of testing was reworded to ***Time to diagnosis*** to address misinterpretation during interviews. Attribute levels were updated to reflect the new wording based on literature (Funston *et al.*, 2020a; Lim *et al.*, 2016).

Figure 6.7: Self-reported attribute non-attendance question



ii. Prominence of accuracy/distraction of accuracy visuals

Accuracy appeared to be dominant for participants during the interviews. Interview findings suggested that the focus on the attribute for some participants is likely a rational and conscious decision based on their strong preference for ***accuracy.***

Contrastingly, other participants commented on the prominence of the attribute, both in terms of placement (first attribute across all choice tasks for all participants) and visually, due to the accompanying visual aid. Other participants commented on the usefulness of the visual prompt so it was retained with a few changes to help interpretation (font size, wording). An additional version of the survey was created for piloting to investigate whether attribute position had any impact on importance. The additional version randomised the order of the attributes in each choice task. A survey using the same constant order across all individuals was also piloted for comparison.

iii. Clarifying relationship between diagnostic interval and survival

Interviews highlighted uncertainty and variations in the interpretation of the implications of longer diagnostic intervals, particularly relating to disease progression and survival. To address this issue, ***time to diagnosis*** was amended to include an explicit statement about any survival impact.

The relationship between time to diagnosis and outcomes for patients with ovarian cancer is currently unclear and evidence is mixed (Figure 6.8). Many studies find no association between diagnostic delays and stage of diagnosis or survival, whilst others find longer waiting times are associated with lower QoL outcomes, decreased satisfaction and reduced survival. A third strand of literature reveals an inverse relationship between diagnostic delays and survival. These studies demonstrate the "Waiting Time Paradox" (Neal *et al.*, 2015), whereby patients with the most severe symptoms (possibly indicating advanced or aggressive cancers) are most likely to be referred quickly, resulting in a faster diagnosis but also a poor prognosis.

Given this uncertainty, an additional research aim was identified—exploring the impact of diagnostic delays on preferences. To address this question, existing versions of the survey included a statement that indicated delays in diagnosis had no impact on

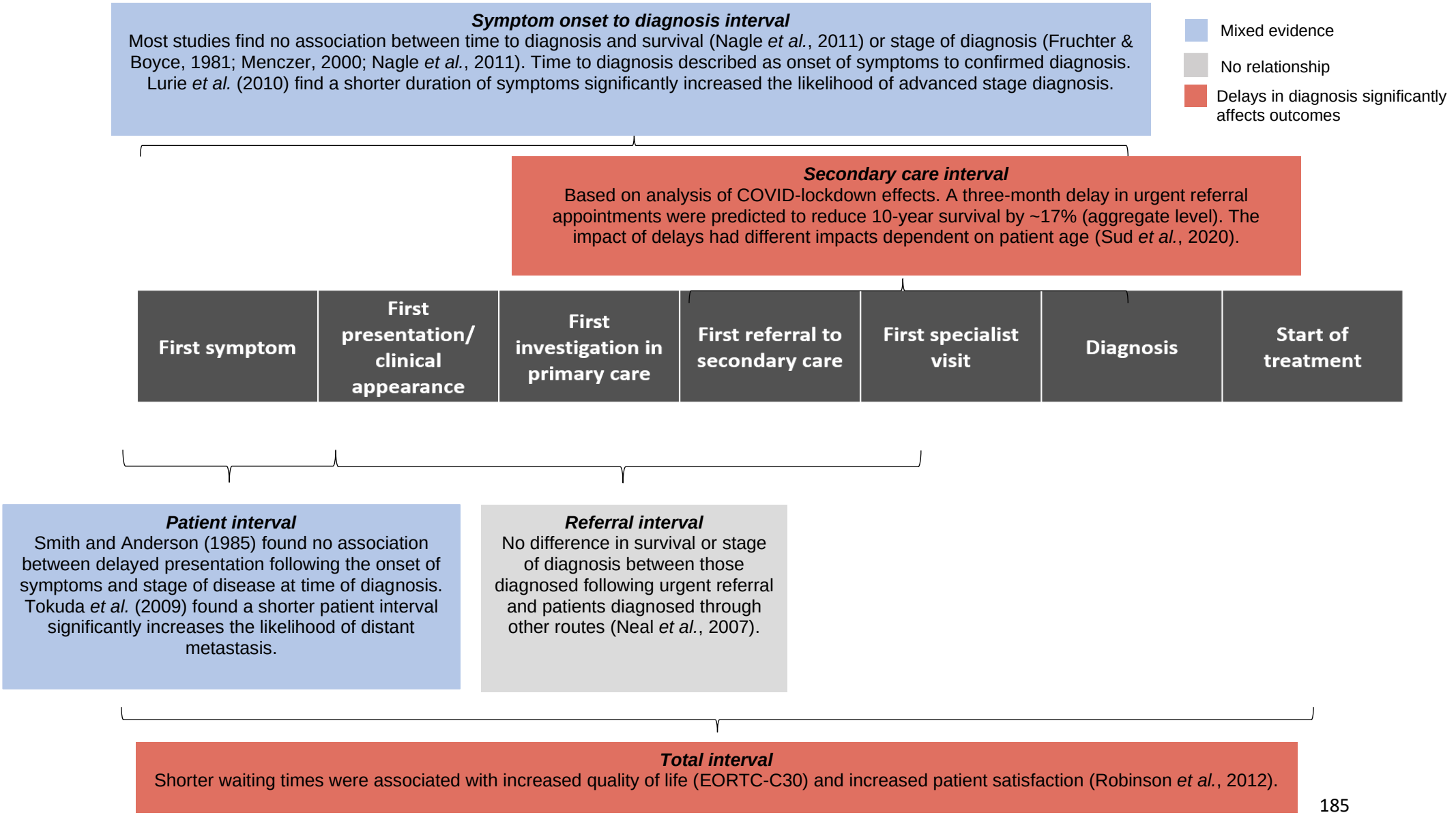
survival. An additional version of the survey was created to explore a potential relationship. In this version, respondents were told that the chance of survival was reduced by 1% per 1 week delay (a delay was a diagnostic time beyond 1 month). This was a placeholder estimate based on unpublished work by a research team member (WH), to assess the feasibility of the research question during piloting.

iv. Additional changes

The interviews also identified some smaller problems relating to the survey instrument which decreased accessibility such as font size and question formats (e.g. the use of a sliding scale rather than checkboxes) which were also adapted prior to the quantitative pilot. The choice certainty question was also reworded to increase the clarity “On a scale of 1-10, how confident are you in your answer?”

.

Figure 6.8: Relationship between delays along the diagnostic pathway and outcomes for ovarian cancer



6.3.5.3 Quantitative pilot study

The pilot study aimed to determine whether the survey was able to be completed online, check trading behaviour and experimental design, and determine sample size requirements for the final study.

Based on the findings and methodological questions arising from the think-aloud interviews, four versions of the survey were piloted:

Version 1: 3% cancer risk, constant attribute order
Version 2: 1% cancer risk, constant attribute order
Version 3: 3% cancer risk, randomised attribute order
Version 4: 3% cancer risk, constant attribute order, time to diagnosis--survival relationship introduced

Each version of the survey was completed by 25 respondents recruited via Prolific (Prolific.co). The sample was limited to women over 40 living in England and Wales with at least one ovary to align with the intended final sample demographics.
Each participant completed the full survey including 12 choice tasks within the experimental design, plus four additional rationality checks

Pilot data were cleaned and analysed using SPSS 26 and Stata 16. Results were analysed using multinomial logit (MNL) model introduced in the previous chapter (McFadden, 1974).

Gauging participant understanding and engagement was a primary aim of the pilot study and was evaluated in a number of ways:

1. Length of time taken to complete the survey
2. Self-reported task difficulty
3. Frequency of attention and rationality check failures
4. MNL results
5. Selection of the opt-out and indifference alternatives
6. Self-reported non-attendance

Results of quantitative piloting

Key summary statistics for respondents of the pilot study are shown in Appendix 6.6. Respondents were aged between 40 and 80 years old, were predominately white (94%) and were mostly test-naïve (88%).

The median completion time across the four survey versions was 15 minutes and 10 seconds (range: 6 minutes 34 seconds – 39 minutes 42 seconds).

i. Self-reported task difficulty

Table 6.2 shows the perceived difficulty of the choice tasks as reported by respondents across the four survey versions. Respondents from versions 1 and 2 were equally split in finding the task easy or difficult. Versions 3 and 4 were viewed as more difficult by respondents, reflecting the methodological variations of these versions. Version 3 (randomised attribute order) was the only version where respondents considered the task to be very difficult.

Table 6.2: Self-reported task difficulty from quantitative pilot study

	V1: 3% cancer risk	V2: 1% cancer risk	V3: Randomised attribute order	V4: Timing-survival relationship
Very easy	-	1 (4%)	1 (4%)	-
Easy	9 (36%)	8 (32%)	6 (24%)	5 (20%)
Neither easy nor difficult	9 (36%)	8 (32%)	8 (32%)	10 (40%)
Difficult	7 (28%)	8 (32%)	5 (20%)	10 (40%)
Very difficult	-	-	5 (20%)	-

ii. Frequency of attention and rationality check failures

In total, two respondents failed the embedded attention check questions. These respondents were excluded and replaced within the final analysis. A summary of the rationality check failures is shown in Table 6.3. Failures were generally low across all versions. The most common failure across all versions was decision-making based on a single attribute (dominant attribute test), specifically respondents who failed based their choices on the alternative with the highest accuracy level in all cases.

Table 6.3: Number of respondents failing rationality checks during the pilot study

	V1: 3% cancer risk	V2: 1% cancer risk	V3: Randomised attribute order	V4: Timing-survival relationship
Monotonicity	0 (0%)	0 (0%)	1 (4%)	0 (0%)
Transitivity	0 (0%)	0 (0%)	1 (4%)	0 (0%)
Stability	0 (0%)	0 (0%)	2 (8%)	1 (4%)
Non-trading	2 (8%)	3 (12%)	5 (20%)	2 (8%)
Flatlining	0 (0%)	0 (0%)	0 (0%)	0 (0%)

iii. MNL results

Results of multinomial logit analysis for each pilot version can be found in Table 6.4. Coefficients generally followed the expected direction of preferences, although results from version 3 (randomised attribute order version) suggested higher utility was associated fair rather than good communication. Despite the small sample size, almost all parameters were significantly significant.

Table 6.4: MNL results from the quantitative pilot study

	V1: 3% cancer risk	**V2: 1% cancer risk**	**V3: Randomised attribute order**	**V4: Timing-survival relationship**
	Coefficient (95%CI)	**Coefficient (95%CI)**	**Coefficient (95%CI)**	**Coefficient (95%CI)**
Accuracy				
Per 1%	0.06*** (0.04 **–** 0.08**)**	0.06*** (0.04 **–** 0.08**)**	0.09*** (0.07 **–** 0.11**)**	0.09*** (0.06 **– 0.11)**
Time to diagnosis				
Per month	-0.25*** (-0.42 – [-0.09])	-0.31*** (-0.48 – [-0.14])	-0.60*** (-0.82 – [-0.37])	-0.52*** (-0.77 – [-0.27])
Identifiable conditions				
Cancer only	Ref	Ref	Ref	Ref
Cancer plus additional related conditions	0.67*** (0.36 – 0.98)	0.41** (0.02 – 0.81)	-0.02 (-0.47 – 0.43)	0.55** (0.11 – 0.99)
Communication				
Poor	Ref	Ref	Ref	Ref
Fair	0.93*** (0.42 – 1.43)	0.57*** (0.16 – 0.97)	0.89*** (0.55 – 1.24)	0.84*** (0.32 – 1.36)
Good	1.34*** (0.90 – 1.78)	0.67*** (0.23 – 1.10)	0.53** (0.03 – 1.03)	1.05*** (0.49 – 1.61)
Neither test	-1.99** (-3.74 – [-0.23])	-1.05 (-2.43 – 0.33)	-1.89*** (-3.19 – [-0.58])	-1.08 (-2.60 – 0.43)
ASC	-0.10 (-0.29 – 0.10)	0.16* (-0.03 – 0.35)	0.05 (-0.21 – 0.30)	-0.14 (-0.39 – 0.11)
Log-likelihood	-163.58	-203.79	-151.39	-161.90
Key: ***significant at 99% confidence level; **significant at 95% confidence level; *significant at 90% confidence level				

iv. Selection of the opt-out and indifference alternatives

The indifference alternative was selected 55 times out of the 1,200 across all versions, representing 4.5% of total choices: 26% of respondents selected the alternative on at least one occasion. The 'neither test' alternative was chosen slightly more frequently (59/1200; 4.9%) but fewer respondents utilised the option (21/100) indicating serial selection was more likely (Table 6.5).

Table 6.5: Selection of indifference and opt-out alternatives during the pilot study

	V1: 3% cancer risk	**V2: 1% cancer risk**	**V3: Randomised attribute order**	**V4: Timing-survival relationship**
Indifference alternative	14/300 (7 individuals)	15/300 (8 individuals)	12/300 (6 individuals)	14/300 (5 individuals)
Neither alternative	5/300 (2 individuals)	19/300 (4 individuals)	13/300 (6 individuals)	22/300 (9 individuals)

v. Self-reported non-attendance

Self-reported attribute attendance was variable across the survey versions. Accuracy was consistently well-attended in all versions of the survey (Table 6.6). Self-reported non-attendance for all other attributes ranged from 24-68% across the survey versions, suggesting non-attendance should be an important consideration within the full version of the survey.

Table 6.6: Self-reported attribute non-attendance during quantitative piloting

	Version 1	Version 2	Version 3	Version 4
Accuracy	1 (4%)	4 (16%)	3 (12%)	2 (8%)
Time to diagnosis	11 (44%)	6 (24%)	10 (40%)	8 (32%)
Identifiable conditions	9 (36%)	6 (24%)	13 (52%)	11 (44%)
Communication throughout the testing process	14 (56%)	12 (48%)	17 (68%)	12 (48%)

6.3.6 Stage 6: Refinement

A summary of all the changes made following piloting are summarised in Table 6.7.

Table 6.7: Summary of changes throughout the piloting process

	Before piloting	Changes after qualitative piloting	Final instrument
Defining the research question			
Research questions	1. What is the relative importance of attributes relating to investigative testing for ovarian cancer in primary care? 2. How does the underlying level of cancer risk impact preferences towards testing? 3. How does the inclusion of an indifference alternative impact responses within a DCE?	1. What is the relative importance of attributes relating to investigative testing for ovarian cancer in primary care? 2. How does the underlying level of cancer risk impact preferences towards testing? 3. How does the inclusion of an indifference alternative impact responses within a DCE? 4. How does the order of attributes affect attribute non-attendance?	1. What is the relative importance of attributes relating to investigative testing for ovarian cancer in primary care? 2. How does the underlying level of cancer risk impact preferences towards testing? 3. How does the inclusion of an indifference alternative impact responses within a DCE? 4. How does the relationship between time to diagnosis and survival impact preferences towards testing? 5. To what extent is attribute non-attendance an issue? How does accounting for attribute non-attendance impact model estimates? 6. How does the number of choice tasks per respondent impact responses? Is there evidence of a learning or fatigue effect?
Survey versions	1% risk of cancer, indifference 3% risk of cancer, indifference	1% risk of cancer, indifference 3% risk of cancer, indifference	1% risk of cancer, indifference 2% risk of cancer, indifference

		3% risk of cancer, attribute order randomised 3% risk of cancer, timing-diagnosis relationship	3% risk of cancer, indifference 3% risk of cancer, no indifference 3% risk of cancer, timing-diagnosis relationship
Attributes and levels			
Accuracy	**Wording**: *Accuracy- If 100 people with ovarian cancer had this test, the test would: correctly identify x (x%) with cancer but miss x (x%) cases* **Levels:** (65%, 75%, 85%, 95%) **Presentation**: Visual aid	**Wording:** unchanged **Levels:** unchanged **Presentation:** unchanged	**Wording:** unchanged **Levels:** unchanged **Presentation**: unchanged
Time to diagnosis	**Wording:** *Time to completion of testing* **Levels:** 2 weeks, 4 weeks, 6 weeks, 8 weeks **Presentation:** Words only Timing-survival relationship: not-specified	**Wording:** *Time to diagnosis* **Levels:** 1 month, 2 months, 3 months, 4 months **Presentation:** words only Timing-survival relationship: no impact (v1-3) 1 week delay = 1% reduction in survival (v4)	**Wording:** *Time to diagnosis* **Levels:** 1 month, 2 months, 3 months, 4 months **Presentation:** words only (v1-4), visual aid (v5) Timing-survival relationship: no impact (v1-v4), age-stratified survival reduction per additional month (v5)
Identifiable conditions	**Wording:** *Identifiable conditions* **Levels:** ovarian cancer only, ovarian cancer plus alternative conditions related to your symptoms, ovarian cancer plus additional unrelated conditions **Presentation:** words only	**Wording:** unchanged **Levels:** ovarian cancer only, ovarian cancer plus alternative conditions related to your symptoms **Presentation:** unchanged	**Wording:** unchanged **Levels:** ovarian cancer only, ovarian cancer plus alternative conditions related to your symptoms **Presentation:** unchanged
Communication throughout the testing process	**Wording:** *Communication throughout the testing process* **Levels:** Good, fair, poor **Presentation:** words only	**Wording:** unchanged **Levels:** unchanged **Presentation:** unchanged	**Wording:** unchanged **Levels:** unchanged **Presentation:** unchanged

DCE composition			
Construction of choice tasks	Full or partial profiles: Full Test alternatives: Test A, Test B (generic) Opt-out alternative: Yes Indifference alternative: Yes (1 version with no indifference)	Full or partial profiles: unchanged Test alternatives: unchanged Opt-out alternative: unchanged Indifference alternative: All versions included indifference	Full or partial profiles: unchanged Test alternatives: unchanged Opt-out alternative: unchanged Indifference alternative: Yes, except v4
Experimental design	Fractional factorial design Main effects only D-efficient design using very small priors 12 choice tasks, no blocking 4 rationality check questions	Updated to reflect change in *identifiable conditions* levels	Fractional factorial design Main effects only Bayesian d-efficient design using priors from pilot study Joint estimation of versions (1,3,5) Optimised for MNL and evaluated for MMNL 16 choice tasks, no blocking 4 rationality check questions
Instrument design			
Survey design	See section 6.3.4	Additional questions: self-reported non-attendance, reasons for non-attendance Question changes: Choice certainty question reworded Other changes: Choice certainty question format changed from slider to radio	Additional questions: reasons for selection of indifference (if selected at least once) Question changes: none Other changes: attribute order varied between respondents but kept constant for each respondent

6.3.6.1 Re-defining the research question

Following piloting, the primary research question remained unchanged—to understand women's preferences towards diagnostic testing for ovarian cancer and understand how these preferences vary according to sociodemographic characteristics. However, as a result of piloting, methodological research questions were refined, and new research aims were introduced.

Additional research aims for the final study are:

i. How does the underlying level of cancer risk impact preferences towards testing?
ii. How does the inclusion of an indifference alternative impact responses within a DCE?
iii. How does the relationship between time to diagnosis and survival impact preferences towards testing?
iv. To what extent is attribute non-attendance an issue? How does accounting for attribute non-attendance impact model estimates?
v. How does the number of choice tasks per respondent impact responses? Is there evidence of a learning or fatigue effect?

i. How does the underlying level of cancer risk impact preferences towards testing?

During piloting, preferences did not appear to significantly differ based on the risk of cancer and/or symptoms experienced. This research question has important implications for clinical practice and future guideline revision. On this basis, alternative versions with differing risk levels were included in the final study. In addition to the 1% and 3% cancer risk versions piloted within this chapter, a 2% cancer risk version was included for completeness. Current urgent referral guidelines are specified on the basis of symptoms suggestive of a 3% cancer risk. Future changes are likely to be incremental, therefore when thinking about improvements to earlier diagnosis reducing this threshold to 2% appears to be a logical step- understanding preferences (and potential demand) at these different incremental risk levels is likely to be useful to policymakers. This approach also follows previous studies (Banks *et al.*, 2014).

ii. How does the inclusion of an indifference alternative impact responses within a DCE?

Exclusion of indifference alternatives may be detrimental to the quality of responses by forcing respondents to decide between test alternatives artificially. In response, all versions of the survey included an indifference alternative. To investigate the drivers of indifference, an additional debriefing question was added following the completion of all DCE questions for those who selected the indifference alternative at least once. Finally, to understand how indifference may influence results an additional version which excluded the indifference alternative added the final study. This additional version is reflective of current practice in healthcare DCEs, which do not typically include an indifference alternative (Chapter 4).

iii. How does the relationship between time to diagnosis and survival impact preferences towards testing?

From the pilot study it was unclear whether differences in the relationship between timing and survival were influential to preferences. However, cognitive interviews clearly demonstrated that this information was important to participants when considering choices. Given the current uncertainty around the relationship between time to diagnosis and survival it was deemed necessary to maintain the two alternative versions; (i) no survival impact and (ii) longer waiting times reduce the chance of survival.

Following piloting, a study assessing the impact of diagnostic delays on survival in the context of COVID-19 was published (Sud *et al.*, 2020). The study estimated age-stratified, 10-year survival probabilities associated with increased delays, specific to individual cancer sites, including ovarian cancer. For this thesis, the results from Sud *et al.* (2020) were extrapolated and used to update the DCE version with a time-survival relationship. Respondents were shown different 10-year survival risks according to their age (Table 6.8). Given the introduction of risk, a visual aid similar to the one used to describe the accuracy attribute was added.

Table 6.8: Age-stratified reduction in 10-year survival from ovarian cancer adapted from Sud et al (2020)

		Age			
		40-49	50-59	60-69	70+
Time to diagnosis (months)	1	0%	0%	0%	0%
	2	3%	4%	7%	8%
	3	9%	11%	12%	12%
	4	14%	18%	18%	17%

iv. To what extent is attribute non-attendance an issue? How does accounting for attribute non-attendance impact model estimates?

Randomising the order that attributes appeared seemed to increase irrational responses, self-reported difficulty and crucially non-trading behaviour therefore, this version of the survey was removed from the final study.

Self-reported non-attendance responses confirmed findings from the think-aloud interviews—that non-trading behaviour would be an important issue in the final study. The removal and/or replacement of the highly non-attended attributes (communication, identifiable conditions) was discussed by the research team, however, all attributes were maintained based on their importance throughout the attribute development stages. However, given the high likelihood of attribute non-attendance by a subsection of the sample an additional research question exploring the extent and impact of attribute non-attendance was added to the final study.

v. How does the number of choice tasks per respondent impact responses? Is there evidence of a learning or fatigue effect?

During qualitative piloting, some participants appeared to exhibit learning behaviour, developing their knowledge and preferences as the choice tasks progressed. Contrastingly, other respondents appeared to become fatigued by the seemingly repetitive choice tasks. The optimal number of choice tasks within DCEs remains

unclear and there is also contrasting literature on whether respondents experience learning or fatigue effects as choice tasks increase (Campbell *et al.*, 2015).

Given the mixed evidence from this pilot study, as well as the DCE literature more widely, the study was adapted to incorporate an investigation of this issue in the context of preferences for diagnostic testing.

Accommodation of this additional research question involved no additional survey versions but did require manipulation of the experimental design and an extension from 12 to 16 choice tasks. Further explanation of the methods used are presented in Chapter 8.

To address all the research aims, there were five sub-versions of the DCE choice tasks embedded within the online survey in total. Surveys were identical in all other ways. Respondents were randomised to a version and could only complete a single version. An example of a choice task from each version is shown in Figures 6.9-6.13.

6.3.6.2 Experimental design

A Bayesian efficient experimental design incorporating the pilot changes was generated in Ngene. Estimates from the pilot study as priors. The aim was to generate a single experimental design to be used across all survey versions. To do so, a multiple model approach was taken during experimental design generation. All available and relevant data was utilised (1% cancer risk, 3% cancer risk and timing-survival pilot results) to maximise the efficiency of the final design across all survey versions to generate the most efficient design. The Ngene syntax for the final design is found in Appendix 6.7. A final design with 16 choice tasks was selected. The final design is provided in Appendix 6.8.

The final survey instrument can be found in Appendix 6.9.

Figure 6.9: Example of a choice task from DCE version 1: 3% risk of cancer

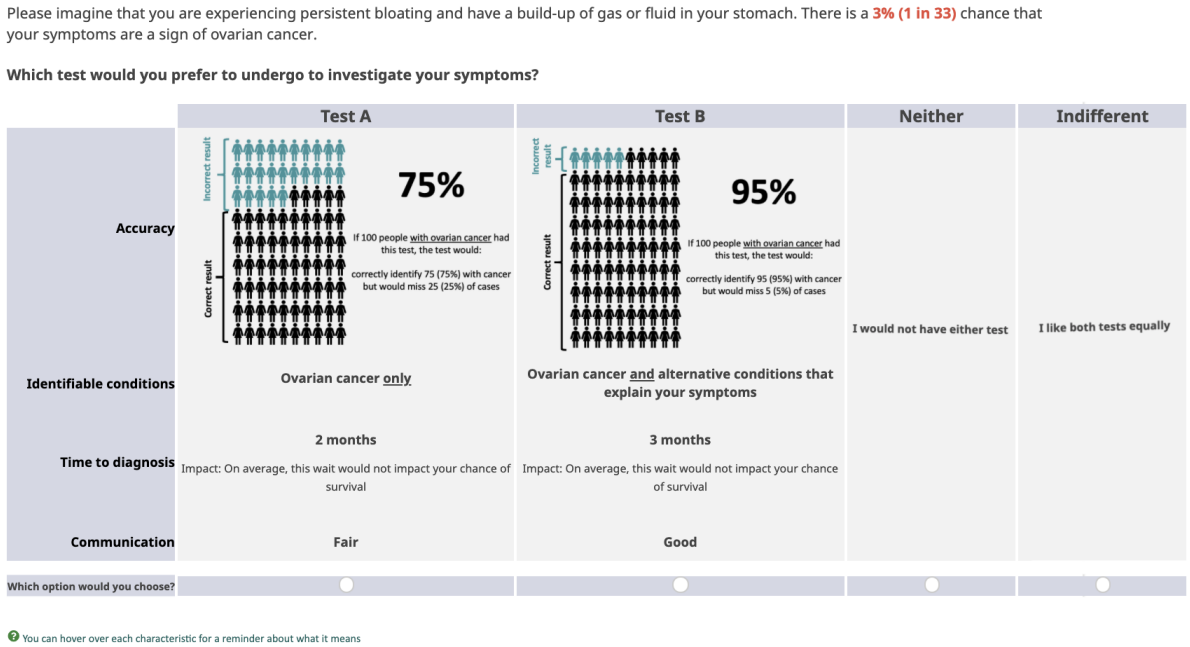


Figure 6.10: Example of a choice task from DCE version 2: 2% risk of cancer

Figure 6.11: Example of a choice task from DCE version 3: 1% risk of cancer

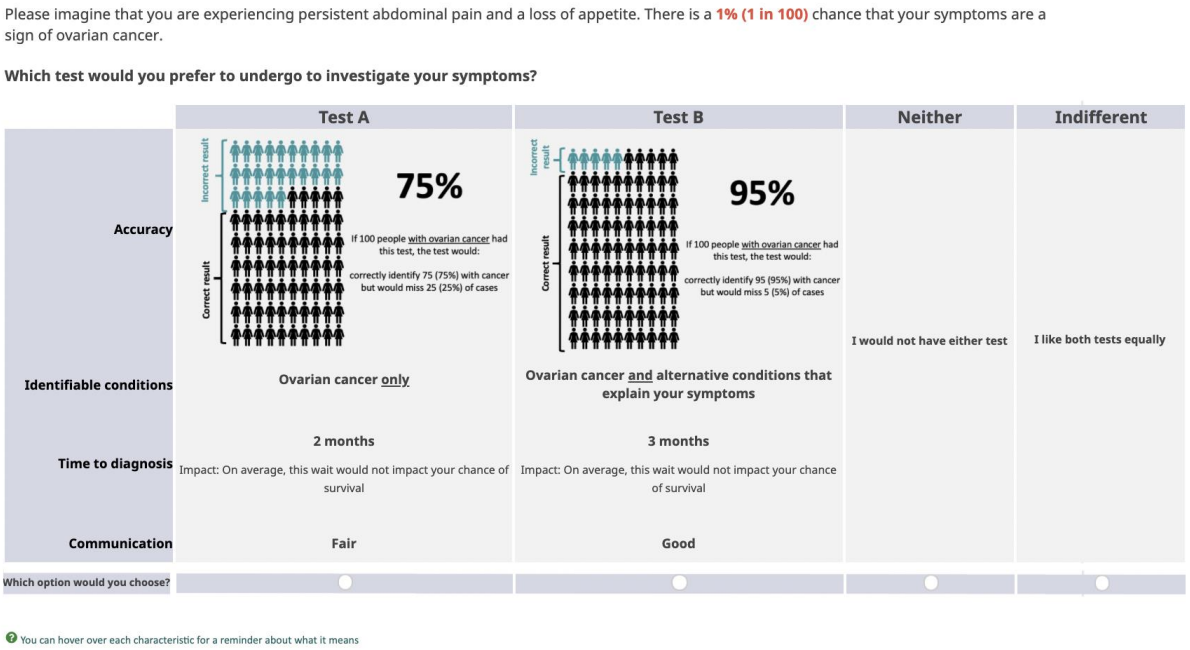
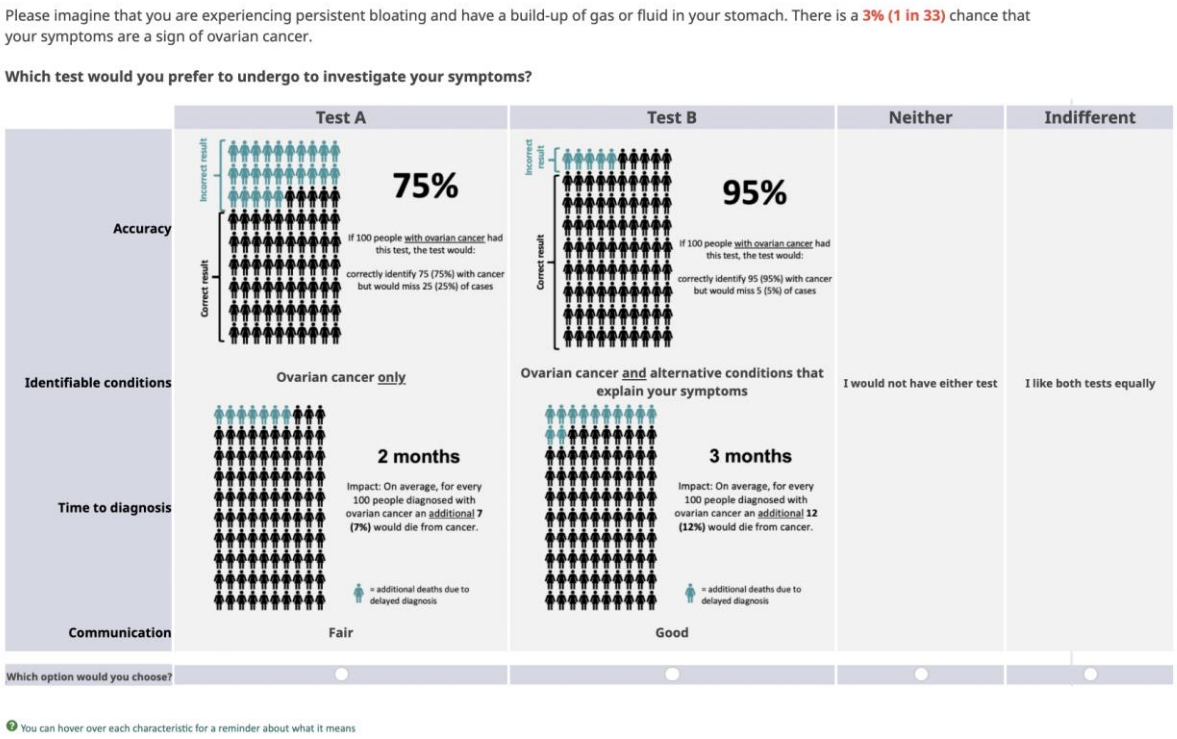


Figure 6.12: Example of a choice task from DCE version 4: 3% risk of cancer, no indifference alternative

Figure 6.13: Example of a choice task from DCE version 5: 3% risk of cancer, increased time-to-diagnosis has an age-adjusted negative affect on 10-year survival



6.4 Chapter summary

This chapter described the process of developing a DCE to investigate preferences towards diagnostic testing for ovarian cancer following best practice guidelines. The rigorous design process and transparent reporting aimed to maximise the quality and validity of subsequent results and are a strength of this chapter.

Two-stage piloting with a qualitative element allowed the survey instrument to be refined, ensuring it was well-understood and completed as intended by the target population. In particular, think-aloud interviews provided a valuable insight into how respondents engaged with the DCE tasks. Results motivated changes to many elements of the DCE design including attribute levels and wording and question formats. Quantitative survey piloting ensured the online format was acceptable to respondents and the survey length was appropriate. Results confirmed assumptions about the directionality of model parameters were correct. Pilot estimates were used to generate an updated Bayesian efficient experimental design for the final study.

In general, DCE instructions appeared to be well understood and respondents were willing and able to engage with the task of trading between attributes within choice sets. However, for a subsection of respondents, attribute non-attendance may be an issue.

The final DCE study on preferences towards diagnostic testing will include five versions of the survey. Each version will utilise the same experimental design and questionnaire but variations in the decision-context, available alternatives and choice task order will allow multiple additional methodological questions to be addressed. These additional questions were identified and refined during the piloting stage of the study. Additional research questions will improve the interpretability of responses within the final study and also add to the evidence and debate within the wider DCE literature.

7 Women's preferences towards testing for ovarian cancer in primary care: Results from a DCE

7.1 Introduction

This chapter describes the final data collection, analysis and results of a discrete choice experiment designed to elicit women's preferences towards diagnostic testing for ovarian cancer. The chapter utilises the survey instrument developed and piloted in the previous chapter. The chapter begins by outlining the research aims, followed by a description of the key methods including sample size calculations and data analysis plan. Finally, results are presented and interpreted in the context of ongoing debates regarding current inefficiencies and uncertainties surrounding investigative testing for ovarian cancer in primary. Research aiming to elicit preferences towards diagnostic testing for cancer is uncommon and results from this chapter are likely to be generalisable to diagnostic testing for cancer more broadly, particularly for sites where primary care-based tests are currently offered or may be offered in the future.

7.2 Aims and objectives

The primary aim of this chapter was to quantify preferences for diagnostic testing for ovarian cancer described in terms of key characteristics (attributes) and how preferences vary according to the level of cancer risk, as indicated by symptoms. Results from pilot testing (Chapter 6) resulted in adding a secondary research question aiming to investigate how current uncertainties in the relationship between the time to diagnosis and chance of survival impacts preferences for testing.

Given these research questions, this chapter sought to address the following objectives:

1. To measure the relative importance of key characteristics relating to ovarian cancer testing for people facing a 1%, 2% and 3% chance of cancer.
2. To examine the willingness to trade between attributes based on marginal rates of substitution at different cancer risk levels

3. To identify sociodemographic characteristics which may influence stated preferences for testing (i.e. explore preference heterogeneity)
4. To understand how uncertainties in the relationship between diagnostic interval and survival influence preferences and demand for ovarian cancer testing

7.3 Methods

7.3.1 Survey versions

Four versions of the survey were used to fulfil the research objectives (see Chapter 6 for more details). Surveys differed only in terms of the risk of cancer described to the respondents in the introductory vignette and the relationship between time to diagnosis and chance of survival. Respondents completed one version of the questionnaire assigned at random following electronic consent. The four versions were:

a. Version 1: Respondents were told there is a **1%** chance their symptoms are caused by ovarian cancer and time to diagnosis has **no impact** survival
b. Version 2: Respondents were told there is a **2%** chance their symptoms are caused by ovarian cancer and time to diagnosis has **no impact** survival
c. Version 3: Respondents were told there is a **3%** chance their symptoms are caused by ovarian cancer and time to diagnosis has **no impact** survival
d. Version 4: Respondents were told there is a **3%** chance their symptoms are caused by ovarian cancer and time to diagnosis has an **incremental age-adjusted impact** on survival

7.3.2 Study population

Participation was limited to women over the age of 40 with one or both ovaries based on current tasting guideline and risk of developing ovarian cancer (NICE, 2015). A general public population (i.e. women with mixed experience of cancer testing and recruited in a non-clinical seetiing) was chosen since the focus was on exploring the preferences of people who may be offered testing in primary care in the future.

7.3.3 Sample size

Sample size requirements for DCEs are a subject of debate. A number of different methods exist, with no accepted best-practice approach (de Bekker-Grob *et al.*, 2015). This study applies the s-estimate approach developed by Rose and Bliemer (2013). The approach utilises the results from the pilot study to estimate the minimum sample size for each attribute estimated as part of the utility function.

The minimum sample size N, for attribute k to achieve parameter estimates with 95% certainty is estimated following the equation below:

$$N_k \geq \left(\frac{1.96 \cdot \ se_1(\bar{\beta}_k)}{\bar{\beta}_k} \right)^2$$

Where $\bar{\beta}_k$ is the parameter estimate from the pilot study and $se_1(\bar{\beta}_k)$ is the associated standard error. The use of a d-optimal experimental design results in some parameters being estimated with much higher levels of reliability (i.e. lower standard errors) than others, since efficiency is globally optimised for the whole utility function. As a result, the minimum sample size will be the lower bound for finding a statistically significant parameter estimate for the attribute with the highest standard error in the pilot study.

Based on this approach, the minimum sample size for the final experimental design was 78 per survey version. This represents the theoretical lower bound for statistically significant parameters at the 95% confidence level, assuming that pilot estimates were completely accurate. A final larger sample size of 150 respondents per survey version was decided based to accommodate the uncertainty around pilot results and to allow for variability in sociodemographic covariates used to assess drivers of preference heterogeneity.

7.3.4 Recruitment

An online questionnaire was developed using the survey platform, Limesurvey (limesurvey.org) as described in Chapter 6. Following the successful pilot study, Prolific (Prolific.co) was used to recruit participants for the study. Participants were screened based on sex and age initially, with a follow up question designed to identify people without ovaries who were ineligible for inclusion. Participants were paid a

completion fee of £2.50 (based on a target average hourly rate of £7) directly into their prolific account. Participants who were interested in taking part were then given further information about the study and completed an online consent form prior to the beginning of any study questions.

7.3.5 Data analysis plan: survey questions

Anonymised data were downloaded and cleaned in SPSS v27 and data analysis was performed using Stata 17.

7.3.5.1 Respondent characteristics

Respondent characteristics were summarised using descriptive statistics to provide an understanding of the sample. ANOVA tests were used to identify any differences between the respondents across the different survey versions.

7.3.5.2 Qualitative analysis- communication

The survey included an open-ended question to understand what “good communication” meant to respondents. Answers to this question were coded into common themes and visually presented.

7.3.5.3 Ranking results

Respondents were asked to rank the attributes within the survey in terms of importance from 1(most important) to 4 (least important). Results were analysed to understand the frequency of ranking position for each attribute and the overall order of ranking in each survey version.

7.3.6 Data analysis plan: stated choice data

7.3.6.1 Preferences based on cancer risk

The first stage of analysis focused on understanding preferences towards ovarian cancer testing for women facing different risks of cancer. This stage related to the

analysis of data from survey versions 1-3 and addressed objectives i-iii. As described in Chapter 6, each choice task included 4 alternatives; test A, test B, opt-out and indifference. Indifferent responses were removed from analysis in this chapter to align with current practice in healthcare DCE literature, where indifference alternatives are not routinely provided (Soekhai *et al.*, 2019). The implications of including or excluding indifference alternatives within DCEs is explored further in Chapter 8.

Initially, main effects multinomial logit (MNL) models were estimated for each version of the DCE separately following the utility function:

$$V = \alpha_B + \alpha_{Opt-out} + \beta_1 \text{Accuracy75} + \beta_2 \text{Accuracy85} + \beta_3 \text{Accuracy95} + \beta_4 \text{Timing2} + \beta_5 \text{Timing3} + \beta_6 \text{Timing4} + \beta_7 \text{RelatedConditions} + \beta_8 \text{CommunicationFair} + \beta_9 \text{CommunicationGood}$$

Relative utility weights for each level are represented by beta coefficients, $\beta_1 \text{to} \beta_9$ in the utility function. The opt-out alternative was assumed to indicate a respondent would not undergo any testing. To incorporate the opt-out alternative included within each choice set, an alternative-specific constant (ASC), $\alpha_{Opt-out}$ was included. This constant term captures any systematic differences between the average effect of unobserved factors on testing and non-testing utility (i.e. opt-out effects). An additional ASC, α_B was included to correct for any left-right bias in respondents' choices.

To investigate the functional form, all attributes were initially assumed to be categorical. Attributes were dummy-coded with the lowest level for each attribute acting as the reference case. Following estimation of the categorical model the specification of timing and accuracy attributes were investigated by checking for between-level linearity using visual checks of plotted coefficients.

The multinomial logit model provides a simple starting point for understanding the structure of preferences within DCEs and also benefits from requiring the smallest sample size of the models utilised within this thesis. However, a major limitation of the model is estimated parameters refer to the average preferences of the population and assume preferences are homogenous across all respondents. In response, after finalising the functional form of the utility function, a mixed logit (ML) model was

estimated to capture preference heterogeneity. Mixed logit models were estimated based on simulations using 1000 Halton draws and all attributes included as random parameters to account for unobserved variation in respondents' preferences. All attributes were assumed to follow a normal distribution. This meant two parameters were estimated relating to each attribute; the mean and standard deviation of the distribution allowing the degree of preference heterogeneity surrounding each attribute parameter to be understood. Alternative model specifications were compared using likelihood-ratio tests.

Further contextual information of the models used to analyse DCE data within this chapter is found in Chapter 2.

Interpretation of findings

Ex-post calculations using the β- coefficients from the choice models were used to aid interpretation of results and allow for easy comparison of preferences for testing at different cancer risk levels. Two measures were calculated: (a) Relative importance scores for each attribute; and (b) Marginal rates of substitution (MRS).

Relative importance scores

Relative importance scores consider how much an attribute contributes to the overall utility of a test relative to the other attributes included in the DCE. The steps to calculate attribute importance are as follows:

1. Calculate the attribute utility range (i.e. utility of biggest level estimate – utility of smallest level estimate)
2. Sum the utility ranges of all attributes
3. Divide each attribute utility range by the total attribute utility range of all attributes

Importance scores can range between 0 and 1, with a higher score indicating greater importance. Estimates are ratio-scaled, meaning comparison is straightforward (e.g.

an importance score of 0.2 indicates an attribute is twice as important as an attribute with an importance score of 0.1).

The scores were calculated using mixed logit coefficients from each survey version. To control for uncertainty, final scores were the average of a simulation of 1000 Monte Carlo draws assuming β-coefficients followed a normal distribution. Confidence intervals were calculated based on 1.96 x standard error for each attribute importance score.

Marginal rate of substitution (MRS)

MRS represents how much more of one attribute respondents are willing to sacrifice in exchange for an improvement in another attribute. ***Time to diagnosis*** was chosen as the most appropriate numeraire, meaning estimates could be calculated by taking a ratio of the marginal utilities of each attribute/level and the parameter estimate for ***time to diagnosis***. Results are interpreted as the number of additional months respondents are willing to wait for a diagnosis in exchange for a given improvement in the associated attribute. For example, $\frac{\beta_8 \text{CommunicationFair}}{\beta_4 \text{Timing}}$ represents the additional length of time (in months) an individual would wait to receive their diagnosis in exchange for an improvement in communication from poor (dummy-coded base level) to fair. MRS was calculated based on mixed logit coefficients. Confidence intervals were generated using the Delta method in Stata v17.

Investigating differences in preferences based on the level of cancer risk

To investigate differences in preferences between cancer risk levels (1%, 2%, 3%) a pooled MNL model which included interaction terms for the risk levels was estimated. Statistically significant interaction terms would indicate differing preferences towards ovarian cancer testing could not be ruled out. In contrast, insignificant interaction terms suggest preferences do not differ across the included risk levels.

Sensitivity checks- response quality and validity

Sensitivity checks were performed to assess the quality and validity of data. Firstly, the models were re-estimated with respondents who completed the survey in less than

10 minutes excluded. This lower time limit was specified based on the average completion time during piloting.

Secondly, data were reanalysed with those who failed one or more of the rationality checks omitted. Estimates from sensitivity checks were compared to MNL results which included the full sample. Next, a logit model was estimated to investigate factors associated with the chance of failing validity questions and included explanatory variables such as completion time, self-reported task difficulty, education and choice certainty.

Finally, subgroup analysis was performed to check for differences in MNL estimates between respondents based on the self-reported task difficulty.

Heterogeneity in preferences for testing

i. Subgroup analysis

Subgroup analyses were performed to investigate the potential influence of specific sociodemographic characteristic of preferences for ovarian cancer testing. Subgroup analyses were based on MNL models and compared on the basis of MRS estimates (willingness to wait) using T-tests. MRS is a ratio calculation therefore potential differences in error variance between subgroups will be mitigated. Included subgroups were informed by the literature (Chapter 4) and are described in Table 7.1. Sociodemographic characteristics were informed by the systematic review described in Chapter 4. To control for the increased risk of type I error caused by performing multiple subgroup analyses. In total 12 subgroup tests were performed. Therefore, the critical value for statistical significance during t-tests was increased to 0.005 (i.e. 0.05/11) (Wang *et al.*, 2021).

ii. Latent class analysis

A latent class logit model (Chapter 2) was used to further explore heterogeneity in preferences. The number of classes was determined by comparing models including between 2-6 classes and involved balancing goodness-of-fit based on model statistics (Akaike information criterion (AIC), Bayesian information criterion (BIC) and log-likelihood) and ability to interpret the findings meaningfully. The sociodemographic characteristics described in Table 7.1 were also included to test the influence of individual characteristics on probability of class membership.

Opt-out behaviour

Analysis of opt-out behaviour was performed to help identify factors that may prevent people from undergoing testing for ovarian cancer. Opt-out behaviour was modelled using a logistic regression where opting-out at least once was the dependent variable. Sociodemographic characteristics such as those in Table 7.1 were considered as explanatory variables alongside task-feedback factors such as task difficulty.

Table 7.1: Sociodemographic characteristics used to investigate preference heterogeneity

Category	Subgroups
Age	• 40-49yrs • 50-59yrs • 60+yrs
Ethnicity	• White • Non-white
Education	• Attended university • Did not attend university
Self-reported health	• Good health • Average/below good health
Ovarian cancer testing experience	• Previously tested • Never tested
Know someone diagnosed with ovarian cancer	• Yes • No
Worried about ovarian cancer	• Yes • No
Current medical decision-making role	• Active (Report having “a great deal” or “a lot” of input in medical decisions) • Passive (Report having a lesser role in medical decisions)
Desired medical decision-making role	• Active (Report wanting “a great deal” or “a lot” of input in medical decisions) • Passive (Report wanting a lesser role in medical decisions)
Confidence in ability to recognise symptoms of ovarian cancer	• Low • High
Task difficulty	• Very easy/easy • Neither difficult or easy • Very difficult/difficult

7.3.6.2 Impact of diagnostic delays on survival

To investigate how a potential relationship between ***time to diagnosis*** and survival may impact preferences, responses were compared across survey versions 3 and 4. In version 3, respondents were asked to assume ***time to diagnosis*** had no impact on the chance of survival, whereas in version 4 chance of survival was linked to the ***time to diagnosis*** and age of the respondent, as described in Chapter 6. As before, results were compared on the basis of attribute relative importance scores and willingness to wait estimates.

Demand for testing

To further understand how the incorporation of survival might impact demand for different tests, choice shares for alternative test profiles were calculated. Analysis centred around whether a less accurate but quicker triage test is acceptable to patients. Specifically, choice share analysis compared demand for alternative test profiles with demand for the CA125 blood test, the current first-line test for people with vague symptoms of ovarian cancer in primary care (NICE, 2015).

Transvaginal ultrasound (TVUS) or combined testing using concurrent TVUS and CA125 tests are two alternative approaches currently utilised by other countries and readily available potential alternatives to the sequential testing using the CA125, currently recommended by NICE (2015). Accuracy of the TVUS is currently unknown in a diagnostic primary care setting but is assumed to be greater than CA125 with screening trial results suggesting ~85% sensitivity in a non-symptomatic population (Menon *et al.*, 2009). TVUS also benefits from broader diagnostic capabilities and is able to identify alternative conditions that may be the underlying cause of symptoms where cancer is not the cause (e.g. fibroids, cysts). However, waiting times for ultrasound scans are longer than a blood test performed in GP surgeries.

The CA125 blood test was represented by a profile assuming a 77% (95% CI: (72.8–80.8%) accuracy rate, ability to detect cancer only, 1-month diagnostic interval and fair communication. A 1 month diagnostic interval is an optimistic estimate based on the Faster Diagnostic Standard target (NHS England, 2016). Current diagnostic intervals are often longer meaning estimates represent the best-case scenario for the CA125 test (Funston *et al.*, 2020a; Lim *et al.*, 2016). Given the differences in test-performance and survival, subgroup analysis of people under 50 years old was performed using an alternative CA125 profile which assumed a lower accuracy level of 62.5% (95% CI:51.0-73.1%) (Funston *et al.*, 2020a). Since the accuracy of transvaginal ultrasound is unknown and waiting times are variable, scenario analysis compared demand between CA125 and tests with 85% or 95% accuracy. The new test was also able to identify alternative conditions. Communication was held constant (fair) across both tests.

Demand for competing test profiles in each scenario were estimated using the share of preference method (Hensher *et al.*, 2005c):

- Calculate the utility for each test profiles and no-test alternative you wish to compare by calculating the sum of the β-coefficients of the associated attribute levels
- Exponentiate the total utility of each test
- Divide each exponentiated test utility by the sum of all the test profiles exponentiated utilities to give the share of demand

Confidence intervals for estimates were calculated using the delta method.

7.4 Results

In total, 610 women completed the DCE, responses from ten participants were removed due to failing the attention check questions (n=6) or stating indifference for all choice tasks (n=4) leaving a final sample size of 600 (150 respondents per version). Due to the nature of recruitment (open advertisement until sample size was fulfilled), response rates are not reported. In total, 451 (4.7% if all responses) indifference responses were received and 169 (28%) respondents indicated they were indifferent to the test alternatives in at least one task, these responses were coded as missing during the analysis relating to the current results.

7.4.1 Participant characteristics

Sociodemographic characteristics

Table 7.2 shows key respondent characteristics across the four versions of the survey. No significant differences were found between respondents across the different survey versions.

Medical knowledge, attitude and behaviours

Respondents typically reported having little-to-moderate input in current medical decisions taking place during primary care consultations, however, the majority

desired an increased role in decisions (“Desired level of input in medical decisions”) (Table 7.3).

Respondents were generally help-seeking, with approximately 75% of people indicating they would seek advice within 1 month of the onset perceived symptoms of ovarian cancer. However, the ability to recognise symptoms was a clear barrier as just 11% of respondents reported feeling confident in their ability to recognise symptoms and familiarity with the five key symptoms ranged from 26% (loss of appetite) to 57% (constant bloating).

Further responses relating to health history and attitude are summarised in Appendix 7.1. Responses did not differ significantly between survey versions, with exception of knowing someone diagnosed with ovarian cancer which was significantly higher for respondents completing survey version 4.

Table 7.2: Sociodemographic characteristics of respondents across the 4 survey versions

	Version 1	Version 2	Version 3	Version 4:	p-value*
Age					
Mean (SD)	50.6 (8.3)	51.9 (8.6)	51.4 (7.7)	51.9(10.2)	0.54
Range	40-73	40-75	40-70	40-81	
Ethnicity, n (%)					
White	140	143	140	133 (89%)	0.81
Other	(93%)	(95%)	(93%)	15 (10%)	
Prefer not to say	10 (7%)	6 (4%)	10 (7%)	2 (1%)	
	-	1 (1%)	-		
Children, n (%)					
Mean (SD)	1.62 (1.2)	1.6 (1.2)	1.32 (1.2)	1.65 (1.3)	0.48
Range	0-5	0-5	0-3	0-7	
Relationship status, n (%)					
Single	12 (8%)	23 (15%)	24 (16%)	21 (14%)	0.79
In a relationship	34 (23%)	21 (14%)	32 (21%)	26 (17%)	
Married/civil partnership	81 (54%)	88 (59%)	69 (46%)	78 (52%)	
Separated/divorce	19 (13%)	14 (9%)	15 (10%)	14 (9%)	
Widowed	3 (2%)	3 (2%)	9 (6%)	8 (5%)	
Prefer not to say	1 (1%)	1 (1%)	1 (1%)	3 (2%)	
Education, n (%)					
No qualifications	0 (0%)	1 (1%)	4 (3%)	2 (1%)	0.47
GCSE	31 (21%)	34 (23%)	35 (23%)	37 (25%)	
A-Level/ College	20 (13%)	20 (13%)	23 (15%)	32 (21%)	
Undergraduate	57 (38%)	47 (31%)	48 (32%)	41 (27%)	
Post-graduate/professional quals	33 (22%)	43 (28%)	38 (25%)	36 (24%)	
Other	9 (6%)	3 (2%)	4 (3%)	1 (1%)	
Prefer not to say	-	2 (1%)	-	1 (1%)	
Employment, n (%)					
Employed, full-time	55 (37%)	46 (31%)	55 (37%)	40 (27%)	0.36
Part-time	31 (21%)	27 (18%)	30 (20%)	31 (21%)	
Self-employed	18 (12%)	23 (15%)	22 (15%)	25 (17%)	
Not employed	8 (5%)	11 (7%)	12 (11%)	6 (4%)	
Retired	12 (8%)	21 (14%)	9 (6%)	25 (17%)	
Other	26 (17%)	21 (14%)	17 (8%)	22 (15%)	
Prefer not to say	-	1 (1%)	5 (3%)	1 (1%)	
Household income, n (%)					
£0-9,999	5 (3%)	6 (4%)	3 (2%)	5 (3%)	0.47
£10,000-19,999	20 (13%)	28 (19%)	24 (16%)	28 (19%)	
£20,000-29,999	27 (18%)	28 (19%)	23 (15%)	23 (15%)	
£30,000-39,999	21 (14%)	25 (17%)	26 (17%)	21 (14%)	
£40,000- 49,999	27 (18%)	18 (12%)	17 (11%)	13 (9%)	
£50,000- 59,999	9 (6%)	7 (5%)	20 (13%)	19 (13%)	
£60,000-69,999	8 (5%)	11 (7%)	3 (2%)	9 (6%)	
£70,000+	22 (15%)	16 (11%)	18 (12%)	12 (8%)	
Prefer not to say	11 (7%)	11 (7%)	16 (11%)	20 (13%)	

*p-value from ANOVA analysis to investigate differences in marginal rates of substitution between survey versions

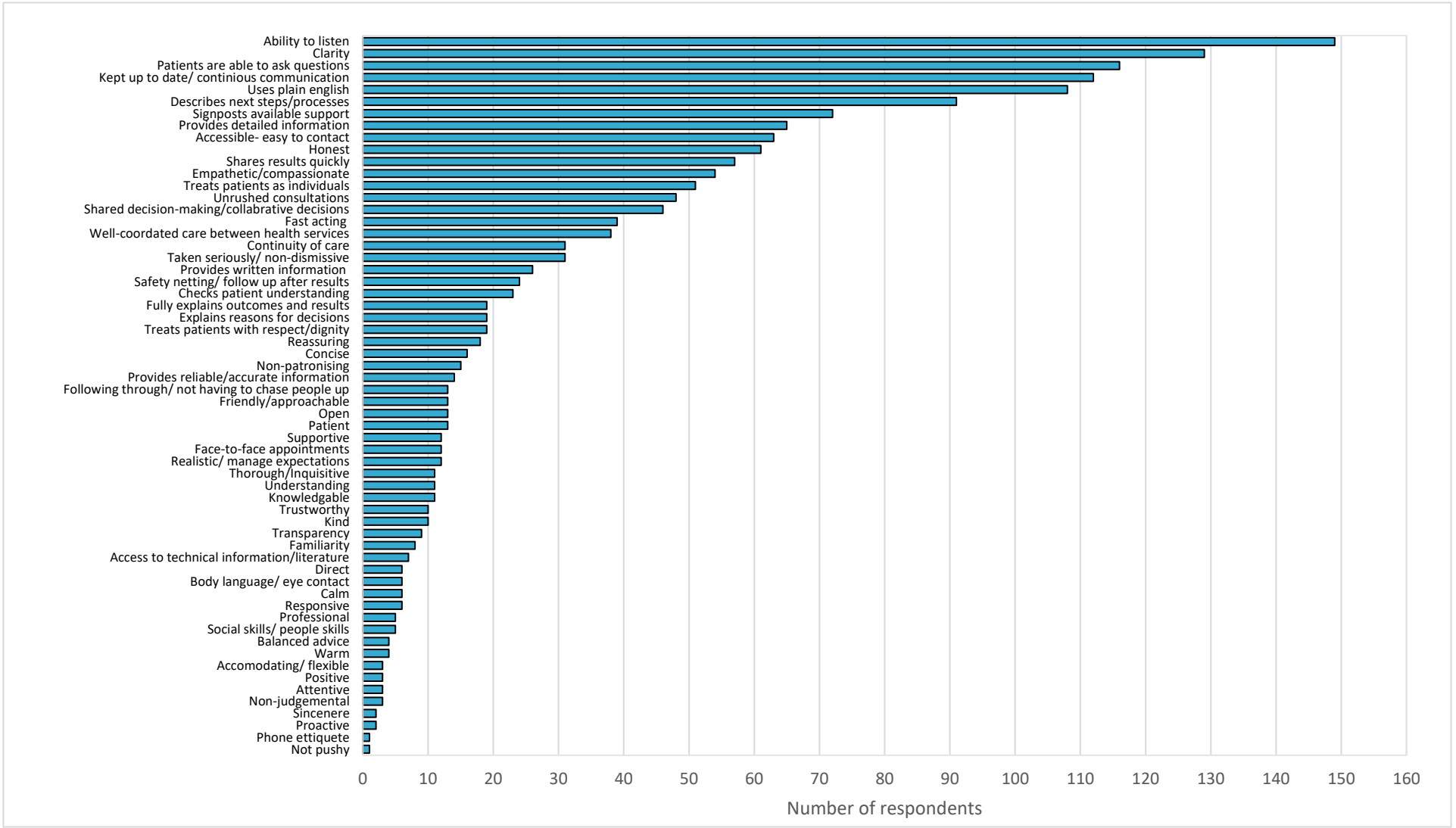
Table 7.3: Responses to selected medical knowledge, attitude and behaviour questions across the 4 survey versions

	Version 1	Version 2	Version 3	Version 4	p-value*
Confidence to recognise OC symptoms, n (%)					
1-Not at all	33 (22%)	38 (25%)	40 (27%)	44 (29%)	0.45
2	78 (52%)	72 (48%)	65 (43%)	67 (45%)	
3	20 (13%)	30 (20%)	22 (15%)	28 (19%)	
4	19 (13%)	10 (7%)	23 (15%)	11 (7%)	
5-Extremely confident	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Symptom recognition (ability to recognise main 5 symptoms), n (%)					
Constant bloating	91 (61%)	87 (58%)	76 (51%)	85 (57%)	0.35
Swollen abdomen	81 (54%)	93 (62%)	79 (53%)	87 (58%)	0.36
Discomfort in the pelvis/abdomen	59 (39%)	62 (41%)	62 (41%)	58 (39%)	0.95
Loss of appetite/ feeling full quickly	39 (26%)	41 (27%)	35 (23%)	40 (27%)	0.87
Increased urinary urgency/ frequency	48 (32%)	45 (30%)	41 (27%)	43 (29%)	0.84
Length of time before consulting GP about OC symptoms, n (%)					0.75
Immediately, no wait	19 (13%)	22 (14%)	20 (13%)	17 (11%)	
Up to 1 week	25 (17%)	28 (19%)	27 (18%)	33 (22%)	
1-2 weeks	31 (21%)	29 (19%)	34 (23%)	30 (20%)	
2-4 weeks	28 (19%)	36 (24%)	38 (25%)	32 (21%)	
More than a month	47 (31%)	35 (23%)	38 (25%)	38 (25%)	
Current level of input in medical decisions, n (%)					
A great deal	15 (10%)	13 (9%)	12 (8%)	17 (11%)	0.63
A lot	28 (19%)	33 (22%)	22 (15%)	22 (15%)	
A moderate amount	57 (38%)	54 (36%)	61 (41%)	54 (36%)	
A little	43 (29%)	38 (25%)	43 (29%)	45 (30%)	
Not at all	7 (5%)	12 (8%)	12 (8%)	12 (8%)	
Desired level of input in medical decisions, n (%)					
A great deal	53 (35%)	64 (43%)	67 (45%)	49 (33%)	0.63
A lot	59 (39%)	52 (35%)	42 (28%)	66 (44%)	
A moderate amount	28 (19%)	28 (19%)	32 (21%)	27 (18%)	
A little	10 (7%)	5 (3%)	7 (5%)	8 (5%)	
Not at all	0 (0%)	1 (1%)	2 (1%)	0 (0%)	
*p-value from ANOVA analysis to investigate differences between survey versions					

7.4.2 What is good communication?

Responses to the open-ended question: *“What does "good" communication from a doctor or healthcare provider mean to you?”* were grouped thematically as shown in Figure 7.1. Responses revealed people most associated a doctor who listens well (149/600), describes things clearly (129/600) and allows the opportunity for questions to be asked and answered (116/600) with “good communication” during medical investigations. Many qualities are dependent on the intrapersonal skills of doctors or constraints of the medical system, however, some easily implementable improvements which were highly valued by participants were also identified including providing written information following verbal communication (4%; 24/600) and signposting available resources such as support groups and charities (11%; 65/600).

Figure 7.1: Coded responses to the open-ended question: "What does "good" communication from a doctor or healthcare provider mean to you?



7.4.3 Ranking

Figure 7.2 shows the frequency of ranking across survey versions and Figure 7.3 shows the overall ranking for each attribute. Accuracy was ranked most important across all survey versions overall and communication was ranked consistently lowest. Timing and identifiable conditions scored similarly in terms of importance and varied between 2nd and 3rd place across survey versions.

Figure 7.2: Frequency of ranking for attributes across survey versions 1-4

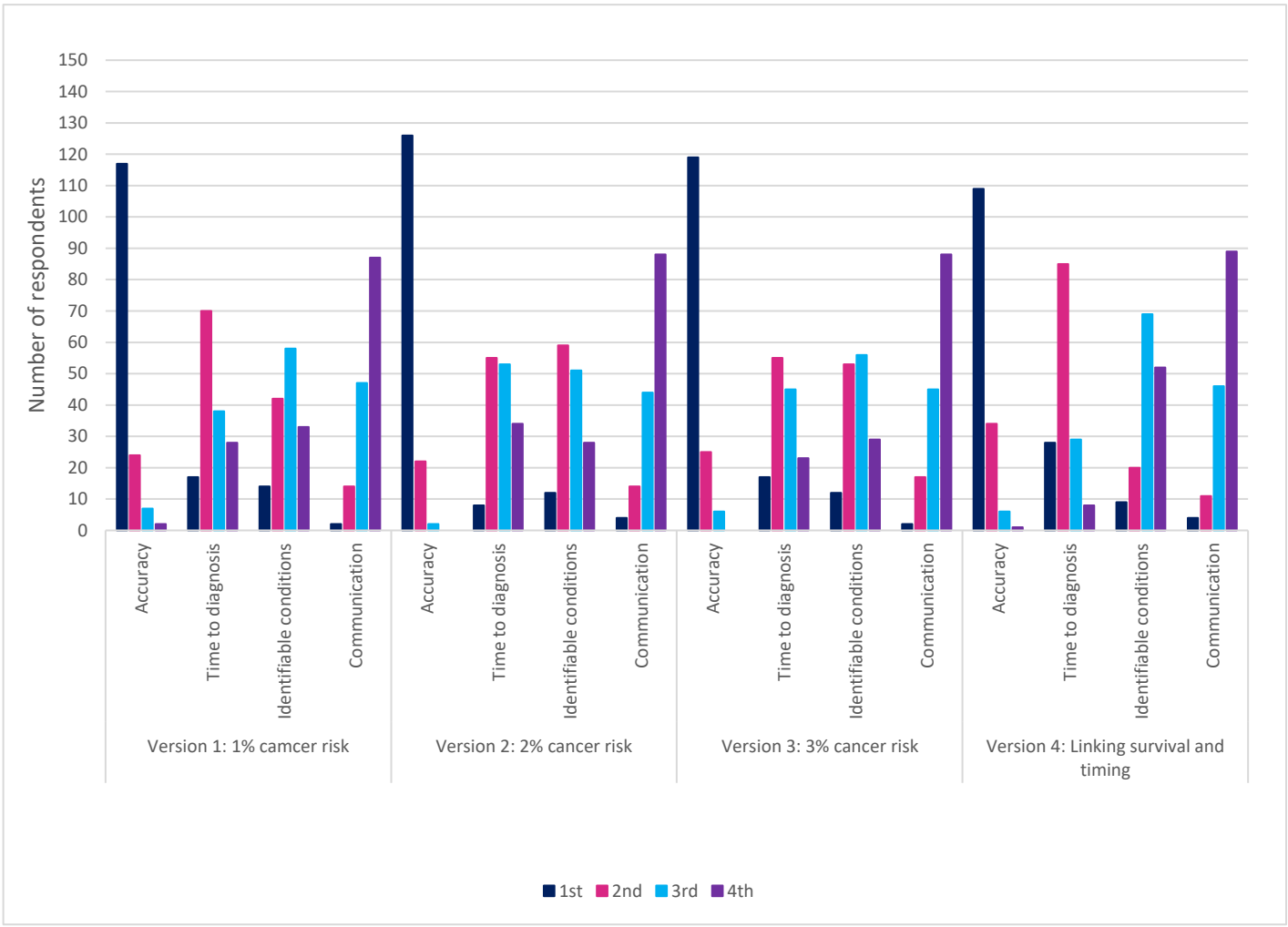
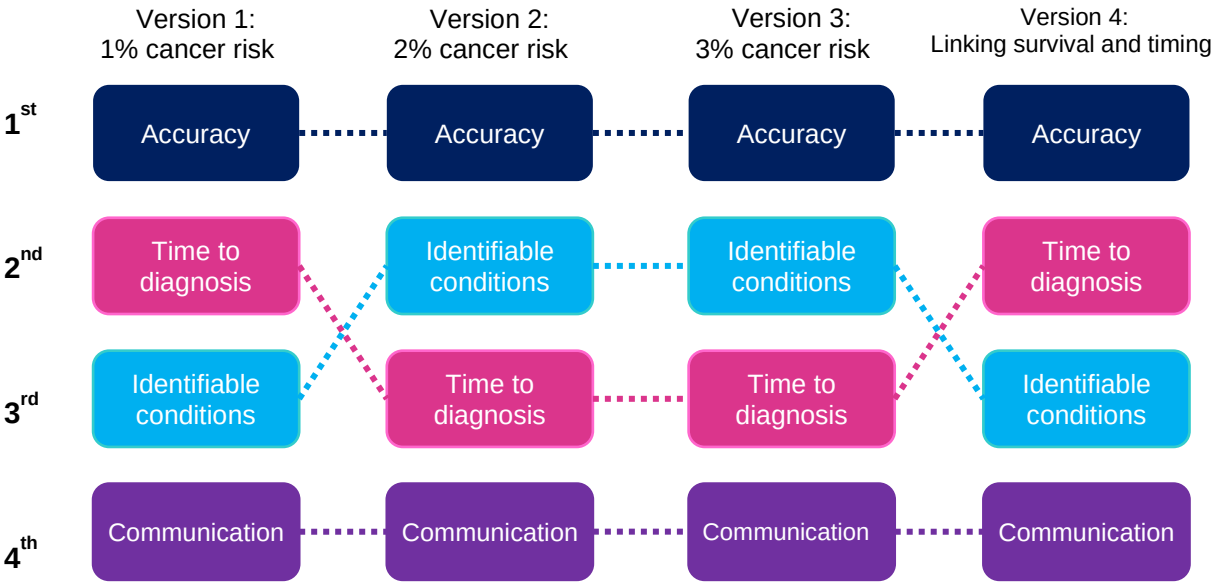


Figure 7.3: Overall ranking of the attributes across survey versions 1-4



7.4.4 Analysis of choice data

7.4.4.1 Preferences towards ovarian cancer testing at different risk levels

Multinomial and mixed logit results

All attribute levels were initially categorically coded using dummy variables to assess the functional form (Appendix 7.2). Visual inspection of accuracy and time to diagnosis level parameters suggested both attributes followed a linear pattern (Appendix 7.3). Both attributes were therefore modelled as continuous linear attributes in all subsequent models.

Table 7.4 shows the results of the individual MNL and ML models for 1%, 2% and 3% risk of cancer. All coefficients were statistically significant and the direction of preferences was as expected. The negative coefficient for the “neither test” ASC demonstrates that in general respondents had a strong preference for testing over no testing. Improvements in communication lead to an increase in utility relative to the base level (poor) in each instance. Similarly, the ability to detect alternative conditions alongside ovarian cancer was preferred to a test that could detect ovarian cancer only. The time to diagnosis coefficient represents the change in utility for each additional

month between consultation and diagnosis and as expected was negative. Increased accuracy was positively associated with utility, with the parameter estimate interpreted as the increase in utility associated with a 1% increase in accuracy.

Data from the three versions were pooled to investigate differences in preferences based on the symptom-based risk of cancer (Table 7.5). Model fit of the pooled model was assessed using a likelihood-ratio (LR) test comparing the pooled log-likelihood with the sum of the likelihoods from the three separate models and results suggested pooling data did not reduce model fit (-24.24, chi2 p-value= 0.86). All parameters associated with level of cancer risk were insignificant in the pooled interaction MNL model (Appendix 7.4) providing evidence that preferences for testing do not appear to vary according to the underlying chance of symptoms being the result of cancer.

The mixed logit models demonstrated significant preference heterogeneity across respondents. For example, in the pooled ML model (Table 7.4) preferences varied significantly across individuals for all parameters with the exception of fair communication (as signalled by the non-significant distribution standard deviation). Most notably, the estimated parameter mean would suggest that respondents have a strong preference for testing as indicated by the large negative no test ASC. However, considering the magnitude of the standard deviation shows that 19% of respondents favour no testing[8]. The large shift in the opt-out coefficient between the MNL and ML specifications further highlights the presence of heterogeneity in opt-out preferences. Model results indicate there may be skewed data, specifically participant generally appear to exhibit a strong preference towards testing with a much smaller proportion of people having a very strong prerfence against testing. Exploration of alternative distributions when modelling opt-out behaviour is an area for future development.

Overall, the importance of capturing such heterogeneity is demonstrated by the model performance with likelihood-ratio tests demonstrating the mixed logit model significantly outperformed MNL model across every survey version. Subsequent ex-post calculations were therefore based on parameter estimates from the ML models.

[8] Based on a z-value of 0.886 (3.76/4.24)

Table 7.4: MNL and ML results for survey versions 1-3

	Version1: 1% risk of cancer				Version 2: 2% risk of cancer				Version 3: 3% risk of cancer			
	MNL		ML		MNL		ML		MNL		ML	
	Coeff	Std error	Mean (SE)	Std dev	Coeff	Std error	Mean (SE)	Std dev	Coeff	Std error	Mean (SE)	Std dev
Accuracy												
Per %	0.09***	0.01	0.16*** (0.01)	0.08***	0.10***	0.01	0.23*** (0.02)	0.14***	0.10***	0.01	0.21*** (0.02)	0.12
Timing												
Per month	-0.37***	0.04	-0.65*** (0.07)	0.58***	-0.34***	0.03	-0.75*** (0.08)	0.57***	-0.41***	0.04	-0.78*** (0.08)	0.71***
Identifiable conditions												
Cancer only	Ref	-	Ref	-	Ref	-	Ref	-	Ref	-	Ref	-
Cancer plus additional related conditions	0.77***	0.07	1.18*** (0.11)	0.82***	0.76***	0.07	1.17*** (0.14)	1.08***	0.71***	0.09	1.21*** (0.14)	1.20***
Communication												
Poor	Ref	-	Ref	-	Ref	-	Ref	-	Ref	-	Ref	-
Fair	0.74***	0.08	1.13*** (0.12)	0.54***	0.65***	0.07	1.11*** (0.12)	0.17	0.69***	0.08	1.22*** (0.13)	0.35
Good	0.83***	0.10	1.24*** (0.14)	0.72***	0.94***	0.10	1.48*** (0.15)	0.79***	0.91***	0.09	1.54*** (0.15)	0.67***
Neither test	-0.58**	0.27	-5.02*** (0.76)	6.99***	-0.52**	0.26	-5.00*** (0.96)	4.79***	-0.63**	0.26	-3.29*** (0.54)	3.40***
Model fit statistics												
LL	-1529.13		-1224.43		-1451.86		-1086.53		-1462.00		-1101.34	
LR test (ML vs MNL)	-		609.40***		-		728.94***		-		721.32***	
Observations	6,840		6,840		6,918		6,918		6,840		6,840	
N	150		150		150		150		150		150	
Key: ***significant at 99% confidence level; **significant at 95% confidence level; *significant at 90% confidence level												

Table 7.5: MNL and ML results for pooled data (versions 1-3)

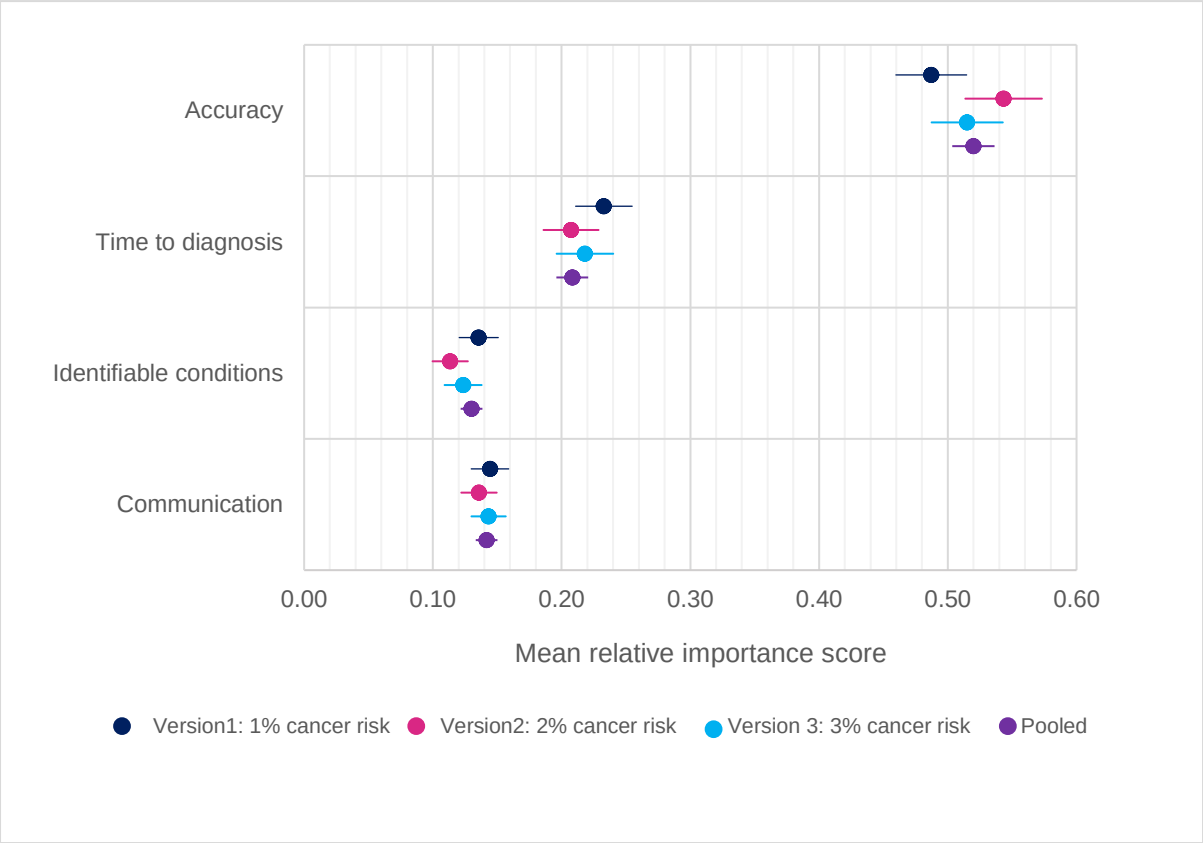
	Pooled-MNL			Pooled-ML		
	Coeff	**Std error**	**MRS (95% CI)**	**Mean (SE)**	**SD**	**MRS (95% CI)**
Accuracy						
Per %	0.10***	0.00	0.26 (0.23–0.29)	0.19*** (0.05)	0.11***	0.28 (0.25–0.31)
Timing						
Per month	-0.37***	0.02	-	-0.68*** (0.04)	0.62***	-
Identifiable conditions						
Cancer only	Ref	-	-	Ref	-	-
Cancer plus additional related conditions	0.74***	0.04	2.02 (1.69–2.35)	1.15*** (0.08)	0.96***	1.70 (1.45–1.96)
Communication						
Poor	Ref	-	-	Ref	-	-
Fair	0.70***	0.04	1.89 (1.59–2.19)	1.13*** (0.07)	0.23	1.67 (1.44–1.90)
Good	0.89***	0.06	2.41 (2.02–2.80)	1.37*** (0.08)	0.78***	2.02 (1.74–2.30)
Neither test	-0.58***	0.15	-	-3.76*** (0.40)	4.24***	-
Model fit statistics						
LL	-4455.11			-3432.17		
LR test (ML vs MNL)	-			2045.86***		
Observations	20,598			20,598		
N	450			450		
Key: ***significant at 99% confidence level; **significant at 95% confidence level; *significant at 90% confidence level						

Interpreting findings

i. Relative importance scores

Relative importance scores for each risk level and the pooled model are shown in Figure 7.4. No evidence of significant differences in estimates between survey versions was found, further demonstrating the high level of concordance between preferences for testing at different risk levels. In all instances, accuracy was deemed most important with a score of approximately 0.5. Time to diagnosis was considered second most important across all models with scores around 0.2. Communication and identifiable conditions were estimated to be of similar performance with scores of around 0.14 and 0.13, respectively.

Figure 7.4: Relative importance scores of attributes across survey versions 1-3



ii. Marginal rates of substitution

Marginal rates of substitution using time to diagnosis as the numeraire are shown in Table 7.6. Estimates represent the number of additional months women would be willing to wait for a diagnosis in exchange for an improvement in each attribute as compared to the base level. For example, based on the pooled ML model, for a 10% improvement in accuracy, women would be prepared to wait an extra 2.8 months (Table 7.5). Alternatively, women would be prepared to wait an extra 1.7 months to receive fair communication as opposed to poor communication throughout the testing process, however, the distinction between fair and good communication is much less important and women would only wait an addition 0.35 months for a further improvement.

Table 7.6: Marginal rates of substitution for survey versions 1-3 based on estimates from mixed logit models.

	Version 1: 1% cancer risk	Version 2: 2% cancer risk	Version 3: 3% cancer risk	p-value*
Accuracy				
Per 1%	0.25 (0.20–0.29)	0.30 (0.25–0.36)	0.29 (0.22–0.33)	0.36
Identifiable conditions				
Cancer only	-	-	-	-
Cancer plus additional related conditions	1.82 (1.38–2.27)	1.56 (1.15–1.97)	1.55 (1.16–1.95)	0.60
Communication				
Poor	-	-	-	-
Fair	1.75 (1.33–2.17)	1.48 (1.12–1.84)	1.57 (1.19–1.95)	0.62
Good	1.92 (1.42–2.41)	1.97 (1.51–2.43)	1.99 (1.53–2.44)	0.98
*p-value from ANOVA analysis to investigate differences in marginal rates of substitution between survey versions				

Sensitivity checks- evaluating the quality of responses

i. Response times

The total completion time ranged from 7-78 minutes, with a median time of 17 mins 33 secs. Sixteen respondents completed the survey in under 10 minutes. Sensitivity analysis found no differences in parameter estimates when excluding individuals in this time (Appendix 7.5).

ii. Rationality checks

Results from the rationality checks are shown in Table 7.7. Overall, failures of the monotonicity (6/450; 1%) and transitivity (8/450; 2%) checks were low. No respondents displayed flat-lining behaviour (i.e. consistently choosing an alternative in the same position). However, 23% (102/450) of respondents displayed unstable results throughout the choice experiment by changing responses across repeated choice tasks. Significant differences in model estimates were found when removing those who failed a rationality check from estimations (Appendix 7.6), specifically those who failed rationality checks exhibited a decreased emphasis on accuracy but overall estimates were comparable in terms of direction, significance and overall importance rankings. Logistic regression was used to identify the relationship between failing one or more rationality checks and respondent characteristics. Increased choice certainty was associated with a small but significant reduction in the chance of failing rationality checks and higher education approximately halved the likelihood of failing checks (OR=0.55, p= 0.01). However, completion time and self-reported task difficulty had no impact (Appendix 7.7).

iii. Task difficulty

The perceived difficulty of the task was diverse. Thirty-six percent of individuals rating the task as very easy or easy whilst almost an equal proportion reported finding the task very difficult or difficult (35%). Results from models stratified according to task

difficulty were similar in terms differences were found when comparing communication coefficients (Table 7.8).

In general, the direction and significance of model estimates was consistent across all sensitivity checks; therefore no responses were removed during the final estimation of models. Three respondents failed the unrelated attention check questions placed throughout the survey and were removed from the analysis.

Table 7.7: Failure of the rationality check questions across survey versions 1-3

	Version 1: 1% cancer risk	Version 2: 2% cancer risk	Version 3: 3% cancer risk	Overall
Monotonicity failures	0/150 (0%)	0/150 (2%)	3/150 (2%)	3/450 (<1%)
Transitivity failures	5/150 (3%)	3/150 (0%)	0/150 (0%)	8/450 (2%)
Stability failures	36/150 (24%)	36/150 (24%)	30/150 (20%)	102/450 (23%)
Flat-lining	0/150 (0%)	0/150 (0%)	0/150 (0%)	0/150 (0%)

Preference heterogeneity- investigating the role of sociodemographic characteristics

Evidence from the mixed logit models suggested significant heterogeneity in preference across respondents in the study. Heterogeneity in preferences according to sociodemographic characteristics was investigated in two ways: (i) subgroup analyses using MNL models; (ii) latent-class logit model. Stable preferences across cancer risk levels allowed the data from versions 1-3 to be pooled to maximise the sample size.

i. Subgroup analyses

The results of the subgroup analyses performed using MNL models are shown in Table 7.8. Subgroup analysis revealed very few differences willingness to wait in improvements in other attributes. However, people who knew someone diagnosed with cancer were willing to wait 0.56 months longer for a diagnosis in comparison with

did not know anyone with the disease in exchange for a test that could identify multiple conditions.

ii. Latent class model

Model statistics suggested a six-class model provided the best fit for data but would also result in a number of small classes (<5% of the population) (Appendix 7.8). Instead, a five-class model was selected based on a parsimonious approach, balancing the level of detail with the interpretability of results given the sample size. Results of the final model are shown in Table 7.9. The model utilised dummy-coding with all sociodemographic characteristics interpreted relative to the fifth class. Mean attribute importance scores were calculated based on model coefficients are shown in Figure 7.5. Test accuracy remained the most important attribute across all classes. However, variations in the order and magnitude of importance of the remaining attributes demonstrated differences in preferences and decision-making style across classes.

Class 1: Accuracy maximisers (45.2%)

Responses in this class are characterised by the overwhelming importance placed on test accuracy, which was considered over 50% more important than the second most prioritised attribute, time to diagnosis. The probability of belonging to this class was the highest estimated within the model at 45%. As a result, these results strongly reflect the sample-level results presented in the previous sections. Responses in this class were significantly more likely to be from paritcipants who identified as white. University attendance was weakly associated with class membership.

Class 2: Test sceptics (6.1%)

Class 2 had the lowest membership probability. Responses in this class were more likely to belong to women who selected the opt-out alternative during the experiment and demonstrated a strong preference against the tests presented within the choice tasks as demonstrated by the large opt-out ASC. The close range of importance of all attributes may indicate that all aspects of a test must meet a high standard for these individuals to be happy to undergo testing. Individuals in this class are more likely to

want an active role in medical decision-making. This may also explain the comparatively high level of importance placed on communication throughout the testing process within this class. This group were less likely to report being in good or very good health.

Class 3: Experience maximising traders (19.6%)

Despite remaining most important, accuracy appeared to be less of a concern for this group in comparison to other classes. Instead, this group valued communication and identifiable conditions higher than any other class. Oppositely, time to diagnosis was comparatively less prioritised by this class. This suggests this class value the overall service and experience of testing and were more willing to balance all aspects of testing. Class 3 was also associated with a large negative neither test ASC indicating a strong preference towards testing. There were no sociodemographic associations with membership for this class.

Class 4: Accuracy-conscious hesitant testers (10.3%)

Class 4 was the second smallest preference class, with a membership probability of 10.3%. Membership was associated with a large positive opt-out coefficient suggesting a tendency against testing. Accuracy was the most important attribute by a significant margin. The overall order of attribute importance followed the population-level findings. Sociodemographic associations with preferences were limited for this class with only increased confidence in ability to recognise symptoms being weakly associated with class membership.

Class 5: Time-conscious satisficisers (22.9%)

The probability of belonging to this class was the second highest identified within the model. This class demonstrated the largest preference towards testing, demonstrated by the large negative opt-out ASC. This group placed the largest importance on timing of all the classes by a significant margin. Increased importance of time to diagnosis, appeared to come as a result of sacrifices in terms of test accuracy, suggesting this class may take a satisficing approach to decision-making, a pragmatic approach, whereby individuals seek to obtain a satisfactory or “good enough” (according to a minimum standard) test available in in a reasonable time frame.

Opt-out behaviour

In total, the “neither test” alternative was chosen in 394 tasks (5%) by 127 respondents. Logistic regression found individuals who reported experiencing a moderate-high level of anxiety about ovarian cancer and those who reported being good or very good health were approximately 50% (OR=0.56) less likely to select the opt-out alternative. Compared to those who were not working, being employed also reduced the odds of opting for no testing but to a slightly lesser degree (OR=0.62). Despite the large number of sociodemographic characteristics hypothesised to impact opt-out behaviour, no other factors were found to be significant (Appendix 7.9).

Table 7.8: Results of subgroup analysis estimated using multinomial logit models and pooled data from survey versions 1-3. Results demonstrate the willingness to wait for a diagnosis in exchange for an improvement in the remaining attributes.

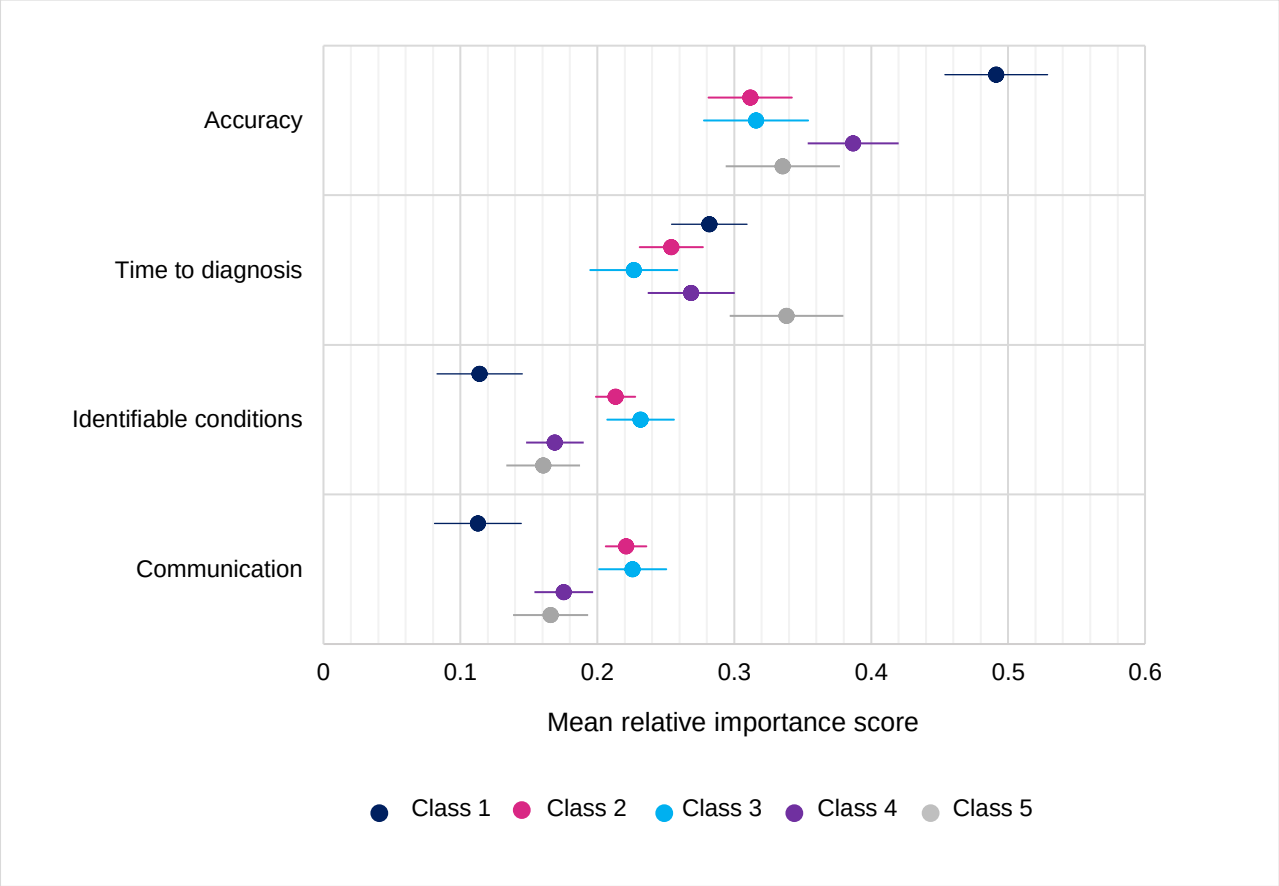
Subgroup	n	Accuracy	Identifiable conditions		Communication		
		Per %	Cancer only	Related conditions	Poor	Fair	Good
Age*							
40-49yrs	235	0.26*** (0.22–0.30)	-	2.14*** (1.65–2.63)	-	1.87*** (1.45–2.29)	2.27*** (1.74–2.79)
50-59yrs	150	0.25*** (0.20–0.30)	-	1.64*** (1.15–2.12)	-	1.68*** (1.23–2.12)	2.31*** (1.70–2.93)
60+	65	0.31*** (0.21–0.41)	-	2.60*** (1.50–3.71)	-	2.64*** (1.61–3.67)	3.33*** (1.91–4.75)
Ethnicity							
White	423	0.32*** (0.28–0.36)	-	2.76*** (2.34–3.17)	-	2.51*** (2.16–2.86)	3.31*** (2.83–3.78)
Non-white	27	0.21*** (0.12–0.30)	-	0.77*** (0.34–1.88)	-	1.65*** (0.52–2.77)	2.13*** (0.90–3.36)
University							
Yes	274	0.28*** (0.24–0.33)	-	1.93*** (1.52–2.33)	-	1.76*** (1.39–2.13)	2.16*** (1.68–2.63)
No	176	0.23*** (0.19–0.27)	-	2.09*** (1.55–2.65)	-	2.05*** (1.57–2.54)	2.71*** (2.06–3.35)
Self-reported health							
Good health	310	0.28*** (0.24–0.32)	-	2.06*** (1.63–2.50)	-	1.80*** (1.43–2.17)	2.27*** (1.79–2.75)
Average or below good	140	0.23*** (0.18–0.27)	-	1.86*** (1.33–2.38)	-	1.77*** (1.29–2.25)	2.52*** (1.86–3.18)
Previously tested for ovarian cancer							
Yes	34	0.20*** (0.12–0.29)	-	1.64*** (0.71–2.58)	-	1.85*** (0.83–2.88)	2.01*** (0.82–3.19)
No	416	0.27*** (0.24–0.30)	-	2.06*** (1.70–2.41)	-	1.90*** (1.59–2.21)	2.45*** (2.04–2.86)
Know someone diagnosed with ovarian cancer							
Yes	73	0.28*** (0.19–0.37)	-	2.49*** (1.45–3.53)	-	1.91*** (1.11–2.71)	2.45*** (1.41–3.49)
No	377	0.26*** (0.23–0.29)	-	1.93*** (1.58–2.28)	-	1.89*** (1.57–2.21)	2.40*** (1.98–2.82)
Current role in medical decision-making							
Active	123	0.28*** (0.24–0.31)	-	2.23*** (1.81–2.64)	-	1.97*** (1.62–2.31)	2.51*** (2.04–2.99)
Passive	327	0.23*** (0.18–0.28)	-	1.50*** (0.95–2.05)	-	1.72*** (1.14–2.29)	2.16*** (1.48–2.83)
Desired role in medical decision-making							
Active	337	0.25*** (0.19–0.31)	-	1.85*** (1.23–2.47)	-	1.74*** (1.16–2.31)	2.56*** (1.71–3.40)
Passive	113	0.27*** (0.23–0.30)	-	2.07*** (1.68–2.47)	-	1.92*** (1.57–2.26)	2.34*** (1.90–2.77)
Worried about ovarian cancer							
Yes	316	0.29*** (0.23–0.36)	-	2.18*** (1.48–2.88)	-	1.90*** (1.30–2.50)	2.21*** (1.49–2.93)
No	134	0.25*** (0.22–0.28)	-	1.96*** (1.58–2.34)	-	1.89*** (1.55–2.23)	2.48*** (2.02–2.95)
Confidence in ability to recognise symptoms of OC							
Low	325	0.28*** (0.19–0.38)	-	2.46*** (1.45–3.46)	-	2.55*** (1.50–3.60)	3.63*** (1.96–5.31)
High	125	0.26*** (0.23–0.29)	-	1.97*** (1.61–2.32)	-	1.81*** (1.50–2.12)	2.26*** (1.87–2.65)
Task difficulty *							
Very easy/easy	162	0.26*** (0.21–0.30)	-	1.78*** (1.28–2.28)	-	1.61*** (1.14–2.08)	2.11*** (1.53–2.69)
Neither easy or difficult	115	0.25*** (0.19–0.31)	-	1.94*** (1.30–2.59)	-	2.21*** (1.58–2.83)	2.83*** (1.98–3.67)
Very difficult/difficult	173	0.28*** (0.23–0.33)	-	2.29*** (1.69–2.90)	-	1.95*** (1.75–3.06)	2.42*** (1.75–3.06)

*Differences between subgroups was assessed using ANOVA.
Key: significant differences at 95% confidence level according to individual sample t-test

Table 7.9: Results of latent class logit regression used to assess preference heterogeneity using pooled data from survey versions 1-3

	Class 1		Class 2		Class 3		Class 4		Class 5	
Attributes	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**	**Coefficient**	**SE**
Accuracy										
Per %	0.21***	0.01	0.08***	0.01	0.05***	0.01	0.20***	0.01	0.10***	0.01
Time to diagnosis										
Per 1 month wait	-0.27***	0.07	-0.40***	0.08	-0.25***	0.04	-0.94***	0.09	-1.04***	0.08
Identifiable conditions										
Cancer only	-	-	-	-	-	-	-	-	-	-
Additional conditions	0.74***	0.07	1.27***	0.20	1.37***	0.10	0.82***	0.18	0.41***	0.12
Communication										
Poor	-	-	-	-	-	-	-	-	-	-
Fair	0.35***	0.11	1.56***	0.26	1.28***	0.11	1.62***	0.24	0.67***	0.13
Good	0.53***	0.11	2.44***	0.27	1.61***	0.13	1.90***	0.30	0.99***	0.16
Neither test	-1.71***	0.41	2.95***	0.36	-2.21***	0.38	2.24***	0.38	-4.21***	0.43
Class probability model										
White	1.96***	0.65	0.13	0.86	0.93	0.68	-0.15	0.59	-	-
University	0.54*	0.30	-0.41	0.49	-0.44	0.36	0.55	0.44	-	-
Good health	0.12	0.32	-0.98**	0.49	0.10	0.38	-0.27	0.44	-	-
Confidence in ability to recognise symptoms	0.99	0.64	0.30	0.97	1.13	0.69	1.39*	0.74	-	-
Want active role in decision-making	0.31	0.32	1.33**	0.67	0.25	0.38	0.35	0.46	-	-
Constant	-1.70	0.74	-1.54	1.06	-0.94	0.79	-1.03	0.77	-	-
Class probabilities										
	45.2%		6.1%		19.6%		10.3%		18.9%	
Model fit statistics										
Log-likelihood	-3406.56									
AIC	6931.12									
BIC	7173.44									
CAIC	7137.80									
Key: ***significant at 99% confidence level; **significant at 95% confidence level; *significant at 90% confidence level										

Figure 7.5: Relative importance scores of attributes for the four latent classes identified during exploration of preference heterogeneity



7.4.4.2 How does incorporating survival impact preferences for ovarian cancer testing?

MNL and ML model coefficients for version 4 are shown in Table 7.10. Linking delays in diagnosis to a decrease in survival led to significant changes in preferences. The relative importance of time to diagnosis increased 65.3% from 0.22 to 0.36. Accuracy remained the most important attribute, however, relative importance was reduced from 0.52 to 0.44. Further reductions in the importance of identifiable conditions and communication came as a consequence of increased emphasis on time to diagnosis (Figure 7.6).

Willingness to wait estimates further highlight the shift in preferences. Willingness to wait for improvements to identifiable conditions and communication are capped at 1 month or less. Willingness to wait in exchange for improvements in accuracy was also

much lower, for instance, for a 10% improvement in accuracy women were now willing to wait 1.2 months (95% CI: 1.1-1.4 months) compared to 2.8 (95% CI: 2.5-3.1) months when survival is not affected (Table 7.6).

Demand for testing - scenario analysis

Results of the choice share scenario analysis are shown in Figure 7.7. The grey lines represent the share of demand for current testing procedures using the CA125 as the first line testing option. The pink line represents demand for a new alternative test aiming to explore the demand for first-line testing using TVUS. The blue line represents the proportion of people who would opt to have no test as the diagnostic interval associated with the alternative test increased. The shaded areas represent variation in demand based on the 95% confidence interval for CA125 sensitivity taken from Funston *et al.* (2020a). Where waiting does not impact survival (Figure 7.7.1 and 7.7.3), testing rates were high and most women expressed a strong preference towards a more accurate test even where diagnostic intervals were long. For a test with accuracy of 95% over 50% of women would rather wait 6 months for a diagnosis but receive a more accurate test rather than receive the less accurate, faster alternative. When asked to consider survival, demand for the more accurate test is initially strong, however, the quicker but less accurate CA125 test is preferred by the majority of women once the diagnostic interval of the new test extends beyond 2-3.5 months, respective to accuracy (Figure 7.7.2 and 7.7.4).

For respondents under 50, when assuming the diagnostic interval has no impact on survival, demand for an alternative, more accurate and broader test is strong. Even where diagnostic intervals reach 6 months, 65% (CI:48-75%) would prefer to wait for the alternative test with 85% accuracy and this proportion increases to 76% (CI: 65-83%) where the accuracy of the alternative test is 95% (Figure 7.8.1 and 7.8.3). The introduction of mortality associated within longer diagnostic intervals, increases the demand for a faster diagnosis, however, demand for the existing CA125 test only becomes dominant once waiting times for the alternative test exceed 4-5.5 months, dependent on the accuracy of the alternative test (Figure 7.8.2 and 7.8.4).

Table 7.10: Results of multinomial logit and mixed logit models used to analyse DCE version 4

	MNL			ML		
	Coeff	**Std error**	**MRS (95% CI)**	**Mean (SE)**	**SD**	**MRS (95% CI)**
Accuracy						
Per %	0.10***	0.01	0.12 (0.11-0.13)	0.15*** (0.09)	0.08***	0.12 (0.11-0.14)
Timing						
Per month	-0.80***	0.06	-	-1.28*** (0.08)	0.64***	-
Identifiable conditions						
Cancer only	Ref	-	-	Ref	-	-
Cancer plus additional related conditions	0.48***	0.07	0.60 (0.41-0.78)	0.60*** (0.10)	0.68***	0.47 (0.32-0.62)
Communication						
Poor	Ref	-	-	Ref	-	-
Fair	0.60***	0.08	0.75 (0.52-1.10)	0.88*** (0.11)	0.17	0.69 (0.52-0.85)
Good	0.88***	0.10	1.10 (0.85-1.35)	1.20*** (0.12)	0.17	0.94 (0.74-1.13)
Neither test	-1.01***	0.21	-	-3.60*** (0.48)	3.61***	-
Model fit statistics						
LL	-1503.66			-1211.98		
LR test (ML vs MNL)	-			583.38***		
Observations	7,053			7,053		
N	150			150		
Key: ***significant at 99% confidence level; **significant at 95% confidence level; *significant at 90% confidence level						

Figure 7.6: Relative importance scores from survey versions 3 and versions 4. Comparison of the importance of attribute when the time of diagnosis has differing impacts on the chance of survival

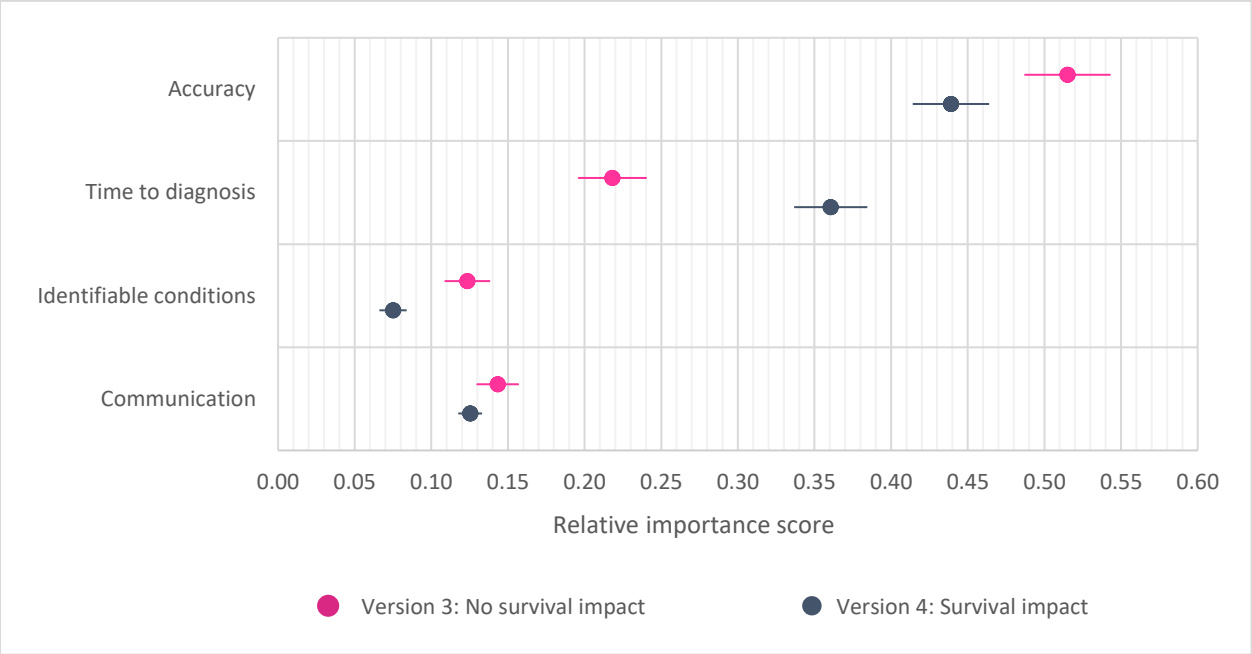


Figure 7.7: Share of demand analysis for the whole population based on estimates from DCE versions 3 and 4.
Graphs show the demand for current testing (CA125 in grey) compared to an alternative test which is more accurate but associated with longer waiting times.

Fig 7.7.1: Time to diagnosis has no impact on survival.
Comparator test in 85% accurate

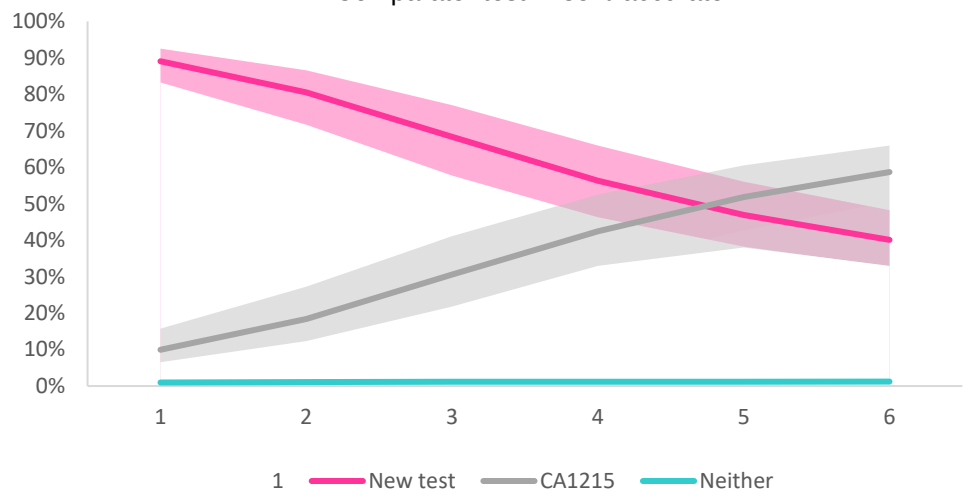


Fig 7.7.2: Time to diagnosis impacts survival.
Comparator test in 85% accurate

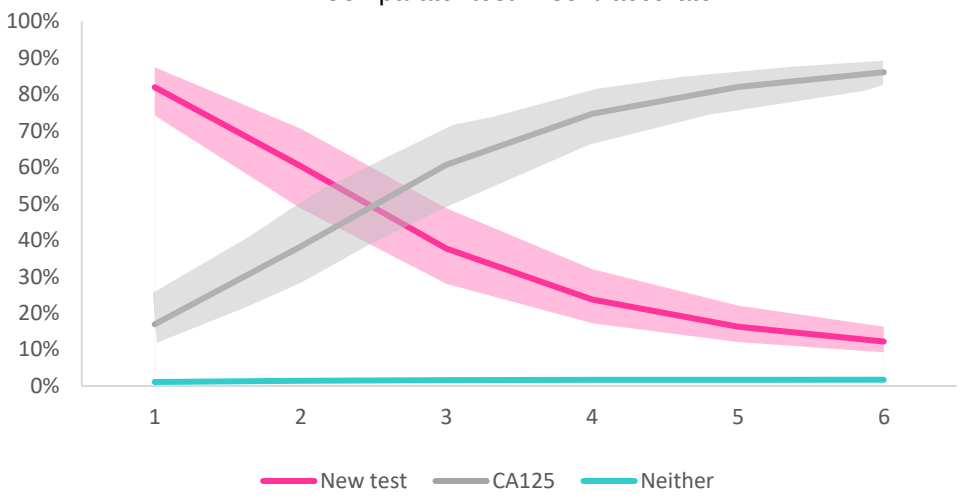


Fig 7.7.3: Time to diagnosis has no impact on survival.
Comparator test in 95% accurate

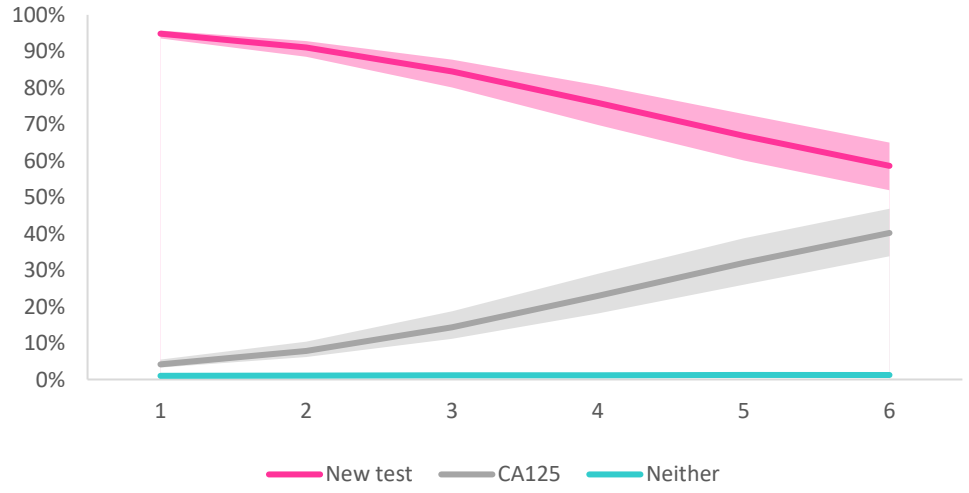


Fig 7. 7.4: Time to diagnosis impacts survival.
Comparator test in 95% accurate

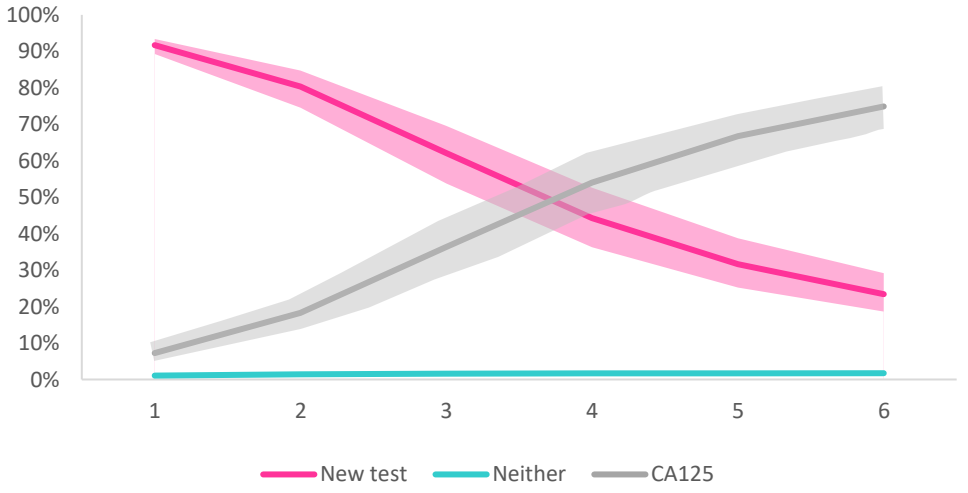


Figure 7.8: Under 50s subgroup share of demand analysis results based on estimates from survey versions 3 and 4.

Graphs show the demand for current testing (CA125 in grey) compared to an alternative test which is more accurate but associated with longer waiting times.

Fig 7.8.1: Time to diagnosis has no impact on survival.
Comparator test in 85% accurate

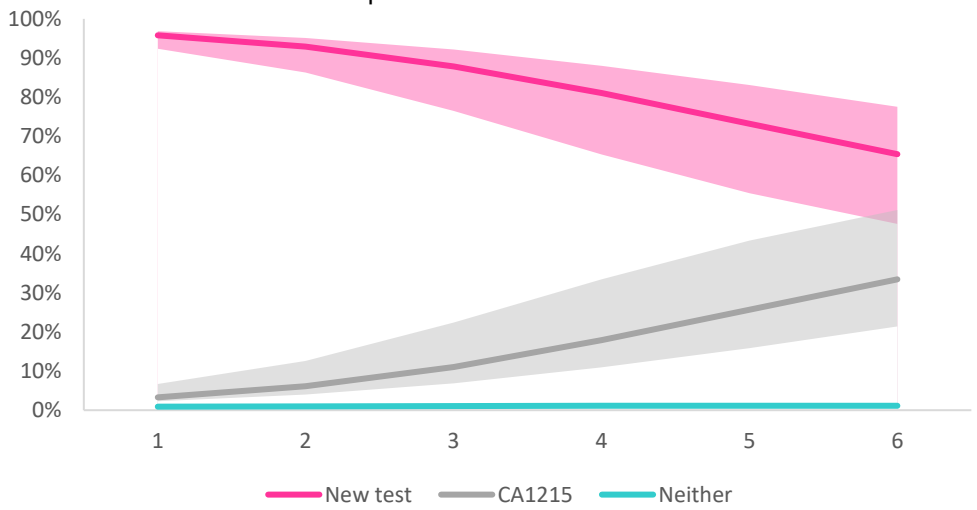


Fig 7.8.2: Time to diagnosis impacts survival.
Comparator test in 85% accurate

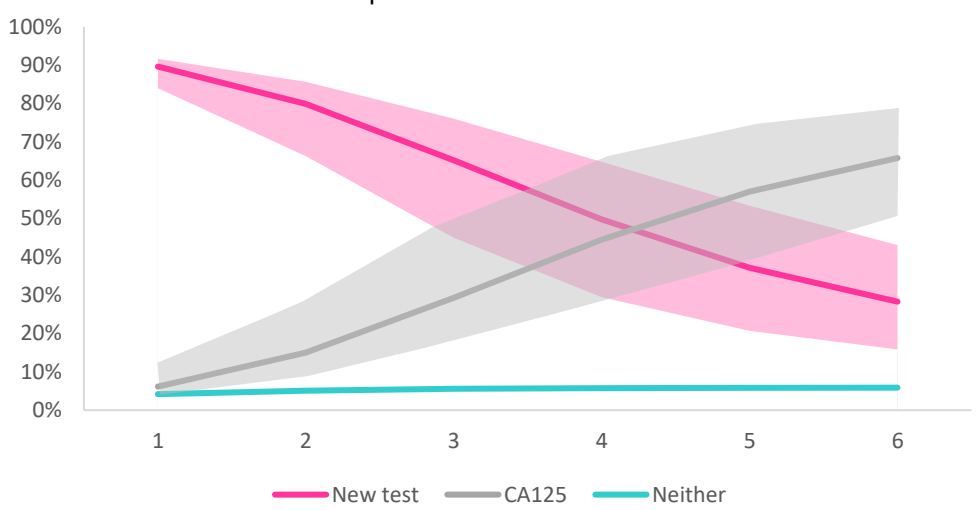


Fig 7.8.3: Time to diagnosis has no impact on survival.
Comparator test in 95% accurate

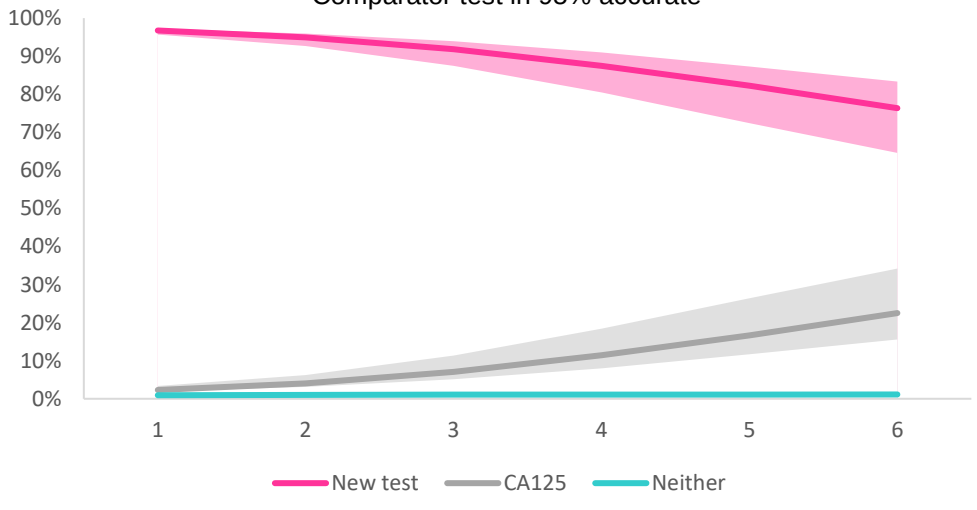
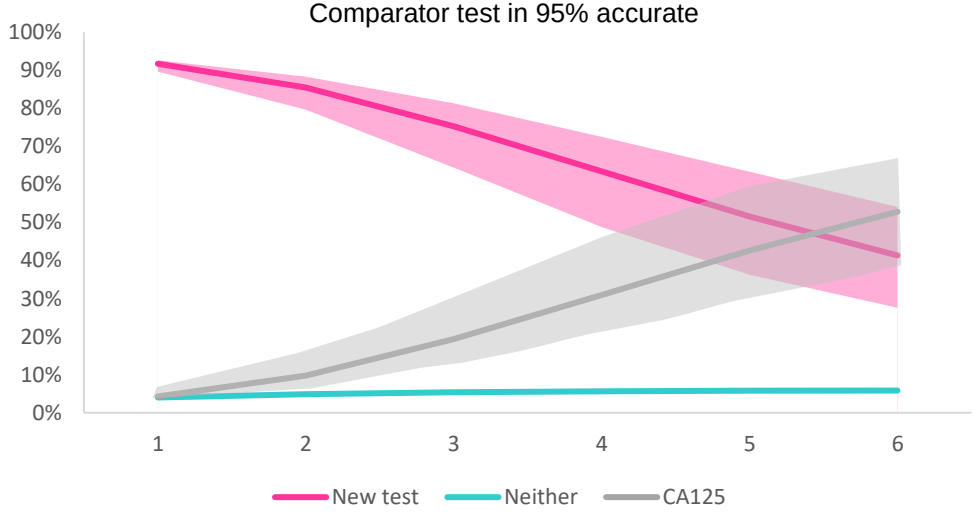


Fig 7.8.4: Time to diagnosis impacts survival.
Comparator test in 95% accurate



7.5 Discussion

7.5.1 Key findings

This study provides important information on the value of different attributes relating to testing patients with symptoms of ovarian cancer presenting in primary care. Overall, demand for testing remained consistently high, even where symptoms suggested a low risk of cancer (1%) with the no-test alternative representing just 5% of responses. The study identified that if experiencing symptoms associated with possible ovarian cancer, women consistently valued test accuracy regardless of symptom severity. Time to diagnosis was found to be second most important attribute to respondents and communication throughout the testing process was rated third in importance on average. Identifiable conditions scored similarly in terms of attribute importance score but overall was considered least important despite the level of cancer risk suggesting symptoms were almost certainly the consequence of an alternative condition given the maximum risk level described to respondents was 3% (i.e. 97% of people would have an alternative condition/reason for their symptoms). These findings remained consistent irrespective of the underlying level of cancer risk described to respondents in terms of symptoms.

The relative importance of attributes shifted once the relationship between delayed diagnosis and chance of survival was introduced. As expected, time to diagnosis became increasingly important, whilst the importance of remaining attributes decreased. However, the overall order of prioritisation remained constant across all survey versions, with accuracy being valued most highly.

Choice share scenario analysis demonstrated demand for an alternative testing strategy to the quick but less accurate CA125 blood test was preferred by a large proportion of people even when the best-case scenario of a 1-month diagnostic interval was assumed. For example, over 50% people would prefer to wait more than 3 months for a test that was 95% accurate and could identify alternative conditions even when this had a significant impact on the chance of survival for those with ovarian cancer.

Mixed logit model estimates highlighted significant heterogeneity in preferences across respondents which was further investigated using subgroup analysis and latent class modelling. Latent class analysis results grouped respondents in to five distinct preference classes, each with different associated relative attribute importance scores and marginal rates of substitution. However, despite these variations, accuracy remained the most important attribute in all classes overall. A limited number of sociodemographic characteristics were associated with the likelihood of belonging to preferences classes; however, sociodemographic associations were limited and weak suggesting preferences around testing may be less systematic and are instead more intrinsically driven, relating to more latent beliefs and/or experiences of each individual.

Low failure rates of the embedded attention check questions (n=6) and high pass rates for the monotonicity (1%) and transitivity (2%) rationality checks indicate high internal validity. Instability in preferences across the study were common but in line with similar studies, with some previous DCEs citing instability rates as high as 81% (Johnson *et al.*, 2019). These 'failures' mirror behaviour observed during the think-aloud pilot study, where participants appeared to adapt and change their preferences as they progressed through the survey as they became more familiar with the task and their preferences towards a previously unfamiliar topic became more established.

7.5.2 Key implications

The study provides an insight into women's preferences towards many aspects of ovarian cancer that continue to be debated as outlined in earlier in the thesis, it is important to acknowledge the strong preferences for fast and quick testing at all risk levels means policy recommendations on preference-based studies alone are unlikely to be feasible in any health system. Instead, results can help guide and motivate guideline revisions if/when they occur to ensure diagnostic guidance is optimised according to patient preference to increase satisfaction and adherence wherever feasible.

7.5.2.1 The accuracy-timing trade-off

The DCE was motivated by the lack of evidence surrounding preferences towards cancer testing in diagnostic settings. Ovarian cancer was selected as an exemplar cancer site at the outset of thesis to explore the extent to which DCEs can inform a key policy question for investigative testing for many cancers; namely, the trade-off between a quick and accessible but less accurate test (CA125) versus a slower test with increased accuracy (TVUS).

During the process of designing and implementing the DCE new research demonstrated CA125 testing performs better than previously anticipated (Funston *et al.*, 2020a). Outstanding uncertainties regarding the test characteristics of TVUS mean it is now unclear whether TVUS provides greater accuracy. As a result, any interpretation of results in terms of directly comparing the use of the two tests are tentative, although results are still useful in guiding clinical practice and policymaking as further evidence on the test performance of both tests continues to emerge. Additionally, the use of a generic, unlabelled approach to the DCE means results still offer valuable insights into preferences surrounding ovarian cancer beyond the choice between CA125 and TVUS and has implications for cancer testing more broadly given the current lack of preference-based studies in this area.

7.5.2.2 Preference for testing even when symptoms vague

Respondents expressed willingness to be tested even where symptoms were vague (e.g. abdominal pain, loss of appetite) and indicated just 1% risk of ovarian cancer. Since the completion of this study, newly published evidence suggests the resource implications of reducing the current urgent referral guidelines to include those with symptoms indicative of a 2% risk of cancer would be modest and manageable within the NHS (Moore *et al.*, 2021). The consistent prioritisation of fast and accurate results regardless of risks within this study suggests patients would respond favourably to lowering the threshold for urgent referral. While under the current system, findings highlight the value of access to testing in primary care amongst low-risk groups. Current guidance recommends the use of CA125 in patients experiencing at least one of four “low risk” (<3% PPV) symptoms (abdominal distension, early satiety or loss of

appetite, pelvic/abdominal pain, increased urination or urinary urgency) on a persistent or frequent basis (NICE, 2015). Findings from this DCE suggest extending the list of qualifying symptoms based on a predictive value as low as 1% would better serve the preferences of patients. Explicitly expanding qualifying symptoms may also increase GP awareness and recognition of symptoms. Demand for testing at low risk levels also has implications for several other cancers where primary care tests are available (e.g. PSA for prostate cancer and faecal testing for colorectal cancer), particularly regarding who and when to test.

7.5.2.3 Sequential versus concurrent testing

NICE guidance currently advocates sequential testing requiring abnormal CA125 and TVUS results prior to referral (NICE, 2015). The motivation for such processes are likely to be resource-driven, aiming to manage costs and reduce waiting times for patients with the highest risks. However, from a patient perspective, the current diagnostic pathway appears sub-optimal. Sequential testing is likely to increase the diagnostic interval (particularly as waiting times for imaging have increased following Covid-19) and decrease the sensitivity of testing. On the other hand, sequential testing is likely = to reduce the number of false positive results and unnecessary follow up testing, and therefore increase the specificity of testing. However, specificity was not prioritised by the target population within the development stages, so much so that it was not included as a final attribute within this study. A finding echoed by existing DCEs in cancer testing which found specificity was rarely the most important characteristic (Chapter 4). Alternative diagnostic strategies such as concurrent testing using CA125 and TVUS with referral recommendations based on abnormal findings from either test or direct referral following CA125 testing appear to better align with women preferences by increasing sensitivity and reducing overall time to diagnosis. Both alternatives appear to be clinically feasible given their use in other healthcare systems (Funston *et al.*, 2019; SIGN, 2018) and based on latest evidence suggesting results from CA125 alone sufficiently can exceed the 3% PPV threshold for urgent referral at appropriate threshold cut points (Funston *et al.*, 2020a).

7.5.2.4 The importance of tailoring policy to the different groups

Whilst evidence suggests CA125 testing is sensitive to ovarian cancer in those over 50, for younger patients test performance is reduced around 65% sensitivity (Funston *et al.*, 2020a). The high importance placed on accuracy within this this study meant that younger patients exhibited higher willingness to wait for more accurate results despite the consequences for survival. This finding suggests that whilst first-line CA125 may be most appropriate for those over 50, who also represent the group most likely to be diagnosed with ovarian cancer (NHS UK, 2020). An alternative approach, using TVUS may accommodate the preferences of younger patients. Given the lower prevalence of ovarian cancer within this population, first-line testing using TVUS also provides an additional benefit of identifying other causes of gynaecological symptoms (although DCE results find this to be of low importance to patients).

7.5.2.5 Symptom awareness as a barrier to early diagnosis

High levels of willingness to be tested across respondents suggests that reluctance in help-seeking is not a major barrier to earlier diagnosis of ovarian cancer, further confirmed by the finding that 69% of respondents would seek advice from their GP within 1 month of onset symptoms. Instead, alongside inefficiencies in the diagnostic process, results of this study suggest a lack of symptom awareness appears to remain a key barrier, with just 20% (122/600) of people recognising the five main symptoms of ovarian cancer. These results highlight the importance of current awareness campaigns largely led by charities and the need for additional avenues for public education.

7.5.2.6 Characterising non-testers

Despite demand for testing being high, examination of characteristics of respondents selecting no test may offer important clinical insights by identifying groups who may be more reluctant to seek help or sceptical of the testing process to help with safety-netting and reducing the risk of missed appointments. Serial selection of the no test option was more likely than a single occasion of opting-out suggesting there may be underlying drivers making reluctance around testing more likely. Analysis of the

selection of the no-test alternative and identification of a “test sceptic” class during latent class analysis identified individuals reporting less-than-good health as a key sociodemographic group who appeared less willing to undergo testing. This finding is important given older age is associated with both a higher incidence rate of ovarian cancer (ONS, 2019a) and a higher prevalence of chronic conditions meaning those with the highest risk of ovarian cancer are also most likely to experience poor health and therefore more sceptical about diagnostic testing (Barnett *et al.*, 2012; Mujica-Mota *et al.*, 2015).

7.5.2.7 Importance of shared decision-making in diagnostic settings

In comparison with other settings such as cancer screening or treatment, decisions regarding testing in primary care are often GP-led and made based on assumptions of patient preferences rather than two-way discussion with previous studies demonstrating patients’ desire for shared decision-making may be underestimated by their GPs (Elwyn *et al.*, 1999; Little *et al.*, 2004). During this study, respondents expressed a desire for an increased role in decision-making during primary care consultations. The importance of communication, particularly within the developmental stages of the DCE may be reflective of the currently unmet desire for an increased role in decision-making leading to feelings of poor communication during medical consultations. Respondents were generally able to meaningfully engage with the choice tasks and made rational decisions regarding medical testing based on their personal preferences formed based on the consideration of seemingly complex concepts such as symptoms, likelihood of cancer and differences in test performance and delivery.

Overall, results suggest patients are willing and able to play a greater role in investigative decision-making and require improved dialogue about the intricacies of testing such as the accuracy of results, which it turn may also improve self-advocacy and safety-netting following negative results but ongoing symptoms. High levels of heterogeneity in preferences highlighted by mixed logit and latent class models further highlights the need for personalised care and patient input during primary care consultations.

7.5.3 Comparison with existing literature

The high demand for testing even at low risks of cancer is supported by previous findings from Banks *et al.* (2014) where demand for investigative testing for colorectal, lung and pancreatic cancer was elicited using vignettes indicating varying degrees of risk (1-10%) described to respondents as symptoms. Findings suggested overall willingness to undergo testing for cancer was consistently high and participants were no more likely to opt for investigation as the riskiness of symptoms increased (with the exception of colorectal cancer where the risk of cancer was balanced against the invasiveness of testing using colonoscopy). Similarly, Whitaker *et al.* (2017) found preferences for GP consultation for perceived cancer risk in primary care were stable between those experiencing "high risk" or "low risk" symptoms with the exception of choice of GP. This current study extends the existing evidence by finding that not only does demand for testing remain high across different cancer risk levels, the priorities around testing are also uninfluenced by risk.

Evidence of increased reluctance to undergo testing amongst those in poor health is limited and mixed. Previous studies have found that in some instances pre-existing conditions (i.e. comorbidities) may lead to delays in help-seeking particularly if symptoms are vague or less burdensome relative to the management of ongoing chronic conditions (Smith *et al.*, 2009); however, findings generally suggest comorbidities do not impact (or even facilitate) reporting of new onset cancer symptoms possibly due to having regular contact with medical professionals (Macdonald *et al.*, 2006; Porta *et al.*, 1996; Salika *et al.*, 2017). However, no studies investigating how and why comorbidities affect the willingness to undergo testing once a risk of cancer has been recognised were identified

This study follows similar studies on cancer screening in finding test sensitivity (or accuracy within this study) was the most important attribute to people facing testing, as highlighted in Chapter 4. However, the lack of importance associated with the communication attribute within this study is somewhat surprising given the strong preference shown during the development stages of the study (Chapter 5) and evidence from previous studies. For example, qualitative studies relating to the

experiences and healthcare preferences of women diagnosed with ovarian cancer during the diagnostic phase highlights a strong desire for improved communication during the diagnostic process (Fitch *et al.*, 2002; Jelicic *et al.*, 2019). These experiential findings are confirmed by results of a previous DCE investigating the preferences of patients seeking a GP consultation for perceived cancer risk, which found communication, namely "listening skills of the GP" to be the highly important, second only to waiting times (Whitaker *et al.*, 2017). Participants were willing to wait an extra 3.5 weeks for an appointment with a doctor with good/very good listening skills (versus very poor listening skills). Despite the low importance placed on communication within this study, the findings mirror Whitaker *et al.* (2017), in showing listening skills were the most important aspect of communication to respondents.

The "identifiable conditions" attribute was similarly considered relatively less important despite preparatory work indicating high levels of relevance and importance to the target population, motivating the inclusion in this current study. The presence of discordance between formative stages (particularly qualitative aspects) and final DCE results has been noted elsewhere (Timmis, 2020), with the suggestion that the use of a mixed methods approach provides the strength of providing multiple insights to a research question but sometimes leads to differing conclusions (Tariq & Woodman, 2013).

Evidence from the marketing conjoint analysis literature may offer an alternative explanation for the mismatch between preferences expressed during the preparatory stages and the final DCE results; relating to the tangibility or concreteness of attributes. Levels associated with the communication attribute, alongside the similarly less-prioritised "identifiable conditions" attribute were expressed linguistically (i.e. "good", "fair", "poor") as opposed to the remaining attributes which utilised numeric levels (e.g. 65%, 75%). Numeric levels are considered to be more "concrete" meaning information is specific, tangible and presented in an easily processible form (Viswanathan & Childers, 1996) . Linguistic attributes are typically more abstract; the meaning is vaguer and evaluation requires further processing. Research shows concrete attributes are easier to understand, process and directly compare (for example, the difference between 1 month and 3 months is directly obtainable than the

difference between “poor” and “fair” communication) (Stone & Schkade, 1991). As a result, concrete attributes such as those with numeric levels, typically have a greater impact on choice. In contrast, abstract attributes require greater cognitive effort leading to selective attribute processing, reduced intra-attribute comparisons and more non-compensatory decision-making, a simplifying heuristic whereby evaluations are made at the alterative rather than attribute-level (Horsky *et al.*, 2004; Huber, 1980; Jiang & Punj, 2010; Nisbett & Ross, 1980).

7.5.4 Strengths and limitations

This study is the first DCE to investigate preferences towards diagnostic testing for ovarian cancer and adds to the limited number of preference-based studies in cancer diagnostic more broadly. A strength of the study is that it was based on a rigorous development and piloting process helping to maximise respondent understanding and relevance to the target population. The results provide valuable insights that can help to guide potential updates to NICE guidance in the face of current inefficiencies in the diagnostic pathway of ovarian cancer. It also adds to the limited number of studies providing evidence that people are willing and able to consider complex decisions regarding cancer risks, symptoms and diagnostic testing, serving for a motivation for improved shared decision-making in this aspect of primary care which is currently neglected.

However, the study is subject to a number of limitations. Firstly, the finding that preferences do not differ between people facing differing risks of cancer is dependent to the risk levels investigated within this study (1%, 2%, 3%). These risk levels were chosen based on the focus on primary care testing as higher risk levels should theoretically result in urgent referral to secondary care. However, differences between included risk levels may not be sensitive enough to reveal differences in preferences according to risk and stability in preferences may not hold at higher risk levels (e.g. 5%, 10%).

The underrepresentation of Black, Asian and Minority Ethnic (BAME) people within the study is a limitation of the study, with just 5% (31/600) of the overall sample self-identifying as non-white. Evidence suggests white people have a higher incidence of

ovarian cancer and worse outcomes (Forman, 2009; Shirley *et al.*, 2014); however, uptake of cancer testing tend to be lower in BAME patients meaning identifying potential barriers should be a priority (Bansal *et al.*, 2012; Jack *et al.*, 2014; Moser *et al.*, 2009). Results of this study follow previous studies in finding differences in preferences for testing according to ethnicity, however, sample size limits the ability to draw a clear conclusion. Additionally, although the survey was open to anyone with ovaries, due to recruitment methods only cisgender women were represented in the final sample. Non-binary people, intersex people and transgender men with ovaries are still susceptible to ovarian cancer and represent groups currently underserved by gynaecological services heightening the importance of understanding the preferences of these populations (Peitzmeier *et al.*, 2014; Stewart *et al.*, 2020; Teti *et al.*, 2021). Importantly, the development stages of the study also primarily utilised evidence from white, cisgender women meaning increased inclusion in the final sample may still misrepresent the preferences of marginalised groups if the included attributes do not match the underlying priorities of these populations.

Communication of risky attributes remains a key challenge within DCEs. The format and presentation of risk within the accuracy attribute was designed and tested using a rigorous development process including workshops, piloting and examples from published DCEs to ensure maximum understanding. The final iteration of the accuracy attribute focused on test sensitivity, specifically the rate of false negatives for those with cancer and levels ranged from 65-95% based on known test performance of the CA125 blood test in primary care population and accommodate for higher levels of accuracy expected for the TVUS but currently not known. This approach is factually correct and utilised by almost all studies containing a “test sensitivity” attribute in Chapter 4. However, stating accuracy in terms of only those **with** the disease may unintentionally have inflated the perceived occurrence of inaccurate (false negative) results given the prevalence of the disease within the population (i.e. between 1-3%) which may have artificially inflated the importance of accuracy. Upon reflection, an alternative approach incorporating prevalence within the accuracy attribute may have been clearer to respondents (e.g. a 65% accurate test would result in 350 in 1000 people with cancer incorrectly told they did not have the disease. However, if disease prevalence is 3%, this equates to just 10-11 people were 1000 people tested).

Ambiguity in the specification of the opt-out alternative within the experiment is a limitation of this study. In particular, caution should be taken when drawing conclusions about uptake. In Chapter 6, piloting participants interpreted the opt-out alternative as intended and as modelled during estimation (i.e. no test would be received), however, it is not possible to say that all respondents in the final study interpreted the opt-out alternative in the same way. For instance, some may understand the choice of "neither" test as a status quo rather than rejection of testing overall.

Finally, current uncertainties surrounding the ovarian cancer testing exist and therefore several aspects of the study are reliant on evidence-based assumptions. For example, the study explicitly assumes that the TVUS is more diagnostically accurate than the CA125. Additionally, the true impact of increased diagnostic intervals on survival is currently uncertain. Comparisons within this study are based on point estimates from a single publication (Sud *et al.*, 2020) which focuses on diagnostic delays in the context of Covid-19. The study aimed to account for uncertainties where possible within time and monetary constraints through scenario analyses and alternative survey versions. Preferences and demand for alternative tests differed substantially according to changes in these assumptions highlighting the need for greater understanding of ovarian cancer diagnosis within primary care (e.g. diagnostic performance of TVUS, sequential testing). Understanding these uncertainties is also crucial to evaluating the feasibility of policy changes better meet preferences.

7.5.5 Future research

Results from this chapter reveal avenues for future research falling into three categories: (i) performance of tests for ovarian cancer; (ii) economic and resource implications of alternative test strategies, (iii) preference elicitation studies.

7.5.5.1 Performance of tests for ovarian cancer

Throughout the thesis the lack of evidence surrounding the performance of investigative primary care testing has been highlighted. Understanding the diagnostic accuracy of alternative testing procedures is crucial for not only meeting the preferences of patients and streamlining investigations in primary care but also

understanding the broader economic and service capacity implications of alternative approaches. Given the preference for testing even at low risk levels, understanding if and how the sensitivity of CA125 testing varies according to the risk profile of symptoms is a further area for future research. Symptoms indicating a 1% risk of cancer are close in nature to asymptomatic population screening, it is unclear whether estimates of test sensitivity in a primary care setting will be upheld particularly given the ineffectiveness of screening in asymptomatic populations (Menon *et al.*, 2021).

7.5.5.2 Economic and resource implications of alternative test strategies

Preference-based results in this chapter add support to calls for updates or amendments to current diagnostic processes including a shift away from sequential testing and/or the use of alternative first-line tests, particularly in populations where the CA125 has been demonstrated to have decreased sensitivity (e.g. women under 50) (Target Ovarian Cancer, 2022b). However, policies cannot be solely determined based on patient preferences. A key area of future research is to understand the capacity and resource implications of alternative test strategies as well as the cost-effectiveness of any proposed changes. Existing evidence on the cost-effectiveness of alternative modalities suggests sequential multimodal screening dominates TVUS screening when used for screening (largely driven by the low disease prevalence in asymptomatic population), however, research in a diagnostic setting is underexplored to date (Menon *et al.*, 2017). Key questions may include economic and service provision implications of sequential versus concurrent versus single test only primary care testing and capacity implications and cost effectiveness of using CA125 and TVUS to test low risk patients in primary care.

7.5.5.3 Future preference elicitation studies

Findings highlight several areas for further preference-based research, specifically relating to ovarian cancer but also diagnostic testing more broadly. Firstly, understanding the barriers and facilitators of testing for marginalised groups is a clear area for future research that should be prioritised (e.g. non-binary/transgender men with ovaries, ethnic minorities). Secondly, whilst decision-making should aim to be as collaborative as possible, diagnostic decisions continue to be largely clinician-driven. This means future research to investigate how the preferences and decisions of

doctors align with patient preferences would be a valuable avenue for future research. DCEs could be an appropriate method to understand GPs perceptions of patient preferences and also how GP preferences align with patient preferences. Encouragingly, results from previous studies examining GP referral behaviour suggest that deviations from GPs who report deviating from guidance by referring patients for further testing regardless of CA125 results or without TVUS appear to be acting in a way that better serves the preferences of their patients (Moss *et al.*, 2013). Finally, given the limited evidence surrounding this topic, similar DCEs could be performed focusing on other cancer sites to understand the generalisability of findings. For other cancer sites, such as colorectal cancer it may also be important to consider additional attributes such as adverse events arising from testing. An attribute that appeared important to respondents during attribute development but is not relevant to ovarian cancer due to the low-risk nature of available primary care tests.

7.6 Conclusions

This study is the first to quantify the preferences towards diagnostic testing for ovarian cancer. At a population-level, accuracy (test sensitivity) appears to be consistently prioritised by people facing testing even where accurate tests are associated with longer waiting times resulting in reduced likelihood of survival. However, analysis also revealed substantial preference heterogeneity highlighting the need for a personalised approach to testing decisions within primary care. The findings of this study may be useful in two ways; firstly, this study suggests the preferences of people facing testing appear to be unfulfilled by the current diagnostic pathway therefore findings may be considered during future revisions of NICE guidelines in order to improve patient satisfaction. Secondly, in terms of immediate clinical practice the study demonstrates the ability and desire for increased patient involvement during decision-making within primary care consultations including the discussion of symptoms, risks and test procedures.

8 Methodological Extensions Part 1- Learning and fatigue effects, and Indifference alternatives

8.1 Introduction

This chapter explores some of the methodological uncertainties highlighted within the previous chapters of this thesis, particularly the systematic review in Chapter 4 and the DCE of diagnostic preferences in Chapter 7. Specifically, the chapter focuses on two methodological areas: i) the inclusion of an indifference alternative within choice tasks, and ii) the relationship between the number of choice tasks per respondent and learning and fatigue. Both of these concepts have unexplored with the healthcare field to date. The two methodological extensions are described sequentially within this chapter.

8.2 An exploration of learning, fatigue and the number of choice tasks

8.2.1 Background

During the qualitative piloting stage (Chapter 6) some respondents appeared to exhibit learning behaviour as they progressed through the choice tasks. For example, answering choice tasks faster over time, revaluating their definitions of attributes in later tasks or in some cases even changing the attributes they considered during deliberations between earlier and later tasks within the choice. Alternatively, some respondents appeared to grow bored or “fatigued” with the seemingly repetitive nature of the tasks. This observation motivated a methodological extension aiming to understand how responses to DCEs vary as respondents progress through choice tasks. In particular, this extension looks for evidence of learning or fatigue effects during the experiment to understand if there is a potential “burn in” stage where respondents are still familiarising themselves with the choice tasks and establishing their preferences and at what point if any, do respondent begin to become fatigued and as a result, the quality of the collected data begins to diminish.

8.2.2 Learning and fatigue effects: existing evidence

The concept of learning and fatigue effects within choice experiments is by no means new and this study was conducted in the context of a large body of existing evidence.

The presence of learning or fatigue effects can affect DCE responses via a number of mechanisms.

8.2.2.1 Model scale and error variance

First, is the effect of increasing numbers of choice sets on model scale. If respondents become fatigued with increasing numbers of choice sets then engagement and attention become reduced, increasing the likelihood of erroneous or random responses resulting in a measurable reduction in model scale (Box 8.1). Alternatively, respondents may become more familiar and comfortable with the choice process as the tasks progress resulting in a learning effect whereby scale increases throughout the tasks due to increased consistency between tasks. It is also possible that both learning and fatigue effects are present, as indicated by an initial increase in scale as newly emerging preferences begin to stabilise followed by a decline in scale as respondents begin tire of the choice process and responses become more random (Hu *et al.*, 2006).

The impact of the number of choice tasks completed on model scale has been the focus of several studies, although investigations in a healthcare setting are limited (Bech *et al.*, 2011). However, there is still no consensus in the literature about the role of learning and/or fatigue effects and the “appropriate” number of choice sets per respondent to offset their presence. An early and frequently cited study by Bradley and Daly (1994), found that in a DCE with sixteen tasks, scale decreased significantly throughout the experiment. Since then a number of additional studies also find higher numbers of choice sets result in increased error variance (Bech *et al.*, 2011; Phillips *et al.*, 2002), whilst other studies find no significant differences in model scale between longer and shorter surveys (Adamowicz *et al.*, 1998; Bateman *et al.*, 2008; Hess *et al.*, 2012). Additionally, a number of studies find evidence of learning effects, as evidenced by a downwards trend in scale as choice tasks progress (Caussade *et al.*, 2005; Holmes & Boyle, 2005; Hu *et al.*, 2006; Kingsley & Brown, 2010; Savage & Waldman, 2008). In almost all instances, this initial learning effect was followed by a subsequent increase in scale indicating a tipping point for fatigue.

Box 8.1: Scale parameter

Comparisons of scale parameter (and error variance) between responses is a central component to analysis of choice responses within this chapter and later chapters of the thesis. This information box provides a background on model scale and accounting for scale differences using the Heteroscedastic logit model.

As part of the specification of the logit model, inferences must be made regarding the shape, location, and dispersion of the distribution of the random error term, ε. Differences in the level of error variance impact the size of parameter estimates in logit models through the presence of a scale parameter, λ. In the IID Grumbel distribution, variance σ_{ni}^2 is defined as $\frac{\pi^2}{6\lambda}$, meaning the scale parameter is inversely related to the variance of the error term (Vass *et al.*, 2018b).

This means the true logit function is actually represented by equation 1:

$$P_{ni} = \frac{e^{\lambda(\beta X_{ni} + \gamma Z_n)}}{\sum_{j=1}^{J} e^{\lambda(\beta X_{nj} + \gamma Z_n)}}$$

(eq. 1)

As a result, estimated β-coefficients for each attribute are actually scaled preferences $\lambda\beta$, indicating the effect of each observable variable relative to the variance of unobserved factors (Train, 2009). As the scale parameter decreases (i.e. variance increases), observed coefficients ($\lambda\beta$) appear smaller, in other words choices become more random.

Within an individual model, the two components are confounded at cannot be separately identified and therefore λ is arbitrarily normalised to 1. This means error variance is assumed to be constant across all individuals withing the MNL model.

However, when attempting to compare preferences across groups where scale (i.e. error variance) differs systematically scale heterogeneity becomes an issue. Scale heterogeneity relates to the "randomness" of choice behaviour or differences in estimated preferences due to differences in the error variance of across respondent choices. Scale heterogeneity can arise for many reasons such as differences in survey design (e.g. number of choice tasks), data collection methods (e.g. electronic vs in-person data collection) or survey populations (e.g. based on level of experience- patients vs general public populations). Confounding between β-coefficients and the scale parameter mean it is not possible to establish whether observable differences in preferences are due to a genuine difference if tastes (i.e. preference heterogeneity) or due to differences in the variance of the error term.

When making comparisons across groups (or in the case of this chapter between tasks within a choice sequence) within this thesis, Heteroscedastic conditional logit models (HCL) are used to estimate and adjust for scale heterogeneity when modelling choice data (Hensher *et al.*, 1998; Hole, 2006). The HCL model allows the scale parameter to be a function of individual, n's characteristics (λ_n). Within the model, λ_n is parameterised as $\exp(Z_n\gamma)$, where Z_n is a vector of individual characteristics and γ is a vector of parameters estimating the effect of individuals' characteristics on the scale parameter. The model collapses to the standard MNL model if $\gamma = 0$, meaning testing if $\gamma = 0$ is equivalent to testing if the error variance is constant across respondents.

8.2.2.2 Rationality of choices

The presence of learning and/or fatigue effects throughout a DCE may also influence the quality of data in terms of the exhibition of “irrational” responses according to the axioms of choice set out in Chapter 2 (e.g. stability, monotonicity, transitivity). Learning effects may result in increased irrational responses during the early stages of a survey when respondents are less familiar with the DCE format whereas fatigue effects may lead to an increase in rationality failures in the later stages. Unlike other aspects of data collection, the relationship between task length and irrational responses has been neglected to date. A few studies discuss the stability of responses however discussions refer to stability in scale rather than test-retest stability. These studies typically find evidence of a brief learning period at the start of DCEs with respondent’s appearing to quickly form stable preferences (Bradley & Daly, 1994; Hess *et al.*, 2012).

8.2.2.3 Decision processing

Observations during the pilot study suggest that decision processing strategies may also vary across choice tasks and how such strategies vary may differ across respondents. Even within the small qualitative sample in Chapter 6 clear differences emerged. Some individuals described re-evaluating attributes they had previously dismissed as unimportant as choice tasks progressed whilst other respondents appeared to disregard attributes which they had previously seemed to consider in the later stages of the survey. This behaviour may be indicative of both learning or fatigue effects. Respondents may disregard attributes as a simplifying heuristic due to fatigue or may instead disregard information they may deem unimportant based on the refinement of their preferences on the topic.

8.2.2.4 Marginal rates of substitution

The previous mechanisms all have potentially important consequences for the final estimates and outcomes of models. In particular, differences in processing strategies and/or the error variance of data due to learning or fatigue may result in variations in the relative importance placed on attributes as the choice tasks progress. Comparisons of willingness to pay or willingness to accept estimates between early and late stage responses provide mixed evidence, although many studies find limited

or no effects based on the number of choice tasks completed (Hensher, 2006) even where differences in scale were present (Caussade *et al.*, 2005). However, there is contrasting evidence to suggest there are significant differences in MRS estimates based on the number of choice tasks completed (Bech *et al.*, 2011; Campbell *et al.*, 2015; Holmes & Boyle, 2005; Plott & Zeiler, 2005) with further evidence that the capacity for learning and fatigue effects (and subsequent variations in consistency and WTP) may differ across respondents (Campbell *et al.*, 2015) (although authors acknowledge that differences may attributable to changes in processing strategies throughout the experiment).

8.2.2.5 Healthcare-based research

Investigations of learning and fatigue in relation to the number of choice tasks demonstrate mixed evidence. Additionally, applications have be primarily conducted in other domains and findings do not appear to be adopted within healthcare contexts to date. It also remains unclear whether such findings *should* be applied to health given differences in the choice context (e.g. health-based DCEs are more likely to involve the valuation of non-market goods, include risky or complex attributes).

Bech *et al.* (2011) is perhaps the most notable example from healthcare perspective. The study investigated how the completion of greater or fewer choice tasks per respondent influenced DCE outcomes and procedures in an experiment examining preferences towards dental services. Results indicated the presence of small fatigue effects once a threshold number of choice tasks was exceeded (17 tasks), indicated by increasingly random responses and changes in MRS. Choice difficulty or certainty did not differ based on the number of choice tasks completed. This single case study is frequently cited in isolation as justification for the number of choice tasks within healthcare studies over a decade later despite conflicting evidence from the broader research field (Buchanan *et al.*, 2022; de Bekker-Grob *et al.*, 2013a; Mühlbacher *et al.*, 2017).

8.2.3 Aims

Given the mixed evidence from this pilot study, as well as the DCE literature more widely. Adaptations to the diagnostic DCE developed and presented in Chapters 5-7 were made to incorporate an investigation of this issue in the context of preferences for diagnostic testing.

This study utilises methods described in previous study to understand variations in DCE outcomes based on the number of choice tasks or stage in the survey but also extends this evidence by also examining how irrational responses vary depending on the number of choice tasks completed.

To summarise, three research questions were addressed:

i. How does performance of rationality check questions differ based on the stage of the survey?
ii. Does scale heterogeneity (i.e. error variance) vary based on the stage of the survey?
iii. Do estimates of the marginal rates of substitution between attributes differ between early and late stage responses?

8.2.4 Methods

8.2.4.1 Study design

Analysis is based on pooled data from DCE versions 1-3 in Chapter 7. Data was pooled across sub-versions since analysis demonstrated no significant differences in preferences based on risk of cancer. The necessary data collection was conducted using six sub-versions of the experimental design. Sub-version varied in terms of question order and position of rationality checks only. The six sub-versions are shown in Figure 8.1.

Firstly, the experimental design was split into 2 blocks, each with 8 choice sets. Each respondent completed both blocks during the survey, however the order the blocks

was randomised evenly across respondents to create two complete response sets to allow for the examination of early (first 8 tasks) versus later responses (last 8 tasks). This approach also controls for fatigue effects in later questions when examining responses across the entire experiment. Next, to explore the relationship between rationality of responses and the number of choice tasks completed, the location of different rationality checks was varied versions. This allowed for the examination of rationality failure rates between early, late and across choice experiments. The separation resulted in 75 responses per sub-version. In total, all respondents completed 20 choice sets (both blocks plus 4 rationality check tasks).

Figure 8.1:Overview of the question structure for each sub-version used to address the research aims. Respondents were randomised to one sub-version

Transitivity in early choices, monotonicity and stability in later choices

1	2	3	T1	4	5	6	7	T2	8	9	10	11	M	12	13	14	15	S9	16

Transitivity later choices, monotonicity and stability in earlier choices

1	2	3	M	4	5	6	7	S1	8	9	10	11	T3	12	13	14	15	T4	16

Transitivity and stability across entire choice set, monotonicity in earlier choices

1	2	3	M	4	5	6	7	T1	8	9	10	11	T2	12	13	14	15	S1	16

Transitivity later choices, monotonicity and stability in earlier choices

9	10	11	M	12	13	14	15	S9	16	1	2	3	T1	4	5	6	7	T2	8

Transitivity in earlier choices, monotonicity and stability in later choices

9	10	11	T3	12	13	14	15	T4	16	1	2	3	M	4	5	6	7	S1	8

Transitivity and stability across entire choice set, monotonicity in earlier choices

9	10	11	M	12	13	14	15	T3	16	1	2	3	T4	4	5	6	7	S9	8

M= Monotonicity check
T1 and T2= Transitivity check based on Task 2
T3 and T4= Transitivity check based on Task 10
S1= Stability check, repeat of Q1
S9= Stability check repeat of Q9

8.2.4.2 Addressing the research questions

Research question 1: How does performance of rationality check questions differ based on the stage of the survey?

The frequency of rationality failures in early, late and cross experiment positions were compared based on count data. Rationality checks included comparisons of monotonicity, transitivity and stability. A manual check for non-compensatory behaviour by examining responses for evidence of decisions made based on a single attribute was also performed. Next, logistic regression was performed to examine how stability of preferences varied at different points across the choice experiment. Similar analysis was not possible for the remaining rationality check questions due to the overall low rate of failures.

Research question 2: Does scale heterogeneity (i.e. error variance) vary based on the stage of the survey?

Heteroscedastic conditional logit (HCL) models (described in Box 8.1) were used to investigate variations in scale that may indicate the presence of learning or fatigue effects. Several models were estimated to investigate differences in scale between early and late responses.

Models 1 and 2 sought to investigate differences in scale at an aggregate level.

Model 1: HCL model including a dummy variable to account for potential differences in scale between early and late responses ("early response dummy" = 1 if for responses to the first 8 tasks and 0 for responses to final 8 tasks). A positive parameter estimate would indicate increased error variance in later stages of the survey (i.e. fatigue effects) whereas a negative parameter estimate would signify evidence of learning effects (i.e. decreased variance in later stages of the survey).

Model 2: HCL model including continuous question order variables. Following previous studies question order was included as both a linear (coded 1-16) and squared function in order to take account of the possible presence of both learning and fatigue

effects (U-shaped form, representing an initial increase in scale followed by an eventual decline) (Bech *et al.*, 2011; Caussade *et al.*, 2005).

Models 1 and 2 also included choice confidence, utility balance and task difficulty as explanatory variables to control for any potential confounding effects on scale (Bech *et al.*, 2011).

Models 3 and 4: HCL model including dummy variables for each choice as an explanatory variable for the scale function. The purpose of these models was to investigate differences in scale between each question, similar to the approach utilised by Hess *et al.* (2012). The scale parameter for the first task was normalised to 1 (by excluding from the model) meaning all question parameters were interpreted relative to this task. Trends across choice tasks are of particular interest. If learning effects are present, a period of increased error variance (i.e. decreasing scale) from question to question while participants adapt followed by a period of stable or decreased error variance (i.e. increasing scale) may be expected. Two models were estimated to control for differences in the question order; 3 for respondents who completed Block A followed by Block B, 4 for respondents who completed Block B followed by Block A. To complete the analysis, a linear regression model was estimated to investigate the relationship between scale parameters and question order. Choice certainty and utility balance between alternatives within each choice set were also included to control for possible confounding effects.

Research question 3: Do estimates of the marginal rates of substitution between attributes differ between early and late stage responses?

To test for differences in marginal rates of substitution, two separate mixed logit models were estimated based on responses from early responses (first 8 tasks) and late responses (final 8 tasks). Following analysis described in Chapter 7, estimates of willingness to wait for a diagnosis in exchange for improvements in the remaining attributes were calculated based on the estimated parameters of each model. Early and late stage estimates were compared using a paired T-test. Since MRS is a ratio-based calculation, any potential differences in scale between early and late responses does not impact estimates.

8.2.5 Results

The median completion time was approximately 30 seconds slower for the earlier tasks compared to later tasks (completion time for rationality check questions was not included in calculations). However, the implications of this finding may indicate respondents become more efficient as they become more familiar with the choice tasks but could equally indicate growing fatigue and less consideration of the choices they are presented over time.

8.2.5.1 Research question 1: How does performance of rationality check questions differ based on the stage of the survey?

Failure of rationality check questions according to placement within choice sets are shown in Table 8.1. Overall, failure rates were low regardless of where the questions were placed within the survey and there did not appear to be any differences in irrational responses based on the number of choice tasks completed. Interestingly, a large proportion of respondents appeared to demonstrate non-compensatory behaviours (always choosing the choice task where one attributes was always dominant, most often accuracy) in either the early or later stages of the choice task, however, only about half of these individuals displayed this behaviour consistently across all choice tasks. For stability failures, logistic regression confirmed the observational finding that of the validity questions was not associated with the probability of displaying unstable responses (Appendix 8.1). Overall, examination of rationality failures provided no evidence of learning effects or fatigue effects through the choice experiment. Additionally, average choice certainty was also similar between early and late responses.

Table 8.1: Summary of rationality failures and key statistics at different stages of the choice experiment

Rationality check	Early responses	Late responses	Cross survey
Transitivity	1 (<1%)	4 (3%)	3 (2%)
Monotonicity	2 (1%)	1 (<1%)	-
Stability	37 (25%)	30 (20%)	35 (23%)
ANA (dominant attribute)	140 (31%)	146 (32%)	68 (15%)
Average choice confidence (1-10)	7.22	7.31	7.27
Median completion time (range)	2m 37s (43s—11m 22s)	2m 8s (49s—12m 58s)	4 m 52s (2 m 12s—15m 31s)

8.2.5.2 Research question 2: Does scale heterogeneity (i.e. error variance) vary based on the stage of the survey?

Table 8.2 shows the results of the HCL models used to examine differences in error variance based on the number of choice tasks completed by respondents. Model 1 includes a dummy variable representing early responses. The coefficient associated with the dummy variable is small but significant. The positive coefficient indicates that error variance was comparatively lower across the first eight choice tasks than the last eight tasks suggesting the possibility of fatigue effects in the later stages of the experiment. Contrastingly, neither the question number nor the square of the question number was significantly related to the scale in Model 2, suggesting the number of choice tasks completed has no impact on the error variance across observations.

In both models, higher confidence in decisions during choice tasks was significantly associated with a small decrease in scale whilst self-reported task difficulty and utility balance between alternatives within choice tasks were not significantly related to error variance.

Scale differences across choice tasks

Analysis of choice task specific scale parameters based on Models 3 and 4 are shown in Figures 8.2 and 8.3, representing the order that question blocks A and B. Each figure plots the scale parameter of each progressive choice task relative to the first choice task where the scale parameter was normalised to 1. Increased scale estimates indicate lower unobserved variance and decreases in scale represent higher levels of variance. This means a downwards trend would indicate fatigue effects whereas an upwards trend as choice tasks progress would indicate learning effects.

Upon initial inspection, scale parameters associated with later choice tasks appear to decrease as choice tasks progress suggesting increasing variance in choices (orange nodes indicate a significant difference in scale relative to the first task), this trend is particularly evident for respondents that completed block A then block B (Figure 8.2). However, changes in scale between adjacent choice tasks appeared more sporadic (triangular nodes indicate significant scale differences between a choice task and the

task prior). Instead, scale parameters also appeared to follow a similar trajectory to variations in average choice confidence for each task (pink line) and utility balance between test alternatives within each choice task (grey line) meaning isolating relationships between scale parameters and the number of choice tasks was not possible based on visuals alone. Regression analysis with scale as the dependent variable demonstrated question order was not significantly associated with variations in scale, however, both utility balance and confidence in choices were significantly related to changes in unobserved variance (Appendix 8.2). Increases in utility balance were associated with decreases in scale whereas increased choice confidence was linked to an increase in scale.

Table 8.2: HCL models used to examine differences in scale based on the stage in the choice experiment

	Model 1 Coefficient (95% CI)	Model 2 Coefficient (95% CI)
Attributes		
Accuracy	**0.05***** (0.03 – 0.07)	**0.06***** (0.04 – 0.08)
Time to diagnosis	**-0.20***** (-0.29 – [-0.12])	**-0.22***** (-0.31 – [-0.12])
Identifiable conditions		
Cancer only	-	-
Cancer plus related conditions	**0.41***** (0.25 – 0.58)	**0.44***** (0.26 – 0.62)
Communication		
Poor	-	-
Fair	**0.38***** (0.22 – 0.54)	**0.41***** (0.22 – 0.58)
Good	**0.48***** (0.28 – 0.68)	**0.51***** (0.29 – 0.74)
Neither test	**-0.33***** (-0.53 – [-0.14])	**-0.36***** (-0.57 – [-0.14])
Scale terms		
Early response dummy	**0.06**** (0.01 – 0.12)	-
Question number	-	**-0.01** (-0.04 – 0.02)
Question number squared	-	**0.00** (0.00 – 0.00)
Choice confidence (0-10)	**0.00***** (0.00 – 0.01)	**0.00***** (0.00 – 0.01)
Utility balance	**0.00** (-0.02 – 0.03)	**0.00** (-0.02 – 0.03)
Task was difficult/very difficult	**0.03** (-0.10 – 0.16)	**0.03** (-0.10 – 0.16)
Model fit statistics		
LL	-4,438.57	-4,439.92
Observations	20,598	20,598
N	450	450
Key: ***significant at 99% confidence level; **significant at 95% confidence level; *significant at 90% confidence level		

Figure 8.2: Variations in scale between progressive choice tasks based on Model 3

(Responses from individuals who completed Block A followed by Block B). Orange nodes= scale was statistically different from the first choice task. Triangular nodes= scale was statistically different from the scale of the previous choice task.

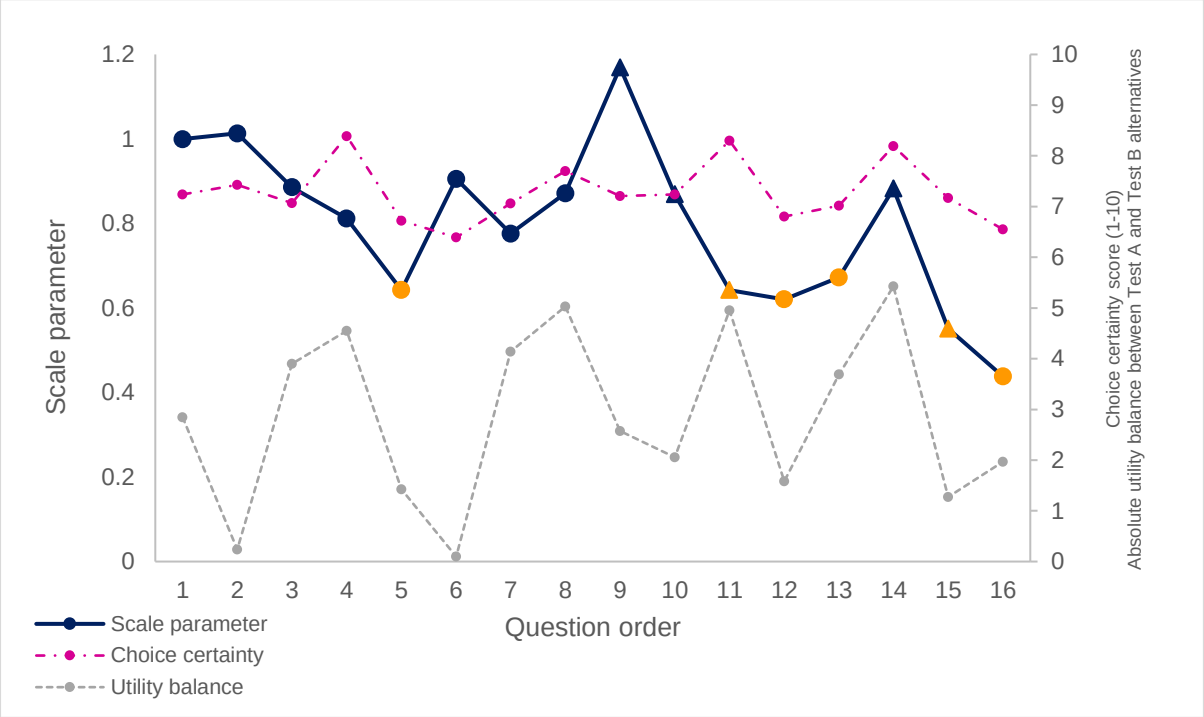
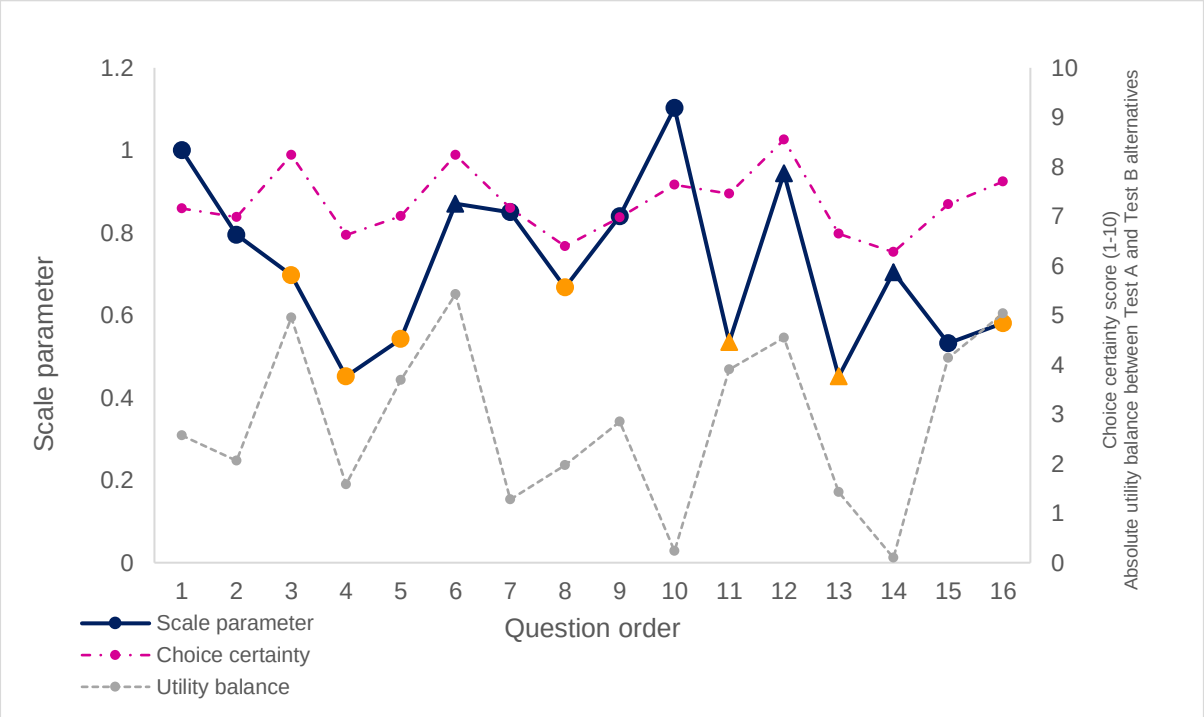


Figure 8.3: Variations in scale between progressive choice tasks based on Model 4

(Responses from individuals who completed Block B followed by Block A). Orange nodes= scale was statistically different from the first choice task. Triangular nodes= scale



8.2.5.3 Research question 3: Do estimates of the marginal rates of substitution between attributes differ between early and late stage responses?

Marginal rates of substitution expressed as the willingness to wait for improvements in the remaining attributes are shown in Table 8.3. Separate estimates were calculated based on responses to early and late estimates. The third column shows full experiment estimates first presented in Chapter 7 for comparison. The final column shows results of a paired T-test comparing estimates from early and late responses. Willingness to wait estimates did not significantly differ between early and late responses for any attributes meaning there was no evidence of either learning or fatigue effects when considering the marginal rates of substation resulting. The models from which willingness to wait estimates were derived are provided in Appendix 8.3.

Table 8.3: Estimates of willingness to wait for an improvement in an attribute according to the stage in the DCE

	Early responses	Late responses	All responses	p-value*
Accuracy				
Per 1%	0.27 (0.30–0.54)	0.28 (0.24–0.32)	0.28 (0.25–0.31)	0.54
Identifiable conditions				
Cancer only	-	-	-	-
Cancer plus additional related conditions	1.72 (1.41–2.03)	1.57 (1.27–1.87)	1.70 (1.45–1.96)	0.51
Communication				
Poor	-	-	-	-
Fair	1.60 (1.30–1.89)	1.48 (1.31–1.91)	1.67 (1.44–1.90)	0.93
Good	1.99 (1.65–2.33)	1.97 (1.58–2.32)	2.02 (1.74–2.30)	0.86

*p-value from a paired t-test to investigate differences in marginal rates of substitution between early and late responses

8.2.6 Discussion

This study adds to existing evidence by examining changes in preferences throughout choice experiments and the role of learning and fatigue effects. In particular, this research tested how the stage in the choice tasks and the number of choice sets completed impacts (i) rationality of choices, (ii) model scale and error variance, (iii) estimates of marginal rates of substitution between attributes.

Investigations of choice rationality based on the stage of the choice experiment added a unique element to the study. Overall findings suggest no evidence of either learning or fatigue effects as the rate of irrational choice behaviours did not vary according to the stage of the survey they were completed (early vs late vs cross survey). However, it is noted total failures were low across the whole study and patterns may emerge in more complex studies or studies with a larger sample size where higher rates of irrational behaviour would be expected.

Comparisons of model scale between early and late responses based on the experimental design blocks (first 8 questions versus final 8 questions) found early responses were associated with a significantly higher scale. This finding implies that error variance increased during the later stage of the survey suggesting respondents became fatigue as choice tasks progressed. However, overall evidence of learning and fatigue effects based on scale were limited since examinations based on question order found no overall trend in scale. Examination of error variance at a micro-level (i.e. variations between adjacent choice tasks) did find significant differences between questions, however, regression analysis revealed this was primarily due to differences in utility balance and choice confidence rather than the progressive number of questions completed. This finding is in line with previous studies that find examination of differences in error variance based on the number of choice tasks can be mistaken or intertwined with caused by order or position effects of choice tasks within the experimental design (Campbell *et al.*, 2015; Day *et al.*, 2012). Failure to randomise the order of tasks between respondents as recommended by Hess *et al.* (2012) is a limitation of this study. Similarly, no differences in willingness to wait were identified between earlier and later responses during the survey.

Failure to find evidence of either learning or fatigue effects according to the number of choice tasks completed within this study is consistent with findings from several previous studies and has important implications for DCEs (Adamowicz *et al.*, 1998; Bateman *et al.*, 2008; Caussade *et al.*, 2005; Hensher, 2006; Hess *et al.*, 2012). In health economics, leading guidance currently recommends an ideal range of 8—16 choice sets per respondents, with studies seldom exceeding this upper limit (Bridges *et al.*, 2011; Soekhai *et al.*, 2019). An increase in the number of choice sets completed

by each respondent means fewer respondents may be needed to achieve sufficient observations to estimate preference (Rose & Bliemer, 2009). This is particularly useful when research relates to the preferences of hard to recruit populations (e.g. patients of specific diseases) or where there are budget constraints. This study adds to growing evidence within a healthcare cases study that respondents are able to complete a large number of choice tasks (in this case 20) without any detriment to the quality of data due to fatigue and increased inattention.

Finally, whilst fatigue and learning effects appeared to be limited at an aggregate level, findings from this study highlight the potential for differences in error variance and preferences during early and late stages of DCEs across individuals as an area for future consideration. In particular, examination of non-compensatory behaviour suggested some individuals may experience evolving decision processing strategies across experiments. For instance, some respondents appeared to initially ignore certain attributes during early decisions before expanding their deliberations to include more attributes, whilst other respondents appeared to do the opposite and narrow down their attributes of interests as tasks increased. From the data alone in is unclear whether such actions are due to simplifying heuristics caused by boredom or attention loss or due to the refinement of preferences as tasks progress. Similar patterns were observed during the think-aloud pilot study in Chapter 6, further qualitative research may add further insights to this phenomenon. Variations in choice processing based on choice task complexity and positioning is a longstanding area of research in choice modelling and is of growing interest in applications to health (Jonker *et al.*, 2018; Pinto-Prades *et al.*, 2019). Further empirical examples are needed to investigate and disentangle these overlapping concepts.

8.3 Investigating the presence of indifferent preferences within discrete choice experiments

8.3.1 Background

Discrete choice experiments, particularly when considered under the dominant paradigm of random utility theory (RUT), rely on the assumption that individuals are rational decision-makers that maximise their utility. This means respondents opt for a

given alternative if, and only if it provides the highest expected utility among all alternatives presented within a choice set (McFadden, 1974; Thurstone, 1927).

Stated preference methods are reliant on experimentally designed choice sets. Choice sets must be carefully designed to ensure realism and maximise the external validity of observed choices. The importance of including (or at least considering) an opt-out option (i.e. “neither” or “no choice” alternatives) within choice sets is widely acknowledged within the DCE literature, particularly in studies aiming to estimate demand and/or elasticity of demand (Bridges *et al.*, 2011; Campbell & Erdem, 2019; Lancsar & Louviere, 2008). Provision of an opt-out option means participants are not forced to choose between potentially unappealing alternatives, increasing the likelihood that participants choose in a way that is consistent with how they would do in a real-life situation (Campbell & Erdem, 2019). Alternatively, the inclusion of a status-quo alternative is also common. Status quo alternatives are an opt-out alternative that allows respondent to state a preference towards their current situation rather than a complete rejection (e.g. a person receiving treatment may prefer to stick with their current medication rather than a new alternative) (Campbell & Erdem, 2019).

However, while the inclusion of non-participation alternatives (i.e. opt-out or status-quo alternatives) is common practice, the inclusion of an indifference alternative has been explored to a lesser extent to date and particularly within healthcare contexts. An indifference alternative (also expressed as “don’t know” or “no opinion”) allows respondents to express indifference or neutrality between the profiles presented within a choice set (Hess *et al.*, 2013). The concept of indifferent preferences is central to alternate preference elicitation methods utilised within healthcare settings. Namely, time trade-off and standard gamble methods, where the objective is to identify the point of indifference between competing alternatives, often health states (see Chapter 2). However, considerations of indifferent preferences within health-based DCEs remains uncommon. For instance, just one inclusion of an indifference alternative was identified during the systematic review in Chapter 4 and it is unclear how such responses were utilised during analysis (Pignone *et al.*, 2012). Examples from the broader healthcare literature are also sparing (Medina-Lara *et al.*, 2014; Robinson *et al.*, 2015). The current omission of indifference alternatives appears to directly conflict

with ISPOR Good Practice Guidance where criterion 5.2 questions: “Did (should) the elicitation format allow for indifference?” (Bridges *et al.*, 2011).

8.3.1.1 Why is consideration of indifferent preferences important?

If experiments do not allow respondents to express indifference respondents may be forced to arbitrarily choose between alternatives introducing stochasticity and reducing the efficiency of information that can be gained from choice tasks. For example, Cantillo *et al.* (2010) demonstrated randomly assigning indifference responses to available alternatives (as respondents would do if no indifference alternative was available), lead to a significant decline in the model's capability to recover the input parameters. Furthermore, studies have demonstrated that the omission of an indifference alternative may artificially increase the selection of other “non-participation” options (i.e. opt-out or status quo alternatives) particularly within choice sets where there was no clearly superior alternative (Dhar, 1997; Fenichel *et al.*, 2009; Tversky & Shafir, 1992). Therefore, Fenichel *et al.* (2009) advocate for the inclusion of indifference alternatives when also offering non-purchase options as a way to reduce bias.

In recent years, the presence of indifferent preferences in healthcare is likely to be exacerbated by the increased utilisation of efficient experimental designs (Bliemer *et al.*, 2009). Since such designs explicitly aim to increase available choice information by generating experimental designs with near-utility balance and minimise the presence of dominant alternatives within choice tasks, making differences between alternatives less perceivable on average (Huber & Zwerina, 1996).

This existing evidence suggests the need to consider the implications of not accounting for the presence of indifference during analysis. However, there are also a number of challenges associated with the inclusion of indifference alternatives, which may contribute to the underutilisation to date. In particular, adding an additional alternative to each choice tasks increases the complexity of decisions (Bridges *et al.*, 2011). Where multiple alternatives (e.g. status quo, optout and indifference) are included the distinction between the different options may become blurred to respondents. Additionally, research also suggests that respondent’s may be tempted

to utilise indifference alternatives as a simplifying heuristic to avoid difficult decisions (Tversky & Shafir, 1992).

8.3.2 Research aims

In response to the challenges and uncertainties surrounding indifferent preferences, this methodological extension seeks to understand how the inclusion or exclusion of indifference alternatives affects the outcomes and processes within a DCE.

Specifically, this methodological extension seeks to address four key questions:

i. How frequent are indifferent preferences and what are the potential reasons for indifferent responses during DCE tasks?
ii. How does the inclusion or exclusion of an indifference alternative impact the on the quality and/or reliability of results?
iii. Does failure to allow indifferent responses during DCE tasks influence unobserved variance of results?
iv. How do marginal rates of substitution vary between DCEs that include an indifference alternative and those that do not?

8.3.3 Methods

8.3.3.1 How frequent are indifferent preferences and what are the potential reasons for indifferent responses during DCE tasks?

Examinations of the rates of indifferent preferences during versions 1-3 of the DCE examining preferences towards diagnostic testing for ovarian cancer presented in Chapter 7 formed the basis for analysis.

Self-reported reasons for indifference selection were investigated using a follow up question after the completion of all choice tasks. Respondents were given a list of pre-established reasons (based on researcher assumptions) to select from but also had the opportunity to express additional reasons using the “other” response option within the question.

Next, since this study utilised an efficient experimental design, the correlation between difference in utility between choice alternatives (Test A and Test B) and selection of the indifference alternative was examined. It was anticipated that tasks with increased utility balance between alternatives will result in a higher number of indifference selections across respondents. The correlation between indifference and choice confidence (1-10) across tasks was also explored. It was expected that higher rates of indifferences would be associated with lower rates of average choice confidence scores.

Finally, logistic regression was used to examine the association between selected respondent characteristics (e.g. education, symptom awareness, medical decision-making style) and choice task characteristics (e.g. selection of opt-out alternative, choice confidence, utility difference), and the selection of the indifference alternative. Included characteristics and hypothesised relationships are described in Appendix 8.4.

8.3.3.2 How does the inclusion or exclusion of an indifference alternative impact the quality and/or reliability of results?

As described in Chapter 6, an additional DCE version was used to address this research question. This additional sub-version was identical to Version 3 described in Chapter 7 except no indifference alternative was provided. As a reminder, choice task 1 from each survey version is shown in Figure 4. Both versions elicited preferences for testing when facing a 3% risk of cancer. The sample size for both versions was 150 respondents. It is important to note that indifferent responses were treated as missing and excluded from model estimates throughout this thesis since the incorporation of indifference responses is an ongoing methodological challenge (see section 8.2.5. below for further discussion).

The quality and reliability of responses was compared between the two versions based on a variety of dimensions:

- Rationality failures (e.g. monotonicity, transitivity and stability of choices)
- Self-reported attribute non-attendance (ANA)

- Self-reported task difficulty
- Average choice confidence
- Number of opt-out alternative selections

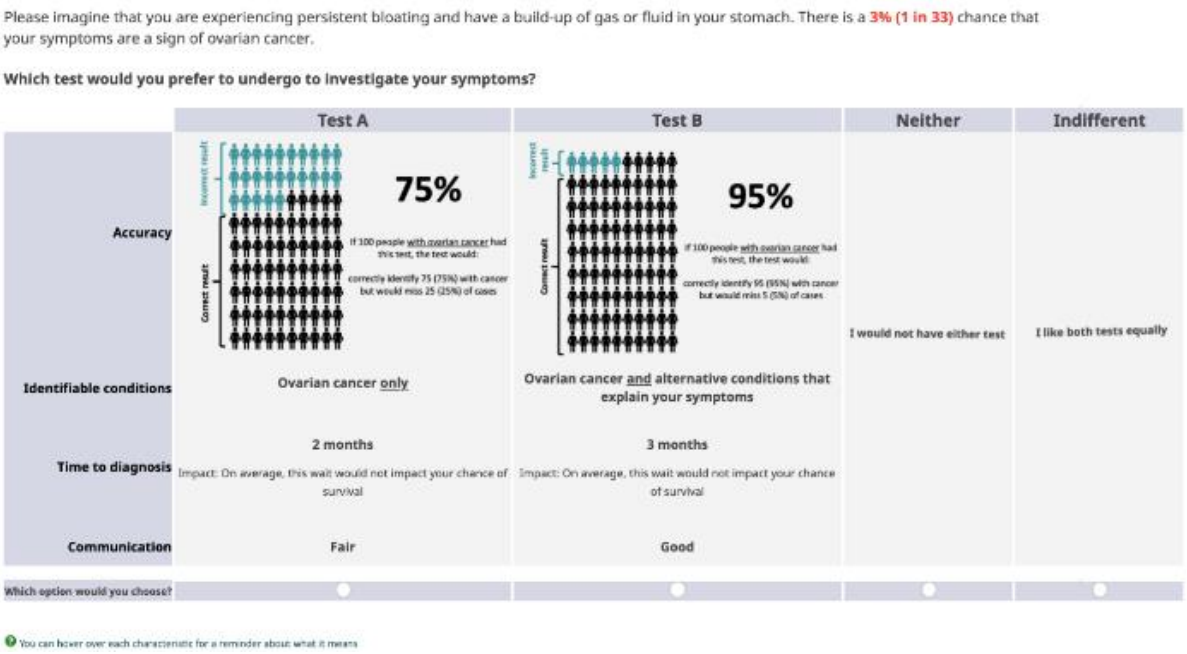
8.3.3.3 Does failure to allow indifferent responses during DCE tasks influence unobserved variance of results?

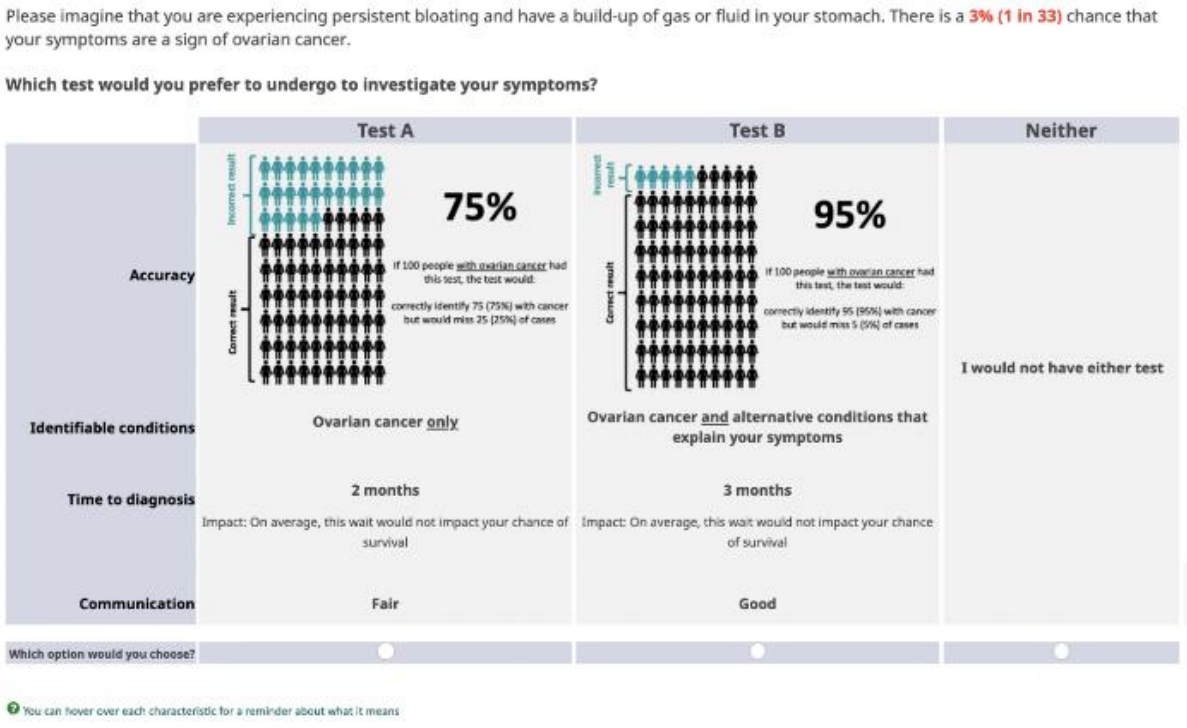
A heteroscedastic logit model was used to investigate differences in unobserved variance between the DCE version including an indifference alternative and the version without. The exclusion of an indifference alternative may increase error variance (i.e. randomness) within the data as respondents with indifferent preferences are forced to artificially express a preference between the available alternatives. On the other hand, the additional choice alternative within each task may add additional complexity in choice tasks which may result in increased error variance.

8.3.3.4 How do marginal rates of substitution vary between DCEs that include an indifference alternative and those that do not?

Marginal rates of substitution were used to compare differences in outcomes between the DCE version including an indifference alternative and the version where indifference was not considered. As with previous chapters, estimates were expressed in terms of willingness to wait and were based on mixed logit models with 1,000 Halton draws.

Figure 8.4: Example of a choice task from the two versions of the DCE used to examine the impact of including or excluding an indifference alternative within choice tasks





8.3.4 Results

8.3.4.1 How common are indifferent preferences and what are the potential reasons for indifferent responses during DCE tasks?

Selection of indifference

The indifference alternative was selected 324 times across all survey versions, accounting for 5% (324/7,200) of total choices. One hundred and twenty-seven of the 450 (28%) respondents selected the indifference option at least once throughout the choice tasks.

Self-reported reasons for selecting indifference are shown in Table 8.4. The most common reason (n=58) for the selection of indifference was: “I thought both options were good/I would be happy with either test”. Interestingly, a common self-submitted answer was a dislike for both options but a desire to undergo testing, nonetheless (n=12).

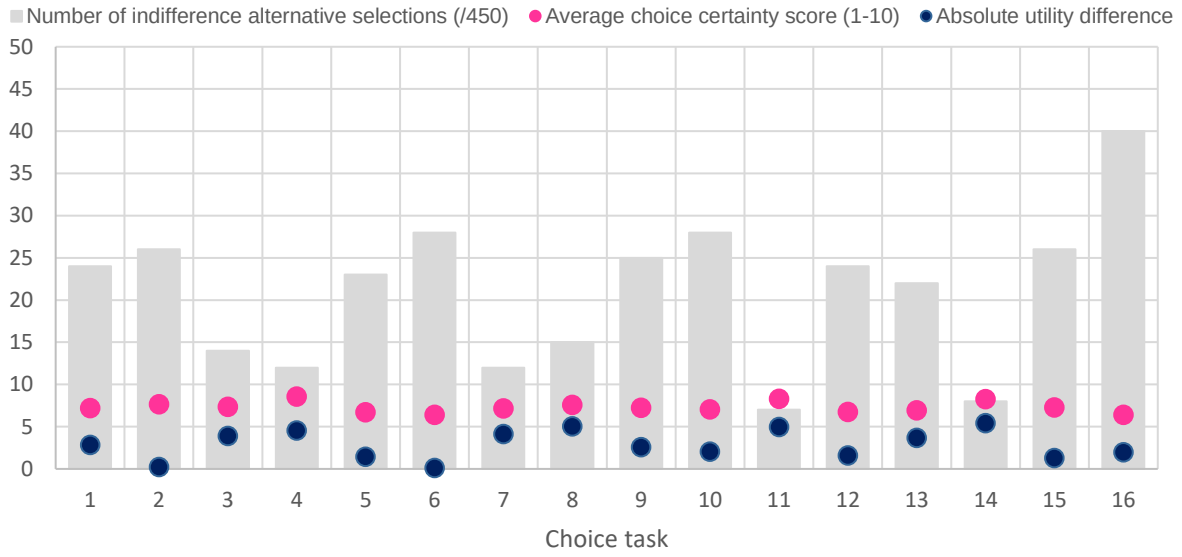
Table 8.4: Self-reported reasons for selecting indifference during DCE versions 1-3

Reasons for indifference	**n**
I liked the two options exactly the same	39
I thought both options were good/I would be happy with either test	58
I count not tell the difference between the two tests	7
I disliked both tests but still want to be tested (write in option)	12
Other	23

Choice confidence, utility balance and selection of the indifference alternative

Figure 8.5 plots the frequency of indifference selection (grey bars), utility balance between alternatives A and B (blue dots) and average choice confidence for each task (pink dots). Two-way correlations between the three measures were all significant. As expected, selection of the indifference alternative was negatively correlated with choice confidence and the difference in utility between choice between alternatives within each choice tasks.

Figure 8.5: Correlations between the selection of indifference, average choice confidence and absolute utility difference between choice alternatives



Pearson's correlations	Average choice confidence score	Absolute utility difference	Number of indifference alternative selections
Average choice confidence score	-	0.67 (p=0.005)	-0.78 (p=0.000)
Absolute utility difference	0.67 (p=0.005)	-	-0.79 (p=0.000)
Number of indifference alternative selections	-0.78 (p=0.000)	-0.79 (p=0.000)	-

Associations with the selection of indifference and demographic and experimental characteristics

Logistic regression revealed a limited number of associations between selection of the indifference alternative and sociodemographic characteristics (Table 8.5). Significant associations primarily related to task-based characteristics (e.g. task difficulty) as opposed to sociodemographic factors (e.g. education). The likelihood of a respondent selecting the indifference alternative reduced incrementally as the number of times a respondent selected the opt-out ("neither") alternative increased (OR=0.93). Respondents who reported finding the task difficult or very difficult were significantly less likely to express indifference (OR= 0.61) perhaps due to an increased effort to differentiate between utility-balanced choice tasks. Somewhat counterintuitively, individuals who indicated taking an active role in their medical decisions were over

50% (OR=1.53) more likely to select the indifference alternative at least once throughout the survey compared to those who adopt a more passive role during GP consultations. Finally, as expected indifference reduced as both self-reported choice confidence (OR=0.84) and the difference in utility between test alternatives increased (OR=0.85).

Table 8.5: Logistic regression used to investigate associations between indifferent preferences and demographic and experimental characteristics

	Full model	**Reduced model**
	OR (std. error)	**OR (std. error)**
Attended university	0.77 (0.18)	
Very good/good health	0.68 (0.17)	
Previously tested for OC	1.01 (0.07)	
Currently have an active role in medical decisions	1.54* (0.41)	1.53*** (0.18)
Desire an active role in medical decisions	1.22 (0.31)	
Symptom awareness	0.98 (0.03)	
Task difficult/very difficult	0.61** (0.15)	0.61*** (0.08)
Total number of optout selections	0.92* (0.05)	0.93** (0.03)
Choice confidence	0.84*** (0.05)	0.84*** (0.03)
Difference in utility between test alternatives	0.85*** (0.03)	0.85*** (0.03)
Constant	0.38* (0.22)	0.27*** (0.06)
Model fit statistics		
Log-pseudolikelihood	-1294.15	-1304.30
Pseudo R2	0.04	0.04
Observations	7,200	7,200
N	450	450
***=1% significant; **= 5% significant; *=10% significant		

8.3.4.2 How does the inclusion or exclusion of an indifference alternative impact the quality and/or reliability of results?

Table 8.6 compares several dimensions relating to the completion of DCE tasks between the two survey versions (with and without an indifference alternatives).

Independent t-tests or chi2 tests were used to identify statistical differences between versions. Overall differences between the two surveys were minimal across all compared dimensions.

Stability failures (i.e. inconsistent choices across repeated questions) displayed the largest absolute difference, with inconsistent choices doubling in the versions which included an indifference alternative.

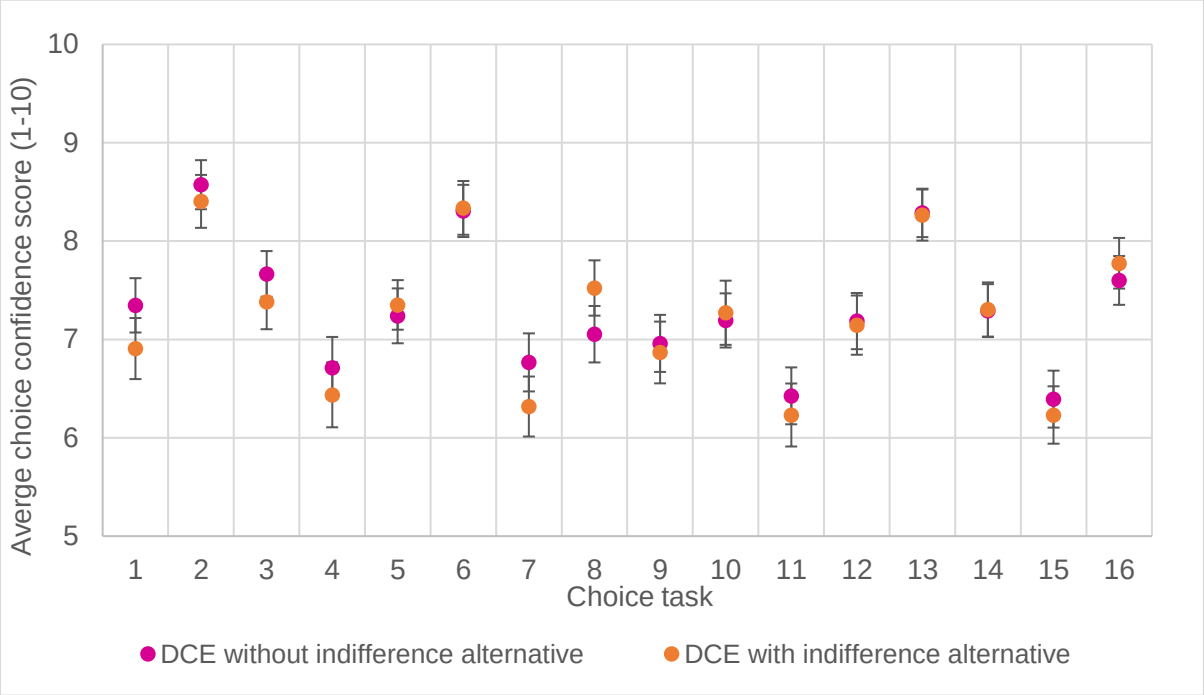
Selection of the "neither" opt-out alternative was slightly higher where the "indifference" alternative was included both in terms of the number of respondents and the number of overall selections, however, differences were not significant. Similar (and even reduced) rates of selection suggest it is unlikely that respondents utilised the opt-out alternative to indicate indifference where an indifference alternative was provided.

Table 8.6: Comparison in response characteristics across DCE versions including and excluding an indifference alternative

	Indifference alternative included	Indifference alternative excluded	p-value[a]
Rationality failures			
Monotonicity failure	0 (0%)	0 (0%)	-
Stability	30 (20%)	15 (10%)	0.02**
Transitivity failure	3 (2%)	4 (3%)	0.70
Stated attribute non-attendance			
Accuracy	16 (11%)	16 (11%)	-
Time to diagnosis	68 (45%)	65 (43%)	0.73
Identifiable conditions	52 (35%)	52 (35%)	-
Communication	95 (63%)	88 (59%)	0.41
Opt-out alternative selections			
Total selections	136/2400 (6%)	110/2400 (5%)	0.89*
Number of respondents	37/150 (25%)	31/150 (21%)	0.41
Self-reported task difficulty			
Very easy	1 (1%)	4 (3%)	
Easy	44 (29%)	49 (33%)	
Neither easy or difficult	29 (19%)	44 (29%)	0.44
Difficult	69 (46%)	50 (33%)	
Very difficult	7 (5%)	3 (2%)	
[a]p-value from an independent t-test used to test for statistical differences across versions. Chi2 test was used to compare self-reported task difficulty responses			

Figure 8.6 shows the average confidence score for each choice task where a higher score indicates higher confidence in their chosen response. Confidence scores were generally comparable across the two versions. Respondents exhibited significantly higher confidence in their choice in the DCE with an indifference alternative on three occasions (tasks 1,7 and 16).

Figure 8.6: Average choice confidence for each choice task for each DCE version. Choice confidence scores were statistically difference for tasks 1, 7 and 16 only.



8.3.4.3 Does failure to allow indifferent responses during DCE tasks influence unobserved variance of results?

The heteroscedastic logit model (Appendix 8.5) found no significant differences in scale between the DCE with and without an indifference alternative. This suggests that the inclusion or exclusion of an indifference alternative had no effect on the stochasticity of decisions between responses.

8.3.4.4 How do marginal rates of substitution vary between DCEs that include an indifference alternative and those that do not?

Sociodemographic characteristics of the additional survey version excluding the indifference alternative are shown in Appendix 8.6. Statistical comparisons of the samples for the two surveys with and without indifference indicated the samples were largely demographically similar. Significant differences were found in two characteristics, knowing someone with ovarian cancer and personal testing history—neither of which appeared to be a source of preference heterogeneity in Chapter 7. All other characteristics did not significantly differ between the two samples.

Willingness to wait estimates from the two survey versions are shown in Table 8.7. Estimates associated from the version with indifference are the same as those presented in Chapter 7 but are repeated here for comparison. The associated mixed logit output for the non-indifference survey version is provided in Appendix 8.7. Independent T-tests revealed no significant differences in the willingness to wait estimates for any attribute between survey versions.

Table 8.7: Comparison of willingness to wait estimates between the DCEs with and without an indifference alternative

	Indifference alternative included	**Indifference alternative excluded**	p-value*
Accuracy			
Per 1%	0.29 (0.22–0.33)	0.23 (0.19–0.26)	0.13
Identifiable conditions			
Cancer only	-	-	-
Cancer plus additional related conditions	1.55 (1.16–1.95)	1.51 (1.21–1.81)	0.85
Communication			
Poor	-	-	-
Fair	1.57 (1.19–1.95)	1.67 (1.34–2.00)	0.69
Good	1.99 (1.53–2.44)	2.13 (1.73–2.53)	0.64

*p-value from independent t-tests to investigate differences in marginal rates of substitution between survey versions

8.3.5 Discussion

This methodological extension adds to the currently limited evidence base examining indifferent preferences in healthcare discrete choice experiments (Medina-Lara *et al.*, 2014; Robinson *et al.*, 2015). The impact of including or excluding indifference alternatives within DCE choice tasks was examined across several domains including rationality failures, stated attribute non-attendance, error variance across responses and willingness to wait estimates. Expressions of indifference towards alternatives within this study, observed rates of indifferent preferences were frequent (5% of all responses in DCE versions 1-3 in Chapter 7). High rates of indifference may be reflective of the diagnostic context of the experiment, where people are used to deferring to their doctor for decisions. Overall, results from this study suggest little-to-no difference in the reliability and validity of responses or the outcomes of DCE based on the accommodation of indifferent preferences. This finding directly contrasts previous studies from the transportation literature where the inclusion of indifference choice option was found to improve model fit and performance (Bahamonde-Birke *et al.*, 2017; Cantillo *et al.*, 2010; Pan & Zuo, 2020).

Encouragingly results suggest that whilst the inclusion of indifference does not appear to affect model outputs in this study, it also did not appear to have a detrimental impact by increasing the perceived task difficulty or introducing noise in the data due to added complexity of an additional alternative presented in each choice task.

Further empirical studies are needed to draw an overall conclusion on the appropriateness of including or excluding indifference alternatives from DCEs in healthcare settings. However, utilisation of indifference alternatives within DCEs remains uncommon in healthcare, despite leading best practice guidance explicitly recommend allowing the capacity to express indifference between alternative during the completion of choice tasks (Bridges *et al.*, 2011).

One clear barrier to the inclusion of indifference alternatives is a lack of guidance on how to best incorporate indifference alternatives into the analysis of choice models. For simplicity, indifferent responses within this thesis were excluded from analysis. However, the development of transparent and user-friendly guidance on indifference

alternative modelling is an area of future development relating to this PhD research. Existing approaches within the transport literature are based on the psychological notion of just noticeable differences (JND) (Coombs *et al.*, 1970). JND stipulates the existence of perception thresholds, below which differences between items of comparison become imperceptible to an individual (Georgescu-Roegen, 1958). Choices between alternatives under this threshold are assigned stochastically to an alternative in forced choice situations (Krishnan, 1977). First introduced by Cantillo *et al.* (2010), the utility-difference threshold approach models indifference through the inclusion of an indifference threshold parameter δ, the absolute value of which represents the minimum level of utility difference between alternative that must be achieved to allow individuals to distinguish between alternatives. However, these models are currently used sparingly due to the complexities of reproduction and also shortcomings when experiments include multiple non-participatory alternatives (i.e. indifference, opt-out and status quo alternatives) since current approaches are based on a binary decision format. The notion of an indifference threshold has further important implications within DCEs, particularly where researchers aim to make policy suggestions to increase demand for services. This is because the presence of an indifference threshold suggests that if the change in utility from a certain policy is too mild, the target audience may remain unaffected resulting in no change in uptake (Pan & Zuo, 2020).

It is important to note that the inclusion of an indifference alternative without a suitable method for analysing such responses may have important implications for the analysis and results. Indifferent responses were removed from the analysis (i.e. specified as missing data) during the analysis of DCE responses presented within this chapter and Chapter 7. This means for 127 (28%) respondents the full experimental design was not utilised resulting in a reduction in the information gained from responses. The lack of differences in model estimates between DCE versions with and without the indifference alternative is encouraging and suggests the exclusion of indifferent responses had little-to-no impact on the ability to identify the parameters within the model. However, missing data reduces the efficiency of any experimental design resulting in diminished precision in model estimates and may mean larger sample

sizes are needed to improve the precision surrounding model estimates (Johnson *et al.*, 2013; Louviere & Lancsar, 2009).

8.4 Chapter summary

This chapter used data collected during the DCE eliciting preferences towards diagnostic testing for ovarian cancer to explore methodological issues relating to the application of DCEs to cancer testing and healthcare more broadly. Namely, the impact of indifferent preferences, and the relationship between the number of choice tasks and learning and fatigue. Both concepts have been explored within the discrete choice literature more broadly, however, explorations within the field of health are more limited. Research in this chapter provides a case study of how these methodological issues apply to a health-focused experiments. It should not be assumed that findings from other fields are automatically transferrable due to differences in the decision-making context. For instance, health focused DCEs often value non-market goods, meaning the formulation and expression of preferences may be less intuitive to respondents. Furthermore, in this instance DCEs of diagnostic testing require the consideration of risks and complex medical concepts increasing the complexity of decisions. As such, further health-based evidence is needed on both these topics to draw a conclusive verdict on both of the topics explored in this chapter.

9 Preferences towards ovarian cancer screening: DCE development

9.1 Introduction

The examination of diagnostic preferences in Chapter 7 demonstrated high rates of willingness to be tested once symptoms of ovarian cancer arise. However, poor symptom awareness and the non-specific nature of symptoms continue to present barriers to early diagnosis. These difficulties relating to the early recognition of ovarian cancer mean efforts to identify an appropriate ovarian cancer screening test continue despite trials demonstrating little-to-no benefit to date. Chapter 3 highlighted the importance of considering prospective users preferences when designing any national screening programme. To date, research on preferences for ovarian cancer screening has typically focused on the acceptability of alternative modalities from the perspective of patients enrolled on screening trials (Drescher *et al.*, 2004; Jenkins *et al.*, 2015; Pavlik *et al.*, 1995). However, evidence suggests these individuals are more likely to have a favourable view of screening and exhibit a willingness/desire to be screen regardless about beliefs surrounding the efficacy of screening meaning results may not be representative of the wider public (Gigerenzer *et al.*, 2009; Hoffmann & Del Mar, 2015). Evidence relating to the preferences of general public, average-risk individuals is more limited despite being the target population for a potential screening programme. In response, the purpose of the second DCE in this thesis was to investigate women's preferences and demand for a potential screening programme. This chapter describes the process of developing an online survey with an embedded DCE aiming to elicit preferences towards ovarian cancer screening. The development process utilises the framework outlined in Chapter 6 when designing the diagnostic DCE with a few modifications.

9.2 Aims

The aim this chapter is to describe the design and pilot testing of an online survey with an embedded DCE to measure women's preferences towards a test which could be potentially used to screen for ovarian cancer before symptoms arise.

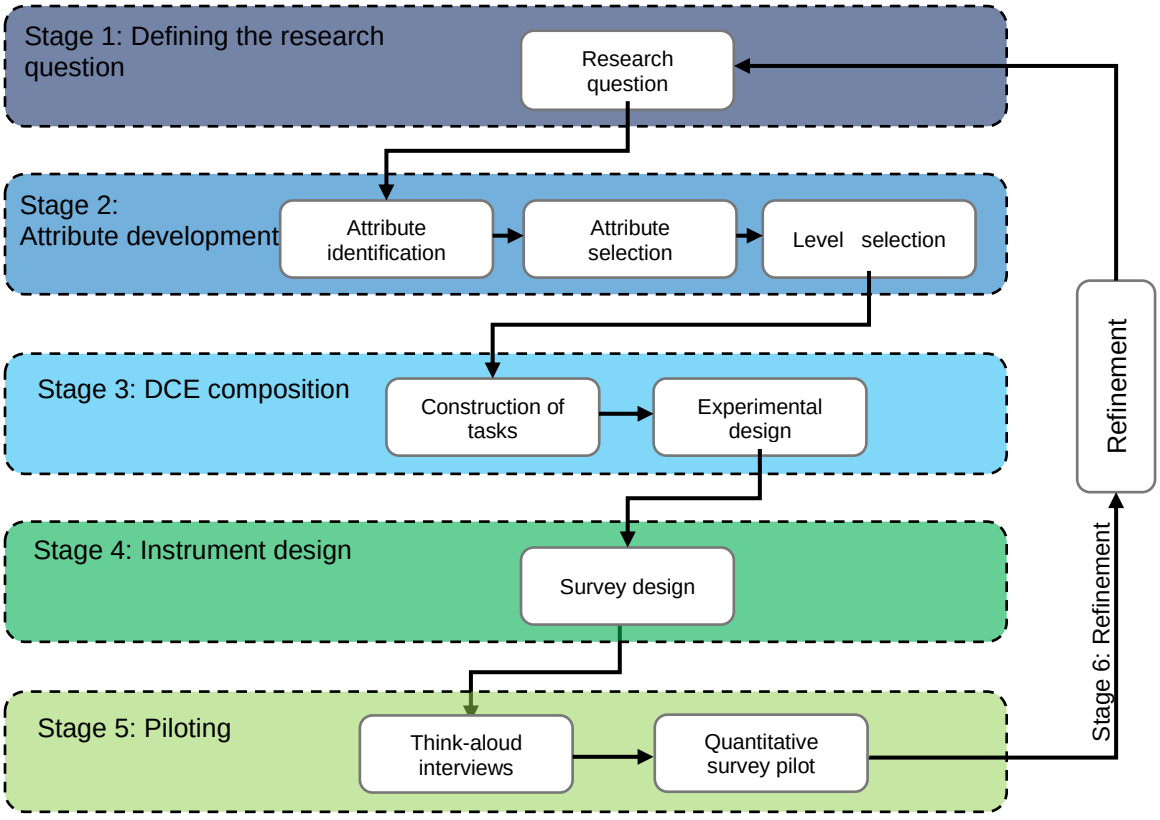
The objectives were:

- To consolidate evidence on the most important attributes to the target population when considering ovarian cancer screening tests
- To develop a discrete choice experiment instrument using leading guidance
- To develop an appropriate accompanying online questionnaire to collect relevant demographic data to allow relationships between personal characteristics and choice behaviour to be explored in the final study
- To test the feasibility and acceptability of the online DCE survey instrument to the target population through pilot testing
- To refine and finalise the survey instrument through an iterative process based on piloting feedback

9.3 DCE development

DCE development followed a similar process used during the development of the DCE on diagnostic testing in Chapters 5 and 6. The approach was adapted based on learnings from this earlier study and to accommodate screening-specific challenges. A reminder of the development stages in provided in Figure 9.1.

Figure 9.1: Summary of the discrete choice development process.



9.3.1 Stage 1: Defining the research question

A screening programme for ovarian cancer would provide an alternative solution to the ongoing barriers to early diagnosis currently experienced in primary care settings (e.g. late presentation, vague symptoms). Screening is reliant on the voluntary participation of asymptomatic individuals meaning the acceptability of tests to the target population is fundamental. Despite continued efforts to develop an effective screening programme for ovarian cancer research on the acceptability of any potential screening programme to female members of the public has been less explored to date.

In response, the purpose of the DCE developed in this chapter was to estimate women's preferences towards a hypothetical screening programme for ovarian cancer.

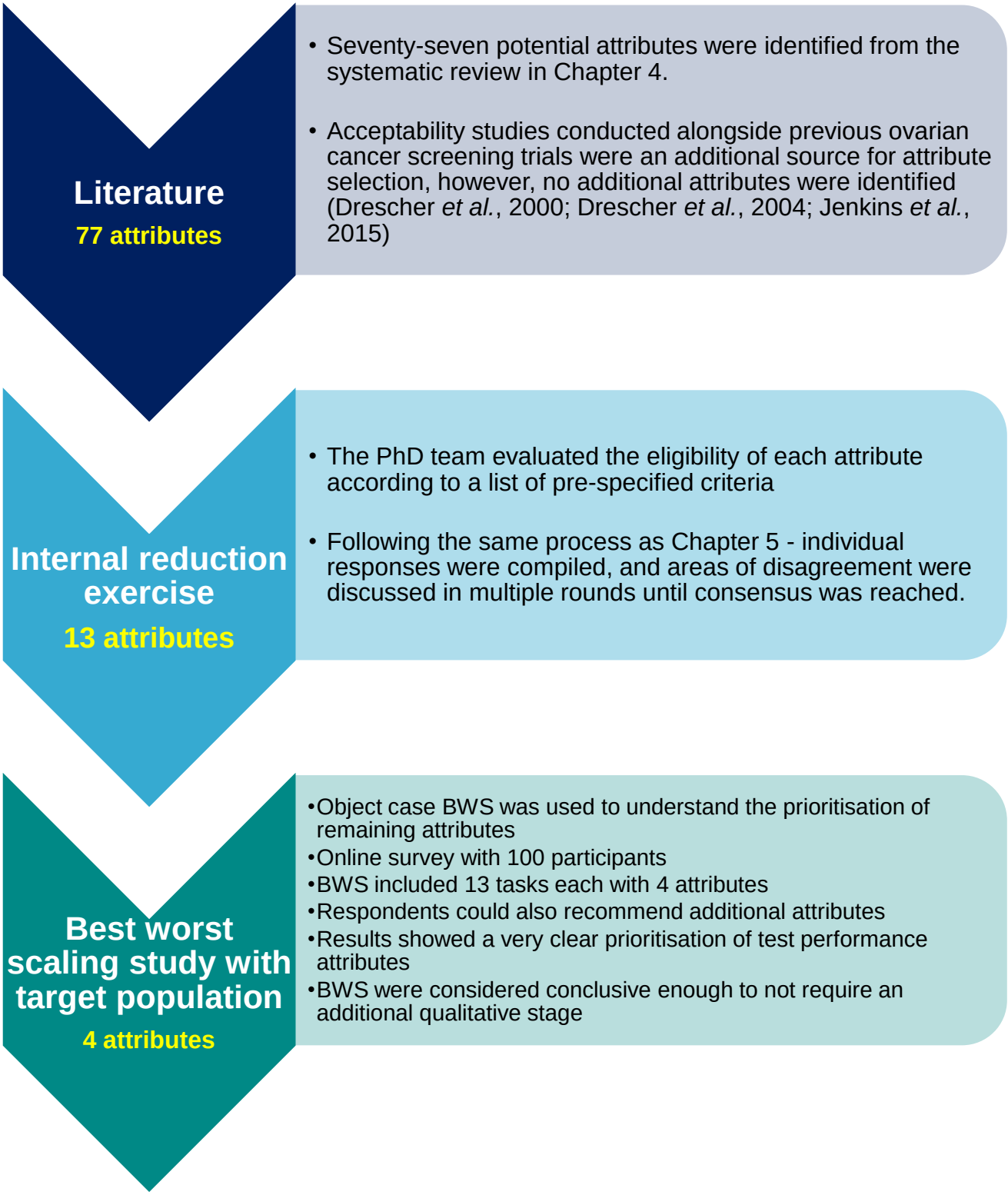
The acceptability of a screening programme may be driven by multiple factors relating to both test performance affecting the perceived benefits of testing (e.g. mortality reduction, false positives, false negatives etc.) and service delivery factors affecting the convenience and experience of undergoing testing (e.g. screening interval, screening modality, location etc.). Due to limited existing evidence, it is unclear which aspects of testing would be most important women and the driving force behind the acceptability of any potential programme. As discussed in previous chapters, the ability to combine attributes from both aspects of acceptability is a fundamental strength of DCE and motivation for the use of the methodology to explore preferences throughout this thesis.

To allow a full exploration of the driving factors of screening decisions, research aims at the beginning of the development process were kept broad with no preconceptions of the attributes to be included. The aim was to design a DCE capable of estimating demand for a potential future screening programme and understand the characteristics of a test may influence demand. (i.e. what are the key criteria to produce a publicly accepted screening test?).

9.3.2 Stage 2: Attribute development

Attribute development largely followed the same process as the earlier diagnostic DCE. Attribute selection adopted an exploratory approach aiming to establish the most influential attributes relating to the acceptability of ovarian cancer screening. This approach was chosen to maximise the relevance of the final attributes to the research question aiming to understand preferences and demand for hypothetical/future screening programmes. The process is summarised in Figure 9.2.

Figure 9.2: Overview of the attribute selection process



9.3.2.1 Identification of attributes

The starting point for attribute development were the previous screening DCEs identified during the systematic review chapter. This evidence was combined with studies relating to the acceptability of screening, typically conducted alongside clinical trials (Drescher *et al.*, 2000; Drescher *et al.*, 2004; Jenkins *et al.*, 2015). In total, seventy-seven potential attributes were identified.

9.3.2.2 Internal reduction exercise

Potential attributes were initially narrowed down by the research team via a reduction exercise following the same structure undertaken in Chapter 5. The criteria used to disqualify attributes were:

i. Relevance to tests which may be feasibly used to screen for ovarian cancer
ii. Relevance to the setting (i.e. England and Wales)
iii. Relevance to the population (i.e. people with ovaries over 40)
iv. Overlap between attributes
v. Attributes must be quantifiable
vi. Attributes must be controllable and/or modifiable in context of ovarian cancer screening

To moderate the influence of research biases the internal reduction exercise was not final and were subject to change based on the remainder of the attribute selection process. Following the reduction process, thirteen potential attributes remained (Table 9.1). Attribute wording and definitions were adapted from existing screening DCEs and patient-facing documents (NHS UK, 2021a) and were assessed for clarity and understanding by two members of the public.

Table 9.1: Remaining candidate attributes following the internal reduction exercise. Attribute and descriptions were further narrowed down using BWS

Attribute	Description
- Type of test	What kind of test you will have. This could be a blood test, internal pelvic ultrasound scan and/or pelvic examination
- Accuracy: chance of **false negative** result	Chance that the test will miss cancer in a patient who actually does have the disease
- Accuracy: chance of a **false-positive** result	Chance of receiving an abnormal or "positive" result when there is no cancer actually present. This will likely cause unnecessary worry and will mean undergoing extra tests
- Chance of needing a follow up test	A follow up test may be needed if you have an abnormal or unclear result. Only a small proportion of these people will actually be diagnosed with cancer
- Chance of dying from ovarian cancer	How much does having the test reduce the chance of dying from ovarian cancer
- Chance of cancer diagnosis	The chance that you will be diagnosed with cancer
- Chance of being unnecessarily diagnosed and treated for a cancer that would never have caused symptoms or death	This is known as overdiagnosis. An overdiagnosed person has cancer but will never have symptoms and will die of other causes. For these people treatment may do more harm than good.
- Screening interval	How often you will be invited for a test. This could range from once in your lifetime to once a year
- Test location	Where the test takes place (e.g. local GP surgery, hospital or specialist centre)
- Action required by you to arrange the result	How you make the appointment- e.g. pre-assigned appointment time, ring, online
- Who performs the test?	How experienced the healthcare professional preforming the test is with the procedure.
- Waiting time for the test	How long you have to wait to have a test once you have been invited for screening
- Waiting time for the result	How long you have to wait to receive your results after having the test

9.3.2.3 Best worst scaling study with the target population

Object-case best-worst scaling (BWS) was used to understand the relative importance of the remaining thirteen potential attributes to the target population (Table 9.1). This section provides an abbreviated version of the methods and results of the BWS study. A detailed summary is provided in Appendix 9.1.

Methods

Survey development

A BIBD was generated in SAS 9.4. The design consisted of 13 choice tasks. Each choice task included a sub-set of four attributes from which participants were asked to select the "most important" and least important". Each attribute appeared four times across the choice tasks. Overall, the experimental design had a d-efficiency of 81.3%. An example choice task is shown in Figure 9.3.

The BWS study was embedded into an online survey hosted on Limesurvey (Limesurvey.com). A full version of the questionnaire is provided in Appendix 9.2. Data was collected in May 2021.

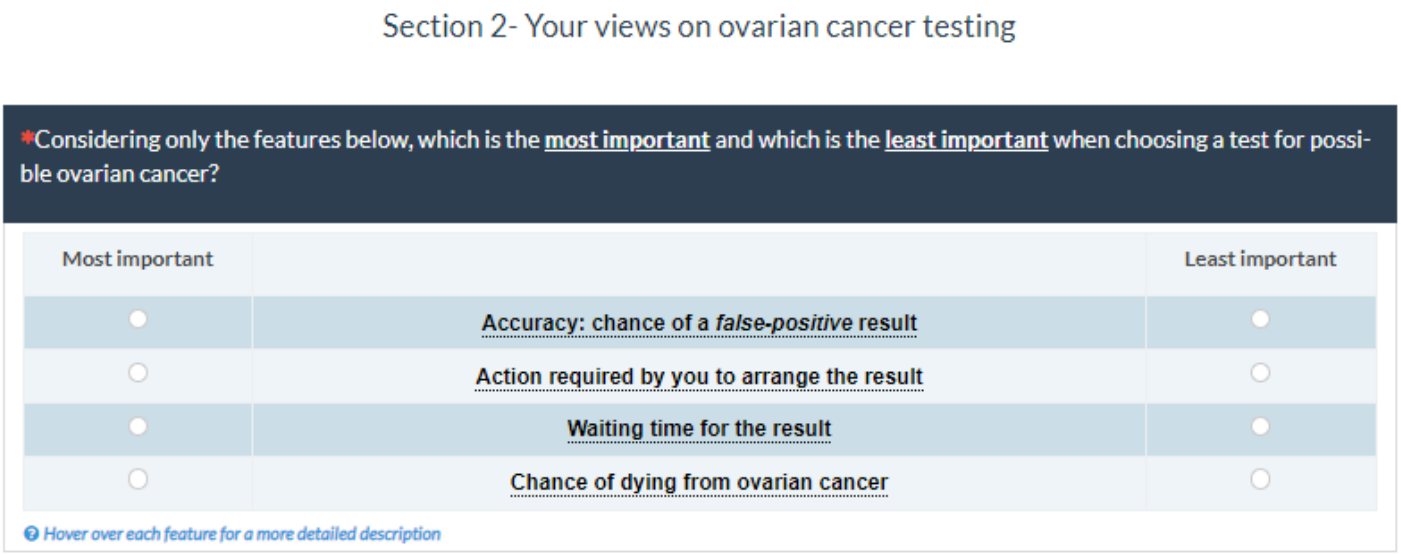
To ensure no relevant attributes were missed and mediate the influence of the internal reduction exercise, an open-ended question was included asking participants if there were any additional attributes they would consider important when making a decision about ovarian cancer screening.

Participants

Participants were recruited using Prolific (Prolific.co). Participation was limited to women and people with ovaries over the age of 40 (no upper limit), living in England and Wales. The minimum sample size estimated alongside the BIBD was 52, given the uncertainty surrounding this estimated a target sample size of 100 was chosen[9].

[9] When generating BIBD in SAS an estimated minimum sample size is provided, however, it is unclear how this estimate is derived. Currently there is no guidance on appropriate sample size for best-worst scaling studies.

Figure 9.3: Example of a choice task from the best-worst scaling study



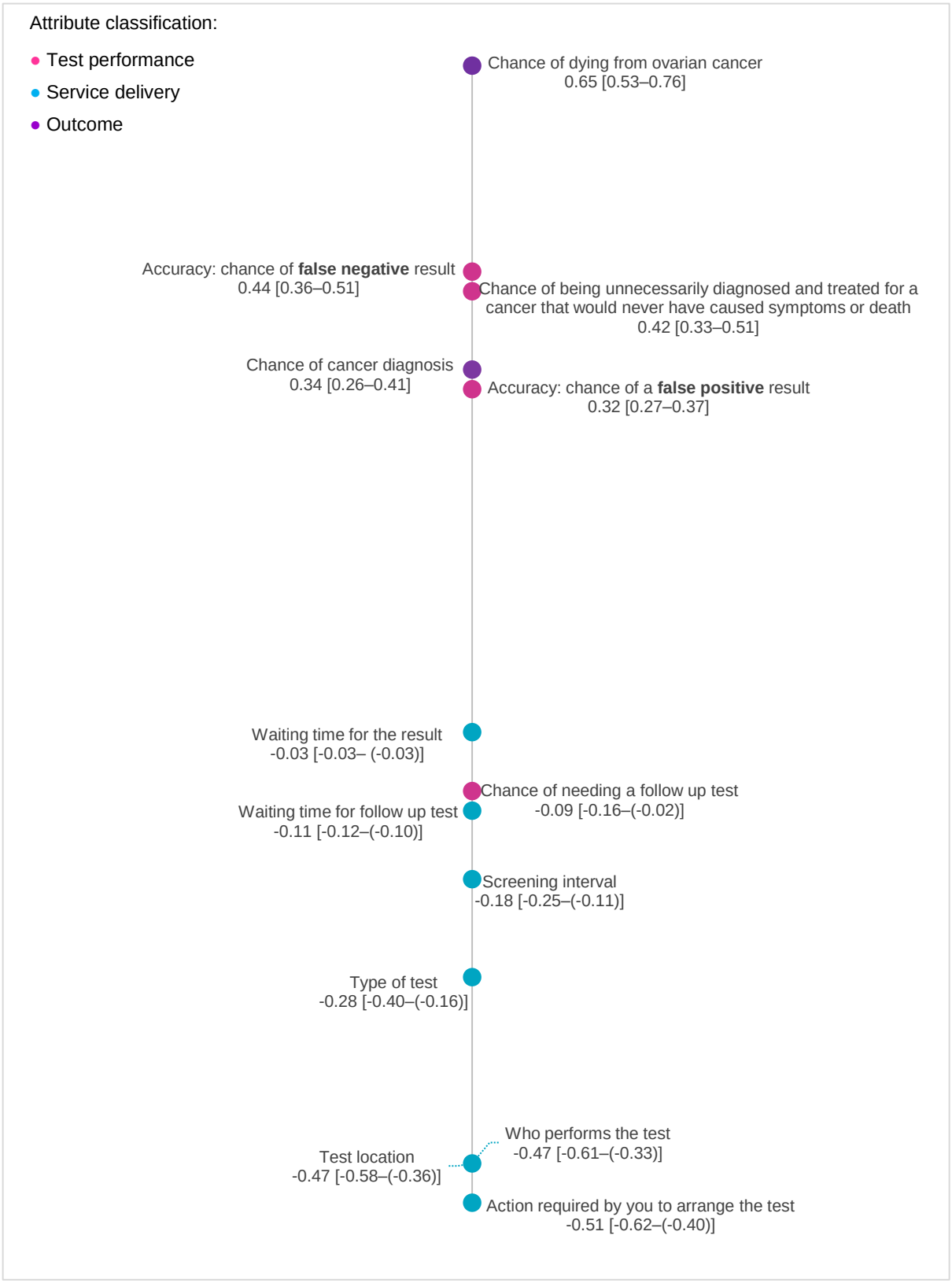
Analysis

BWS data were analysed using the counting approach. Raw counting scores ranged between -400 and +400 (4 attribute appearances x 100 respondents), with a higher score indicating greater importance. To aid interpretation, scores were standardised to between -1 and +1.

Results

Counting results are shown on a spatial scale in Figure 9.4. Importance scores formed two distinct groups. Attributes in cluster 1 were the most important to respondents are related to the performance characteristics of tests. Attributes in the second cluster were distinctly less important to respondents and related to service delivery aspects of screening (except for “chance of needing a follow up test”).

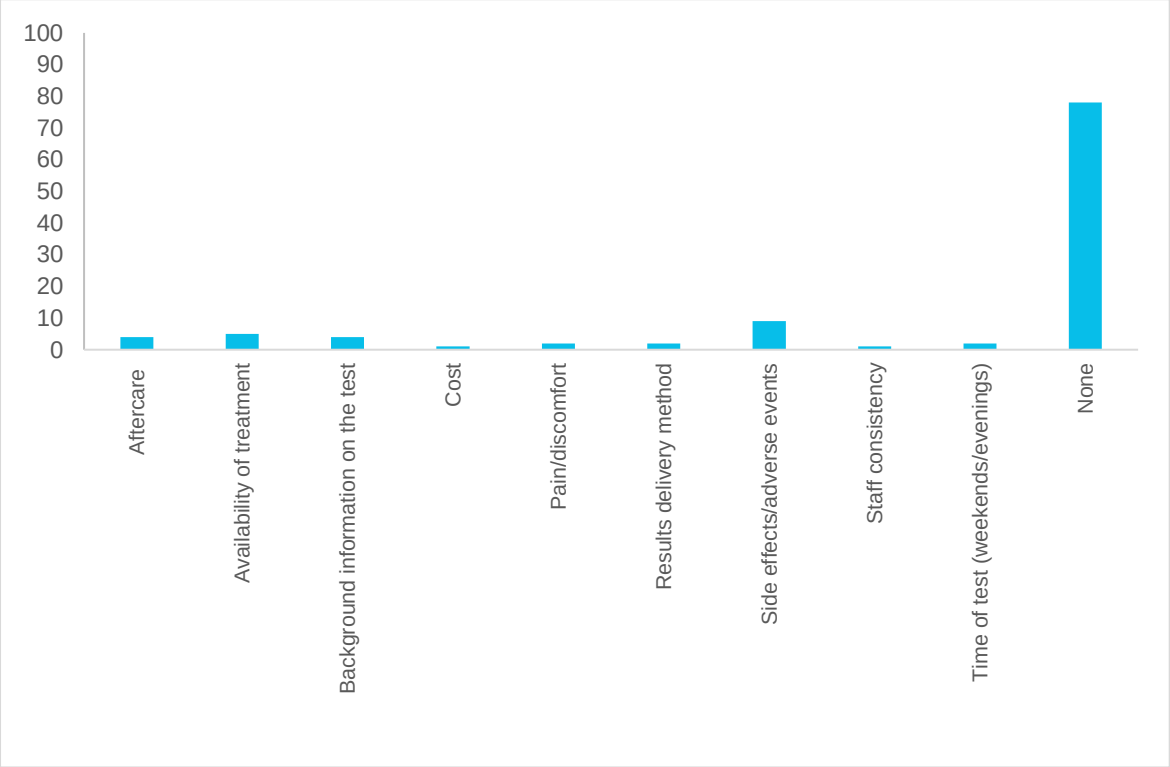
Figure 9.4: Best-worst scaling counting analysis results



Additional attributes

Seventy-eight of the 100 participants did not specify any additional attributes that would be important to their decision to screened. Suggestions of additional attributes provided by the remaining participants were categorised into nine general themes (Figure 9.5). With exception of staff continuity, all additional suggestions had been recognised during attribute identification stage and ruled out during attribute identification based on the previously described criteria.

Figure 9.5: Additional attributes identified from the free text question inviting respondents to add any additional attributes they consider important



9.3.2.4 Interpretation and final selection of attributes

Best-worst scaling results demonstrated a stark separation between the importance of test performance characteristics and service delivery attributes. The large differences in preferences indicated it would not be appropriate to combine attributes from the two clusters in a single experiment as this would likely encourage non-trading behaviour and attribute non-attendance. Given the decisiveness of results, it was

decided further attribute selection methods such as the online workshops utilised within the diagnostic testing DCE were not necessary.

The attributes selected for the pilot study are shown in Table 9.2. Cancer risk was not included as an attribute despite its importance within the BWS study due to difficulties in determining individual-level risks within a general population sample. Potential side effects/adverse events were the most commonly suggested write-in attributes but was excluded during the internal attribute reduction exercise based on relevance to ovarian cancer since risks associated with candidate screening tests (blood tests and/or ultrasound scans) are low (Cancer Research UK, 2022). However, results from RCTs demonstrated significant risks of serious adverse events associated with confirmatory testing in secondary care following a false positive result (Menon *et al.*, 2021). The risk of adverse events was therefore incorporated into the description of the false positive attribute within the final survey rather than a separate attribute.

9.3.2.5 Level assignment

Levels for false positive and false negative attributes were assigned based on evidence from two large-scale ovarian cancer screening randomised controlled trials; The Prostate, Lung, Colorectal and Ovarian Cancer Screening Trail (PLCO) and The United Kingdom Collaborative Trial of Ovarian Cancer Screening trial (UKCTOCS). The PLCO trial was based in USA and included 78,000 women (Buys *et al.*, 2011; Pinsky *et al.*, 2016). Participants in the intervention arm underwent an annual CA125 blood test and transvaginal ultrasound for a period of 3-5 years with a median follow up period of 14.7 years. The UKCTOCS trial took place in the UK and recruited over 200,000 participants (Jacobs *et al.*, 2016; Menon *et al.*, 2021). Participants in the intervention arm received either an annual CA125 test interpreted using the Risk of Ovarian Cancer Algorithm (ROCA) or an annual TVUS. Median follow up time was 16.3 years.

Mortality from ovarian cancer in the control arm of the UKCTOCS trial was 0.4% (Menon *et al.*, 2021). This figure was used to describe the consequences of no screening. Results from both trials suggest that current approaches to ovarian cancer screening have no impact on deaths. A reduction in mortality is the minimum

requirement for any potential screening programme, therefore levels assuming a 25%, 50% and 75% reduction in mortality were assigned.

Overdiagnosis was estimated to be 28% for type II tumours and 72% for all other ovarian tumours. However, the definition of overdiagnosis differed from the DCE attribute and was defined as cancers that would not have been detected without screening (number of tumours detected during the follow-up period/number of screen-detected tumours). The UKCTOCS trial did not formally measure overdiagnosis however, authors speculated overdiagnosis was not an issue since ovarian cancer incidence did not significantly differ between treatment arms and due to the often-aggressive nature of the disease. Given the uncertainty in overdiagnosis in ovarian cancer and challenges in the quantification of overdiagnosis in cancer screening more broadly (Carter *et al.*, 2015) , a range of 0-25% was chosen for the DCE. This figure was based on current evidence surrounding breast cancer screening (Bulliard *et al.*, 2021; Puliti *et al.*, 2012). Percentages were converted into numeric values based on a prevalence of 65 cancers per 10,000 people without screening.

Table 9.2: Attributes and levels for the screening DCE

Attribute wording	Definition	Levels
Ovarian cancer deaths	The number of people who will die of ovarian cancer over the course of 10 years	10 per 10,000 people screened 20 per 10,000 people screened 30 per 10,000 people screened 40 per 10,000 people screened (No screening)
False positive results	The number of people who do not have cancer that will receive an incorrect positive result over the course of 10 years	0 per 10,000 people screened (No screening) 1,000 per 10,000 people screened 2,000 per 10,000 people screened 3,000 per 10,000 people screened 4,000 per 10,000 people screened
False negative results	The number of people with cancer who will receive an incorrect negative result over the course of 10 years	0 per 10,000 people screened (No screening) 3 per 10,000 people screened 7 per 10,000 people screened 10 per 10,000 people screened 13 per 10,000 people screened 16 per 10,000 people screened 20 per 10,000 people screened
Overdiagnosed cancers	The number of people who will be unnecessarily diagnosed and treated for cancer that would never have killed them or even caused symptoms over the course of 10 years	0 per 10,000 people screened (No screening) 3 per 10,000 people screened 7 per 10,000 people screened 10 per 10,000 people screened 13 per 10,000 people screened 16 per 10,000 people screened

9.3.3 Stage 3: DCE composition

9.3.3.1 Construction of choice tasks

Full or partial profiles: A full-profile format was chosen since the total number of attributes within each choice task was low (n=4).

Number of alternatives: Each choice task contained two generic test alternatives "Test A" and "Test B". A format comparing a single varying test alternative to a "no screening" alternative was considered given a primary research question is the demand for screening (i.e. the decision between testing and no testing). A two-profile approach was ultimately selected to maximise the trade-off information obtained from each task since it was anticipated that some respondents would always choose to be screened regardless of the attribute levels based on personal beliefs.

Opt-out and indifference alternatives:

Screening programmes rely on voluntary participation. People may choose not to be screened for an array of reasons such as fear, procrastination, inconvenience or low perceived benefits (Bennett *et al.*, 2018). Given the low prevalence of ovarian cancer and the potentially high rates of harms, it was considered important to include an opt-out alternative, particularly as estimating the demand for screening at different benefit-harm ratios was a fundamental research question. This study builds on the findings from previous chapter by refining the wording of the opt-out alternative in order to avoid any ambiguity. The opt-out alternative was labelled "No screening" and was assigned constant levels based on the control arm of a large-scale RCT (Menon *et al.*, 2021). There is no existing screening programme so a "status quo" option was also not considered (although arguably "no screening" is the current status quo).

Inclusion of an indifference alternative was considered, however, was ruled out in order to limit the complexity of the choice tasks. This also follows current standard practice within the screening literature. Chapter 8 demonstrated that the inclusion or exclusion of an indifference alternative did not alter estimates of preferences is the context of ovarian cancer testing, however, this is one of a few limited studies within a healthcare setting so results may still be susceptible to biases of excluding the optout alterative observed in wider DCE literature.

9.3.3.2 Experimental design

A full-factorial design would result in $3^1 \times 4^1 \times 6^2 = 432$ choice tasks and 93,096 potential paired combinations. Instead, an efficient fractional factorial design was

generated using Ngene 1.2 (Choice Metrics). An initial design used during piloting was generated using small directional priors. Several designs with different numbers of choice tasks were considered. A design using 12 tasks was selected for the pilot to allow level balance across all attributes. It was unclear at this stage if blocking would be necessary given the complexity and challenges associated with risk-based attributes. The design was therefore split into two blocks of 6 choice tasks as a precaution and respondent acceptability was tested during piloting. Following piloting, the design was updated to a Bayesian efficient design using pilot results as priors.

9.3.3.3 Choice task presentation: risk communication

How to best communicate the risk-based attributes within the DCE was an important consideration during the development stage. If risk information is not well understood or well-presented the validity of any findings is diminished (Harrison *et al.*, 2014). Risk communication is an ongoing and debated area of research meaning there is currently no standardised method (Trevena *et al.*, 2021; Zipkin *et al.*, 2014). The most appropriate risk communication is very situational dependent of the target population, intended purpose and magnitude of probabilities. Effective communication should ensure information is presented clearly to maximise understanding for the largest proportion of the target audience. Since the aim of the DCE was to measure underlying preferences, particular attention to ensure was the risk format was informative rather than persuasive was needed. However, the primary challenge in the context of ovarian cancer screening is the frequency of events. The prevalence of ovarian cancer is low (~7,500 new cases annually) compared to cancers where a screening programme currently exists such as breast cancer (~56,000 new cases annually) or colorectal cancer (~43,000 new cases annually) (Cancer Research UK, 2019). As a result, associated benefits and risks are also much lower meaning risk presentation formats used in screening decision aids particularly icon arrays are not directly transferable.

Before developing the risk communication format for the DCE, a list of specifications was compiled based on recent publications. A full explanation of the specifications and related evidence is provided in Appendix 9.3.

In brief, the specifications when developing the risk communication format were:

- Absolute risks not relative risks

Relative risks have been shown to manipulate or persuade audiences by magnifying risk perceptions and decreasing understanding (Akl *et al.*, 2011; Garcia-Retamero & Cokely, 2017; Zipkin *et al.*, 2014).

- Frequencies not percentages

Research suggests rare events (i.e. those occurring less than 1%) may be less well understood when represented as percentages due to the use of decimal points meaning simple frequencies are preferable (Trevena *et al.*, 2021).

- Consistent denominator for frequencies

To aid understanding and comparisons between alternatives uniformity in the size of the denominator is recommended both across and within attributes (Garcia-Retamero & Cokely, 2017; Garcia-Retamero & Galesic, 2009; Garcia-Retamero *et al.*, 2010).

- The population of interest should be the same across all attributes

In practice, this means attributes such as false negatives should be expressed in terms of the number of people screened taking into account disease prevalence rather than in terms of only those with the disease (i.e. approach used in the diagnostic study) (Garcia-Retamero & Cokely, 2017).

- Consistent framing of attributes

To avoid biases in the willingness to trade between attributes, a consistent framing perspective should be adopted. (Akl *et al.*, 2011; Michalovic *et al.*, 2018; Zipkin *et al.*, 2014)..

- Visual aids

To maximise the effectiveness of visual aids there are some universal guidelines that aid the interpretability of graphics:

a. visuals should be supplemented by numerical risks (Garcia-Retamero & Cokely, 2017; Okan *et al.*, 2015; Trevena *et al.*, 2021)

- ensure spatial features of visuals (e.g. height of bars, axis scales) are conventional and representative (e.g. avoid truncated scales) (Trevena *et al.*, 2021)
- Use incremental risk format showing the (risk with and without intervention displayed in same array
- Depict both positive and negative outcomes within the same visual (e.g. stacked bar charts or icon arrays showing outcomes for the entire population)
- assess the graph literacy of the target audience (Garcia-Retamero & Cokely, 2017)
- validate visual aids by conducting usability studies with the target audience before implementation (Okan *et al.*, 2015; Woller-Carter *et al.*, 2012).

Development process

To ensure risk presentation within the DCE was optimised for the most people four potential versions were designed based on the evidence from existing literature and adaptations of existing decision aids (Figures 9.6-9.9). The alternative versions were tested in an online questionnaire with 50 women over the age of 40. Participants were shown all four versions and asked which they preferred—specifically, which they found easiest to understand.

Figure 9.6: Risk communication option A: classic DCE format

Consequences of annual screening over the course of 10 years using different test strategies:

	Test A	Test B	No screening
Chance of dying of ovarian cancer	28 in 10,000 (0.28%)	35 in 10,000 (0.35%)	42 in 10,000 (0.42%)
Chance of receiving a false-negative result	4 in 10,000 (0.04%)	6 in 10,000 (0.06%)	0 in 10,000 (0%)
Chance of receiving a false-positive result	1986 in 10,000 (20%)	1650 in 10,000 (17%)	0 in 10,000 (0%)
Chance of being overdiagnosed	7 in 10,000 (0.07%)	10 in 10,000 (0.1%)	0 in 10,000 (0%)

Figure 9.7: Risk communication option B: Expected frequency tree

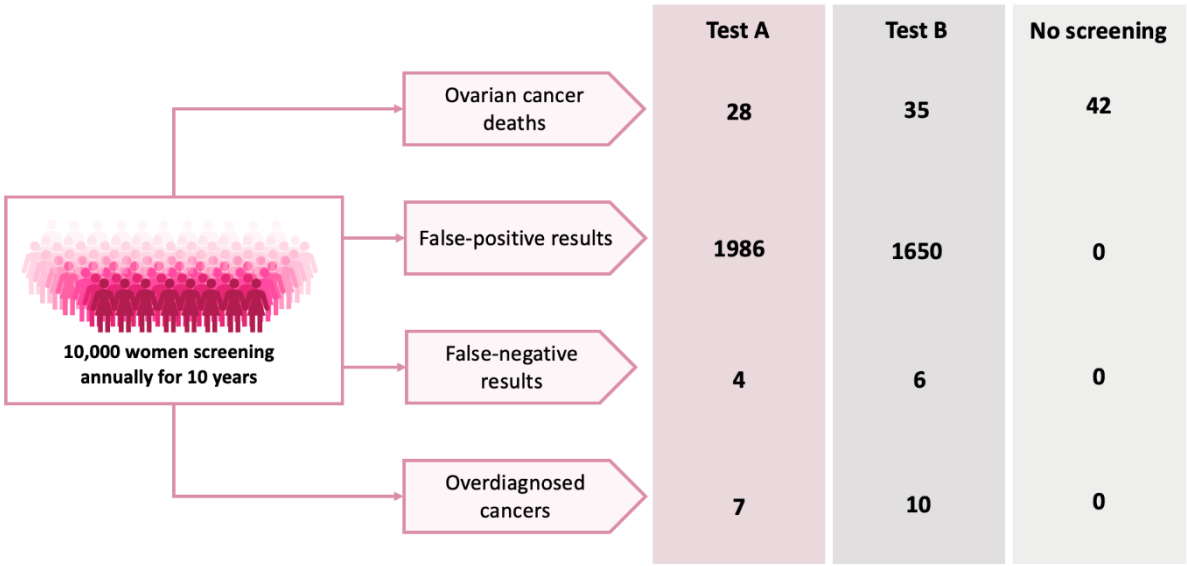


Figure 9.8: Risk communication option C: Icon array

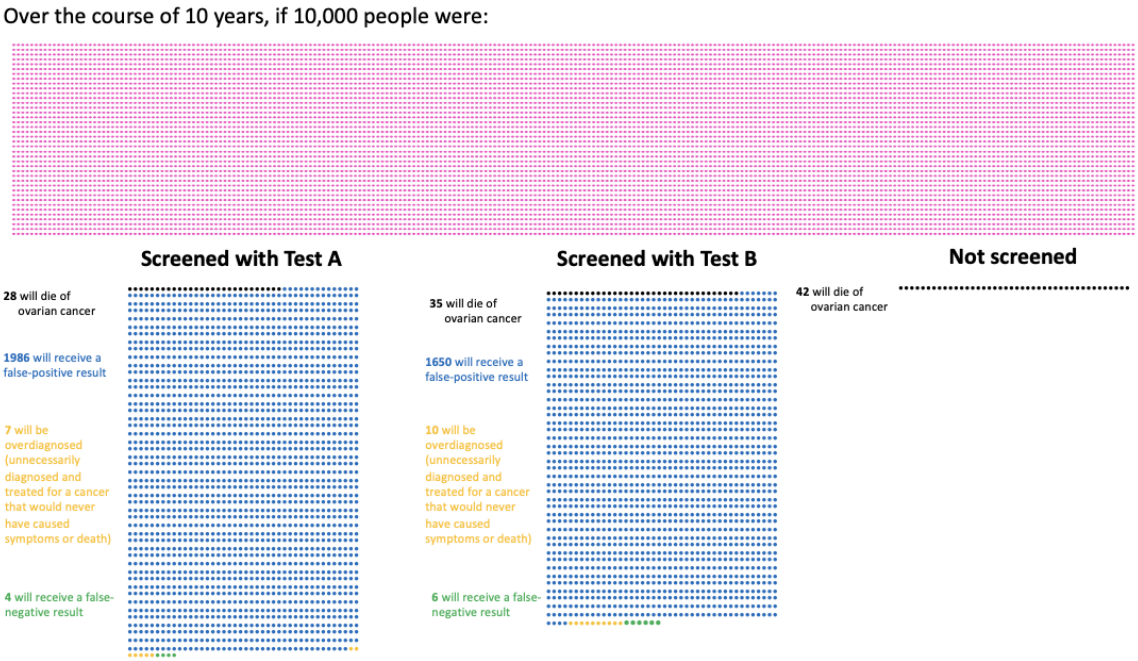
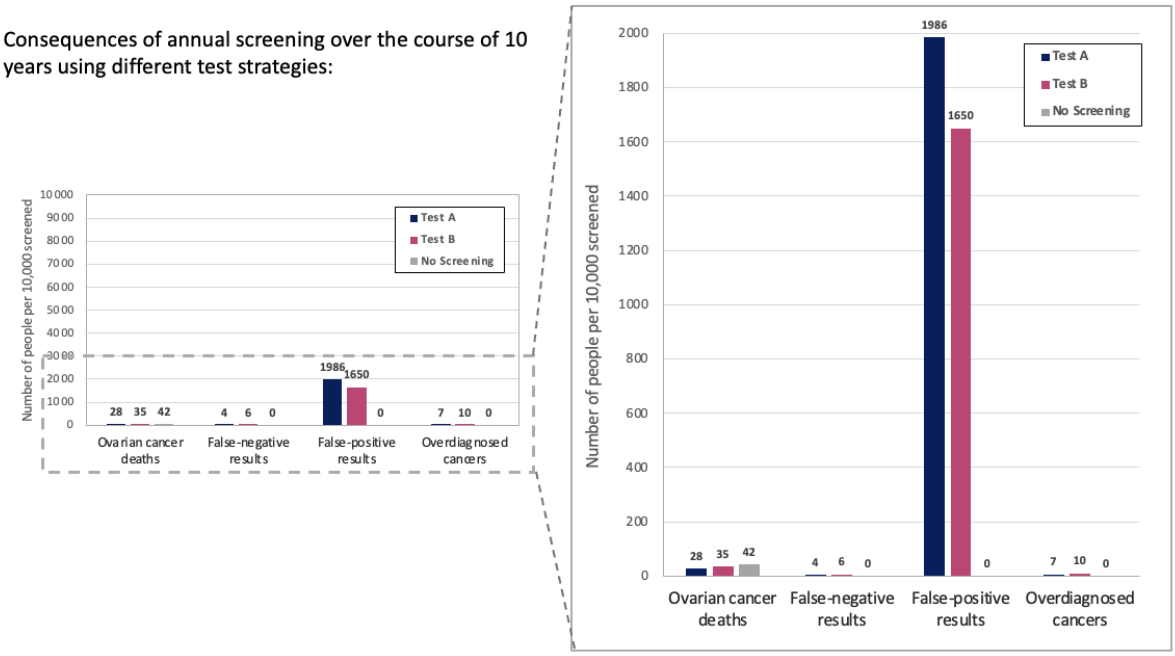


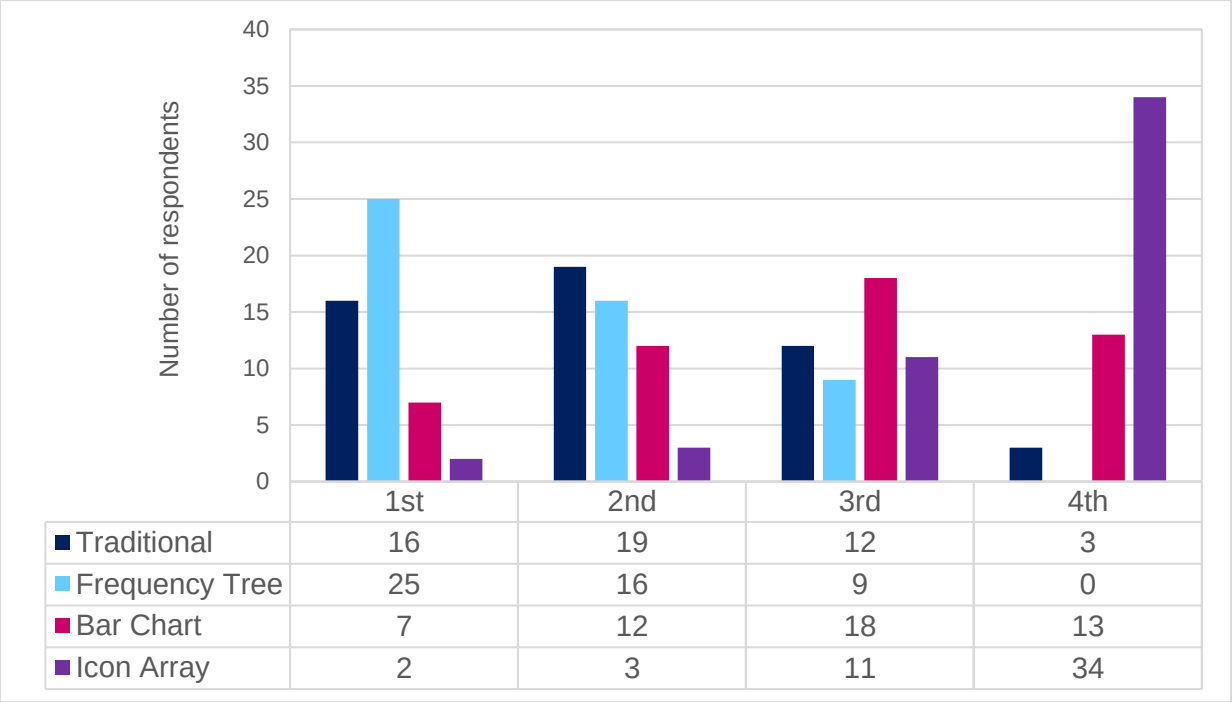
Figure 9.9: Risk communication option D: Bar chart



Results from the questionnaire comparing risk formats are shown in Figure 9.10. The adapted decision tree format was the most preferred option overall (mean ranking: 1.87), followed by the traditional DCE format (mean ranking: 2.27). Bar chart (mean score: 3.04) and icon array (mean score: 3.93) formats were ranked lowest overall.

Based on these results the decision tree format was selected for the DCE. This format benefits from being comparable to the traditional DCE format with a few modifications. No graphics are used therefore differences in graph literacy across the target population do not present a challenge.

Figure 9.10: Ranking results for the alternative DCE formats (n=50)



	1st	2nd	3rd	4th
Traditional	16	19	12	3
Frequency Tree	25	16	9	0
Bar Chart	7	12	18	13
Icon Array	2	3	11	34

9.3.4 Stage 4: Instrument design

The discrete choice experiment was embedded in an online survey hosted on Limesurvey (Limsurvey.org). Given the cross-over in topic, the survey shared many similarities with the diagnostic testing DCE. A summary of key changes made to the survey is provided in Table 9.3. The wording of the survey was checked by two members of the public and further refined following the piloting stage.

Table 9.3: Summary of changes to the survey instrument

Added	Removed	Explanation
Background information		
Additional information on the format of DCE questions and instructions on how to complete choice tasks		A significant proportion of respondents in the diagnosis DCE reported finding the task difficult or very difficult. Additional instructions were added to help familiarise respondents with the DCE tasks prior to the start of the questions
Sociodemographic questions		
No changes		
Health related questions		
Two questions relating to screening behaviour for breast and cervical cancer. Questions were tailored to the age of the respondent based on screening guidelines		It is anticipated that current screening behaviour will be linked to intentions to undergo additional screening for ovarian cancer
	If you had a symptom that you thought might be a sign of ovarian cancer, how long would it take you to go to the doctors from the time from first noticed the symptom?	This question was not relevant to screening behaviour
	How much confidence and trust do you have in general practitioners (GPs)?	This question was not relevant to the screening context

	When seeking help for medical issues, how much do you wish to be able to be involved in decisions about the medical process?	This question was not relevant to the screening context
	When seeking help for medical issues, how much do you feel able to be involved in medical decisions about the medical process?	This question was not relevant to the screening context
	BFI-10 used to measure personality traits of respondents.	No associations between any personality traits and preferences was found within the diagnostic DCE. This set of questions was removed to reduce the survey length and burden for respondents.
Rationality checks		
	Transitivity and stability checks were removed	Failures within the diagnostic DCE (and broader DCE literature) had a limited impact of model estimates. Inclusion of these checks involved the addition of 3 choice tasks, increasing the survey length and complexity. Given the risk-heavy attributes checks were removed to manage complexity. A choice task

		with a dominant alternative was included to check for monotonicity failures.
Qualifying/debriefing questions		
Five assess respondents numerical/risk literacy. Questions were adapted from published instruments (Schapira *et al.*, 2012; Schwartz *et al.*, 2005)		The importance of understanding the numerical literacy of the target population given the risk-heavy attributes (Trevena *et al.*, 2021). Responses were used during analysis to assess any differences in responses based on numeracy ability but were not used to exclude respondents. There was no penalty for completing these questions incorrectly (this was made clear to respondents) and a "I am not sure" option was provided.
	Reasons for selection of indifference alternative	An indifference alternative was not provided in this survey
Attention checks		
No changes. Two attention check questions were included to verify the quality of respondents to accommodate the online format of the questionnaire.		

9.3.5 Stage 5: Piloting

Prior to final data collection, a two-stage pilot study consisting of think-aloud interviews and quantitative survey testing was conducted. The purpose of the pilot study was to determine the:

i. Acceptability and understanding of task instructions
ii. Understanding of attributes, levels and descriptions
iii. Decision-making processes of respondents when completing choice tasks and identify any patterns in non-trading behaviour
iv. Optimal number of choice tasks per person
v. Minimum sample size for final data collection
vi. Acceptability of the online format and wider survey instrument

Results of the pilot study were used to refine the DCE choice tasks and accompanying survey instrument using an iterative process.

9.3.5.1 Qualitative piloting: Think-aloud interviews

Five interviews with members of the public formed the first stage of piloting. Following the diagnostic testing DCE, interviews used the think-aloud technique where respondents verbalised their thoughts whilst completing the choice tasks.

The sample was limited to women over the age of 40. Respondents were recruited via social media (e.g. Facebook) and personal connections. All respondents received an information sheet about the survey and completed an electronic consent form at least 48 hours prior to the scheduled interview. Interviews were conducted online via Zoom. Interviews were held in December 2021.

Subjects completed all 12 choice tasks to allow the appropriate number of choice tasks to be assessed. Following completion of the DCE tasks, several follow up questions were asked to further investigate the decision-making processes of participants and identify any difficulties or areas requiring further refinement. A full interview schedule in provided in Appendix 9.4.

Interviews were digitally recorded and transcribed verbatim. Interview findings, including participant responses and notes on observed behaviour were analysed thematically.

The five interview participants all identified as women, were aged 43-65 years old and lived across England, primarily the North-East region (n=3). Interviews lasted between 35 and 52 minutes.

Summary of changes following qualitative piloting

Analysis of pilot interviews is shown in Appendix 9.5. Several changes were made prior to quantitative piloting in response to findings from the think-aloud interviews.

All changes related to the introductory information which contextualised the choice task and provided instructions for survey participants. Firstly, additional information was added to provide respondents with further information on the prevalence of ovarian cancer. Respondents were told that without screening 65 people per 10,000 screened were diagnosed with ovarian cancer. This figure was determined based on data from the UKCTOCS trial (Menon *et al.*, 2021).

Next, an additional introductory page was added describing potential risk and protective factors of ovarian cancer. The purpose of this information was to allow respondents to form a perception of their personal risk of ovarian cancer in order to further contextualise screening decisions and willingness to endure risks of testing.

To avoid any anxiety arising from the hypothetical choice tasks, both introductory and debriefing information pages were updated to emphasise screening related to individuals without symptoms and reiterate the availability of testing for people experiencing symptoms.

When competing choice tasks, all respondents converted levels from frequencies to percentages. The addition of percentage information alongside numerical levels was considered. However, inconsistencies in how figures were converted and compared

between and within participants meant this was not feasible. This challenge was further compounded by the magnitude of any potential percentages when attempting to express percentages consistently in the context of the 10,000 people screened. Risk communication literature suggests small percentages are difficult to understand (Trevena *et al.*, 2021). Encouragingly, interview findings suggest that for the most part, participants were able to comfortably convert numerical figures into their preferred format successfully in the absence of percentage information.

Finally, respondents appeared to easily complete all twelve choice tasks without signs of fatigue or disengagement therefore blocking did not appear necessary although this would be confirmed based on the completion time and feedback during quantitative piloting.

9.3.5.2 Quantitative piloting

Quantitative piloting aimed to test the acceptability of the online survey instrument, check trading behaviour, refine the experimental design and determine the sample size requirements for the final study. A sample of 40 respondents recruited via Prolific to complete the pilot study. The sample reflected the target population for the final study. Each respondent completed the full survey including thirteen choice tasks (12 plus an additional dominance check question).

Results were analysed using a multinomial model. A number of additional analyses were carried out to assess the suitability of the survey instrument and identify any potential problems prior to final data collection. Additional summary analyses included:

1. Self-reported task difficulty
2. Length of time taken tom complete the survey
3. Frequency of failures of the dominance question

Results of quantitative piloting

A summary of responses to sociodemographic questions from the pilot study is shown in Appendix 9.6.

Completion times ranged from 7 minutes 56 seconds to 51 minutes 3 seconds with a median time of 18 minutes 22 seconds.

Self-reported difficulty responses for the pilot study are shown in Table 9.4. Responses varied as expected 40% (16/40) reported finding the DCE tasks difficult to complete. Encouragingly only three respondents (8%) selected “very difficult”.

One respondent failed the rationality check by choosing the alternative that was objectively worse.

Estimates from the model using continuous coding for all attributes are shown in Table 9.5. All coefficients followed the expected direction of preferences, however, significance of estimates varied presumably due to the sample size. An additional model using categorical dummy-coding for all attributes was also estimated, however, almost all coefficients were insignificant meaning a full assessment of model specification could not be carried out at this stage.

Table 9.4: Multinomial logit results from pilot study

Attribute	**Coefficient (95% CI)**
Ovarian cancer deaths	-0.03 (-0.10 – 0.04)
False negative results	-0.12** (-0.23 – [-0.01])
False positive results	-0.00*** (-0.00 – [-0.00])
Overdiagnosed cancers	-0.06 (-0.13 – 0.01)
No screening	-3.76 (-8.52 – 1.01)
Model fit statistics	
LL	-402.83
Observations	1,440
N	40

9.3.6 Stage 6: Refinement

Quantitative piloting did not highlight any major problems within the survey therefore attributes and levels did not undergo any further changes (Table 9.2).

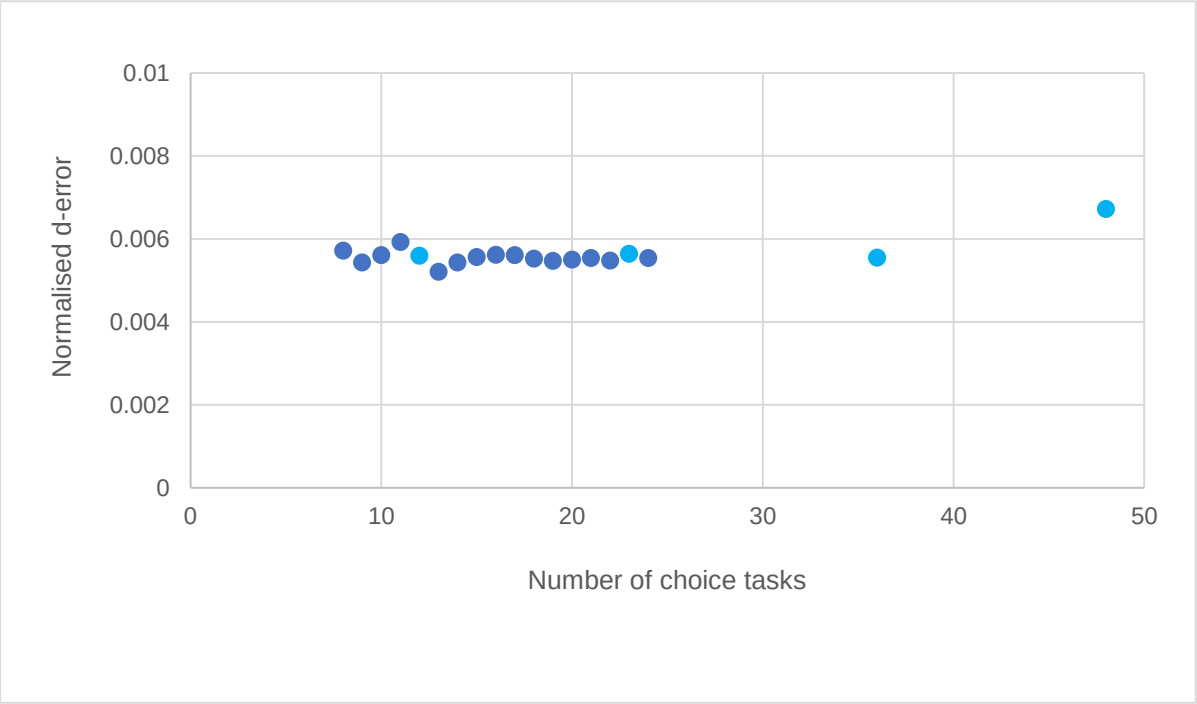
All changes related to the debriefing questions immediately following the DCE tasks. Firstly, an additional question was added asking those who selected “no screening” for all choice tasks why they did so. The purpose of this question was to verify whether this behaviour was a genuine preference or a simplifying heuristic.

A supplementary question explicitly asking if respondents ignored any attributes was added to complement the existing attribute non-attendance” question (“which characteristics did you base your choices on?”). The conditional follow-up question relating to reasons for non-attendance was also reworded (“Why did you ignore certain characteristics?”). The purpose of these changes was to improve the clarity in order to address the discordance between self-reported attendance questions observed during quantitative piloting (see Chapter 11 for further explanation).

Experimental design

The experimental design was updated using a Bayesian Efficient approach (main effects only). Results of the continuous model were used as priors, however, investigation of the appropriate data specification were carried out during the final data analysis. Examination of the d-error when varying the number of choice tasks confirmed 12 choice tasks was appropriate (after this point increasing the number of choice tasks provided no further information per choice task) (Figure 9.11). Both stages of pilot study confirmed 12 choice tasks was acceptable to respondents and blocking was not necessary. The Ngene syntax for the final design is shown in Appendix 9.7.

Figure 9.11: Normalised d-error for experimental designs with increasing numbers of choice tasks



A copy of the final survey instrument is provided in Appendix 9.8.

9.4 Chapter summary

This chapter described the process of developing a DCE to investigate women's preferences towards a potential screening programme for ovarian cancer. DCE development aimed to be rigorous and transparent and refined the evidence-based development framework outlined in previous chapters.

This chapter utilised a simplified attribute selection process in comparison to the earlier DCE on diagnostic testing by excluding the online qualitative workshop phase. This decision was made following the results of the best-worst scaling study. Given the dominance and clear separation of test-performance characteristics to the comparatively large sample in the BWS study (compared to potential sample of focus groups/workshops). It was determined that any attempt to make participants trade-off between attributes falling in to the two separate clusters would resulting in non-attendance and non-trading behaviour, a large issue in the diagnosis DCE, likely to be partially caused by the merging of findings from the quantitative and qualitative

findings from target population engagement. Whilst this approach avoided the challenge of reconciling multiple streams of evidence the absence of qualitative methods means the depth of exploration during attribute selection was more limited.

The initial research questions of the DCE were broad in nature. The aim was to understand the key drivers of demand for screening and estimate how uptake may vary as barriers and facilitators were varied. Ultimately the selected attributes meant the research aims of the DCE were refined to a specific aspect of acceptability. Namely, what level of test performance is acceptable to women? What is the acceptable balance of benefits and harms of testing? This means results from the DCE address two questions that are fundamental for the assessment of any future screening programme by the UK National Screening Committee (UK NSC);

- The acceptability of a screening programme to the public
- The point at which the benefit of testing outweighs the harms

Furthermore, DCE results may also be useful in evaluating the feasibility of screening programmes beyond patient acceptability since measures such as efficacy or cost-effectiveness are dependent on achieving a threshold level of uptake.

The design of the DCE improves on the shortcomings of the diagnostic DCE in several aspects. For instance, efforts were made to remove any ambiguity regarding the meaning of the opt-out alternative and the use of constant levels aims to clarify the consequences of opting for no screening. Secondly, additional debriefing questions were added to allow further investigate aspects of choice behaviour. For instance, questions relating to serial non-testers and additional questions relating to stated ANA.

Given the prevalence of risk-based attributes, risk communication was a primary focus for this chapter was risk communication. This chapter builds on the lessons from the diagnostic DCE by explicitly incorporating disease prevalence into risky attributes. However, the low occurrence of ovarian cancer presented an additional challenge. By incorporating public consultation during the development process it is hoped risk communication is clear and understandable to the maximum number of respondents.

However, studies have demonstrated the most preferred risk presentation format is not always the most effective. Comparisons of DCE responses using different risk formats appears to be an important area for future research.

Finally, this chapter aimed to ensure understanding of the DCE and survey instrument by undergoing a two-stage iterative pilot study prior to final data collection. Results suggest that despite the complexity of information included within the DCE, respondents appear to be engaged in the choice tasks and able to make trade-offs between attributes.

10 Women's preferences towards a potential ovarian cancer screening programme: Results from a DCE

10.1 Introduction

This chapter describes the data collection process, analysis and results of a discrete choice experiment designed to elicit public preferences towards a potential ovarian cancer screening programme. The chapter utilises the survey instrument developed and piloted in the previous chapter. DCE results were used to calculate several welfare measures to aid the interpretation of results such as scenario analysis to predict screening uptake. The results from this chapter provide guidance on the acceptability of hypothetical tests used to screen for ovarian cancer. In particular, the results suggest a threshold of test performance candidate screening tests must achieve to be acceptable to women.

10.2 Aims and objectives

The primary aim of the survey was to examine women's preferences towards ovarian cancer screening. Specifically focusing on test performance characteristics and the trade-offs between the benefits (i.e. reduced mortality) and harms of testing

The primary objectives of this chapter were:

(i) Measure the relative importance of test performance attributes relating to ovarian cancer
(ii) To examine the willingness to trade between attributes based on marginal rates of substitution
(iii) To estimate the uptake of screening for tests with different balances of risks and benefits using scenario analysis
(iv) To explore sociodemographic characteristics which may influence stated preferences for screening
(v) To investigate factors that may influence non-participation in screening

10.3 Methods

10.3.1 Study population

The target population for the DCE was women with at least one ovary over the age of 40 years old. Women recruited from the general public was selected since this would be the target population eligible for screening if/when a universal screening becomes available. To overcome a limitation of the earlier diagnostic DCE, an ethnicity quota was introduced. Responses from white respondents were limited to 85% of the total sample. This limit was determined based on English and Welsh population statics from the 2011 census (ONS, 2021). No other limits to participation were applied.

10.3.2 Sample size

A minimum sample size of 45 participants was estimated based on the s-estimate approach described in Chapter 7 (Rose & Bliemer, 2013). A final sample size of 250 to account for uncertainty in pilot results and allow investigation sociodemographic drivers of preference heterogeneity during analysis.

10.3.3 Recruitment

Following the success of previous studies, Prolific (Prolific.co) was used to recruit participants for the experiment. Participants were presented with an electronic information sheet and consent form prior to the beginning of the survey. Participants were paid a completion fee of £2 directly into their prolific account.

10.3.4 Analysis plan

Anonymised data were downloaded and cleaned in SPSS v27 and data analysis was performed using Stata 17.

Respondent characteristics were summarised using descriptive statistics to provide an understanding of the sample.

10.3.4.1 Ranking results

Respondents were asked to rank the attributes within the survey in terms of importance from 1(most important) to 4 (least important). Results were analysed to understand the frequency of ranking position for each attribute and an average ranking score between 1 and 4, where a lower score indicates a higher level of importance.

10.3.4.2 Analysis of stated choice data

To begin, a main-effects multinomial logit (MNL) model including dummy-coded attribute levels was estimated. Parameter coefficients were assessed to determine the functional form of each attribute.

Following confirmation of the correct functional form, a mixed logit (ML) model assuming a continuous linear specification was estimated based on the following utility function:

$$V = \alpha_B + \alpha_{Opt-out} + \beta_1 \text{mortality} + \beta_2 \text{falsenegative} + \beta_3 \text{falsepositive} + \beta_4 \text{overdiagnosis}$$

Beta coefficients, $\beta_1 - \beta_4$ represent the relative utility weights of the four test performance attributes. The opt-out alternative was incorporated in the model using an ASC, $\alpha_{Opt-out}$. An additional ASC, α_B was included in account for any left-right bias in respondent choices.

Interpretation of findings

Results from the ML model were used to calculate three measures to aid interpretation and application to policy questions:

i. Relative importance of attributes

Relative importance scores for each attribute were calculated to assess how much each attribute contributes to overall utility when considering the level range of each attribute. Scores range between 0 and 1 with higher scores indicating greater importance.

ii. Marginal rate of substitution (MRS)

MRS calculations were used to explore the willingness to trade between the potential benefits and harms of screening. Specifically, trade-offs between attributes were calculated as the level of additional risk (i.e. overdiagnosed cancers, false positives or false negatives) that would be accepted to avoid one extra death.

For example, eq. 1 provides an example willingness to accept (WTA) calculation and is interpreted as the additional number of false negative results per 10,000 people screened that would be accepted in exchange for one ovarian cancer death avoided over a 10-year period.

$$\frac{\beta_1 \text{mortality}}{\beta_2 \text{falsenegative}} \qquad \text{eq.1}$$

iii. Predicted uptake

Finally, expected uptake for different screening tests with varying levels of benefits and harms were estimated using the methods described in Chapter 7. Since no feasible candidate screening tests have emerged following clinical trials to date, scenarios were hypothetical and chosen to demonstrate potential changes in uptake as each attribute varies.

Heterogeneity of preferences

A latent class logit model was used to further explore preference heterogeneity. The goodness-of-fit of models including between 2-7 classes were compared. The appropriate model was selected based on a balance between interpretability and goodness-of-fit statistics (AIC, BIC, CIAC and log-likelihood). Sociodemographic characteristics described in Table 10.1 were included within the model to test the influence of individual characteristics on the probability of class membership. Sociodemographic characteristics of interest were selected based on the systematic review presented in Chapter 4.

Table 10.1: Sociodemographic characteristics included in subgroup analysis and latent class logistic regression models

Category	Subgroups
Age	• Continuously coded
Ethnicity	• White • Non-white
Education	• Attended university • Did not attend university
Employment	• Currently working • Not currently working
Children	• Continuously coded
Self-reported health	• Good health • Average/below good health
Ovarian cancer testing experience	• Previously tested • Never tested
Perceived risk of ovarian cancer	• High risk • Average risk • Low risk
Worry about ovarian cancer	• Low • Moderate • High
Know someone with ovarian cancer	• Yes • No
Cervical screening attendance	• Always attends • Sometimes attends • Never attends
Risk attitude	• Risk averse (score of 1-4 on willingness to take risks) • Risk neutral (score of 5-6 on willingness to take risks) • Risk seeking (score of 7-10 on willingness to take risks)
Worried about ovarian cancer	• Yes • No
Symptom awareness	• Continuously coded between 0-12 representing the number of symptoms recognised
Low confidence in ability to recognise symptoms of ovarian cancer	• Yes • No
Task difficulty	• Very easy/easy • Neither difficult or easy • Very difficult/difficult
Numerical ability	• Continuous variable 0-5 representing the number of correct responses

Sensitivity analysis

To check the validity of responses, several sensitivity checks were performed. Firstly, individuals who failed the dominance rationality check were excluded to test the influence on parameter estimates. Next, subgroup analysis was performed to explore how self-reported task difficulty and numerical ability influenced model estimates.

Finally, a heteroskedastic logit model including failures of the numeracy question and/or dominance task and self-reported task difficulty as scale factors was estimated to explore whether any of these factors influenced the error-variance (i.e. randomness) of responses.

Analysis of opt-out behaviour

Selection of the "no screening" alternative was examined to identify any common characteristics associated with increased non-screening behaviour. Reasons for serial non-participation across all choice tasks were summarised narratively. Finally, a logistic regression model with opt-out choices as the dependent variable and sociodemographic characteristics as the independent variables was estimated.

10.4 Results

10.4.1 Sociodemographic characteristics

In total, 258 individuals began the survey. Four people dropped out part way through and an additional 4 people were excluded after failing the attention check questions, leaving a final sample size of 250 respondents. Key respondent characteristics are summarised in Table 10.2.

Health history, behaviours and attitudes

Responses to key health-related questions are provided in Appendix 10.1. Ovarian cancer worry was generally low across the population, with 74% of respondents stating little-to-no worry (184/250). Over three-quarters of respondents (192/250; 77%) indicated that they were not confident in their ability of recognise symptoms of ovarian

cancer. Rates of symptom recognition ranged from 22% (55/250) to 68% (169/250), and 12% (29%) did not recognise any of the key ovarian cancer symptoms. Current screening behaviour was varied, however 63% (157/250) of respondents reported undergoing cervical screening every time they received an invitation.

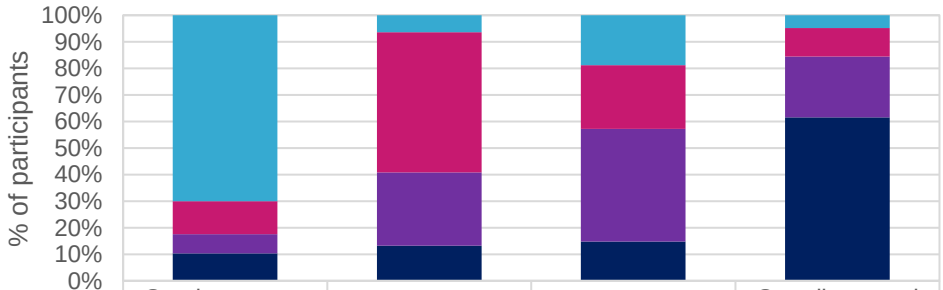
Table 10.2: Sociodemographic characteristics of respondents completing the DCE survey

Characteristic	
Age	
Mean (SD)	52.9 (8.7)
Range	40-80
Ethnicity, n (%)	
White	198 (79%)
Mixed-white and black Caribbean	6 (2%)
Mixed- white and Asian	5 (2%)
Asian- Indian	5 (2%)
Asian- Chinese	8 (3%)
Black- African	9 (4%)
Black- Caribbean	10 (4%)
Other	6 (2%)
Prefer not to say	3 (1%)
Children	
Mean (SD)	1.52 (1.2)
Range	0-5
Relationship status, n (%)	
Single	36 (14%)
In a relationship	46 (18%)
Married/civil partnership	125 (50%)
Separated/divorce	34 (14%)
Widowed	9 (4%)
Education, n (%)	
No qualifications	2 (1%)
GCSE	53 (21%)
A-Level/ College	51 (20%)
Undergraduate	86 (34%)
Post-graduate/ professional quals	52 (21%)
Other	4 (2%)
Prefer not to say	1 (0.4%)
Employment, n (%)	
Employed, full-time	84 (34%)
Employed, part-time	58 (23%)
Self-employed	35 (14%)
Not employed	9 (4%)
Retired	31 (12%)
Other	31 (12%)
Prefer not to say	2 (1%)
Household income, n (%)	
£0-9,999	12 (5%)
£10,000-19,999	36 (14%)
£20,000-29,999	41 (16%)
£30,000-39,999	43 (17%)
£40,000- 49,999	38 (15%)
£50,000- 59,999	16 (6%)
£60,000-69,999	17 (7%)
£70,000+	25 (10%)
Prefer not to say	22 (9%)
Willingness to take risks (1 not at all – 10 completely willing)	
Mean (SD)	4.3 (2.2)
Task difficulty, n (%)	
Very easy	11 (4%)
Easy	59 (24%)
Neither easy or difficult	63 (25%)
Difficult	104 (42%)
Very difficult	13 (5%)

10.4.2 Ranking

Figure 10.1 shows the ranking frequency for each of the attributes. Ovarian cancer deaths was ranked first most frequently and was also the most important attribute overall with an average importance score of 1.58 out of 4. False positives and false negatives differed in the distribution of ranking position but appeared similar once ranking scores were averaged (false negatives=2.47 and false positives= 2.53). Overdiagnosed cancers was ranked least important overall with an average score of 3.41.

Figure 10.1: Frequency of ranking positions of each attribute in the stand-alone ranking question



	Ovarian cancer deaths	False negatives	False positives	Overdiagnosed cancers
1st	175	16	47	12
2nd	31	132	60	27
3rd	18	69	106	57
4th	26	33	37	154

10.4.3 DCE results: multinomial and mixed logit results

A multinomial logit model using dummy-coded levels was initially estimated to check the functional form of all attributes (Appendix 10.2). Coefficient plots for each attribute were examined and continuous linear coding appeared to be acceptable based on visual inspection (Appendix 10.3).

Multinomial logit and mixed logit results using a continuous linear specification are shown in Table 10.3. The likelihood-ratio test demonstrated the mixed logit model, which accounts for preference heterogeneity resulted in a significant improvement in model specification, therefore interpretations focus on results from this model.

All attributes were significant and followed the expected direction. Since all attributes were negatively framed, the negative coefficients indicate an increase in incidence, for example, additional people dying from ovarian cancer, leads to a reduction in utility. The ASC associated with the “no screening” alternative was negative and large in magnitude, demonstrating an overall preference to be screened. However, the large standard deviation (4.74) indicates high levels of heterogeneity across respondents with almost 40% of respondents showing a preference towards no screening (based on a z-score of 2.29/4.74=0.48). Heterogeneity in preferences across the remaining parameters was also observed but at much lower levels as indicated by the smaller standard deviations.

10.4.3.1 Relative importance scores

Relative importance scores for each attribute are shown in Table 10.3. Ovarian cancer deaths (0.42) was most important overall. Interestingly, the order of importance differed from the ranking exercise with false negatives being least prioritised within the choice experiment.

10.4.3.2 Marginal rates of substitution

Table 10.4 shows the willingness to accept extra harms of testing to avoid one additional ovarian cancer death. Results relate to the number of additional harms per 10,000 people screened over a 10-year period. For example, respondents were willing to accept an additional 205 false positive results over 10 years in exchange for one life saved.

Column 2 converts the WTA estimates into percentages based on an incidence rate of 65 clinically meaningful cancers per 10,000 people over a 10-year period (Menon *et al.*, 2021). Estimates demonstrate the acceptable percentage increase in harms exchanged for a 1% reduction in mortality. For example, participants were willing to accept a 1.59% increase in false negative results for a 1% reduction in mortality over a 10-year period.

Table 10.3: Multinomial logit and mixed logit model results

	Multinomial logit		Mixed logit		
	Coefficient (95% CI)	**Relative importance**	**Mean (95% CI)**	**SD**	**Relative importance**
Ovarian cancer deaths	**-0.08***** (-0.09 – [-0.07])	**0.47** (0.44 – 0.51)	**-0.14***** (-0.16 – [-0.12])	**0.10***** (0.08 – 0.12)	**0.42** (0.40 – 0.44)
False negative results	**-0.02***** (-0.03 – [-0.01])	**0.11** (0.08 – 0.13)	**-0.05***** (-0.07 – [-0.04])	**0.05***** (0.03 – 0.07)	**0.14** (0.12 – 0.15)
False positive results	**$-3.11x10^{-4}$***** ($-3.73x10^{-4}$ – [$-2.47x10^{-4}$])	**0.27** (0.27 – 0.27)	**$-6.81x10^{-4}$***** ($-8.10x10^{-4}$ – [$-5.52x10^{-4}$])	**$6.53x10^{-4}$***** ($4.91x10^{-4}$ – $8.16x10^{-4}$)	**0.30** (0.30 – 0.30)
Overdiagnosed cancers	**-0.03***** (-0.04 – [-0.02])	**0.15** (0.14 – 0.16)	**-0.06***** (-0.07 – [-0.05])	**0.04***** (-0.18 – 0.35)	**0.14** (0.13 – 0.15)
No screening	**-0.31***** (-0.17 – [-0.02])	-	**-2.29***** (-3.01 – [-1.57])	**4.74***** (3.75 – 5.72)	-
Model fit statistics					
LL	-2881.70		-1913.08		
LR test (ML vs MNL)	-		1937.2***		
Observations	9,000		9,000		
N	250		250		
Key: ***significant at 99% confidence level; **significant at 95% confidence level; *significant at 90% confidence level					

Table 10.4: Willingness to accept additional harms (per 10,000 people screened) to avoid an additional ovarian cancer death over a 10-year period.

	WTA per 1 additional death avoided (95% CI)	WTA per 1% reduction in mortality (95% CI)
Number of additional false negative results	**2.59** (1.82 – 3.36)	**1.59%** (1.13-2.07%)
Number of additional false positive results	**205.20** (161.89 – 248.51)	**0.83%** (0.65-1.00%)
Number of overdiagnosed cancers	**2.35** (1.76 –2.94)	**1.40%** (1.05-1.73%)

10.4.3.3 Predicted uptake

Table 10.5 shows the predicted uptake for different hypothetical screening tests described in terms of the four DCE attributes. Predicted participation levels were generally high. Scenario 1 represents the current performance of ovarian cancer screening based on outcomes of the UKCTOCS trial (Henderson *et al.*, 2018; Menon *et al.*, 2021). Despite the significant risk of harms and no impact on mortality, results suggest 78% (95% CI: 66-89%) of respondents would still choose to be tested, demonstrating a strong preference to undergo screening.

10.4.3.4 Sensitivity checks

In total, twelve respondents chose the inferior alternative in the dominance rationality check choice task. Sensitivity analysis revealed no significant changes in any model parameters when failing respondents were excluded from the analysis (Appendix 10.4). On this basis, all responses were maintained.

Subgroup analysis revealed differences in parameter estimates based on both self-reported task difficulty and numerical ability. For example, those who found the DCE tasks easy and/or displayed high numeric ability showed a preference for no screening and placed higher importance on ovarian cancer deaths. A heteroskedastic logit model was used to investigate the effects of both factors on scale heterogeneity (i.e. error variance) (Table 10.6). Scale terms associated with both task difficulty and numerical ability were insignificant suggesting error variance did not differ across individuals

based on these factors. Interaction models revealed that differences in model estimates between the subgroups remained even after controlling for scale heterogeneity, suggesting differences were due to genuine preference heterogeneity rather than response variation based on task understanding or burden.

Table 10.5: Scenario analysis estimating uptake of ovarian cancer screening tests with differing levels of benefits and harms

	Attributes				**Predicted uptake**
	Ovarian cancer deaths	False negatives	False positives	Overdiagnosis	% Participation (95% CI)
1*	40 in 10,000 (0% reduction)	7 in 10,000 (10%)	994 in 10,000 (1%)	0 in 10,000 (0%)	77.5% (65.8-89.1%)
2	30 in 10,000 (25% reduction)	23 in 10,000 (35%)	994 in 10,000 (1%)	28 in 10,000 (30%)	52.6% (34.5-70.7%)
3	36 in 10,000 (10% reduction)	7 in 10,000 (10%)	1987 in 10,000 (2%)	0 in 10,000 (0%)	75.3% (63.1-87.5%)
4	10 in 10,000 (75% reduction)	13 in 10,000 (10%)	3974 in 10,000 (4%)	16 in 10,000 (20%)	89.3% (81.5-97.1%)
5	20 in 10,000 (50% reduction)	7 in 10,000 (10%)	1987 in 10,000 (2%)	0 in 10,000 (0%)	96.6% (94.3-99.0%)
6	10 in 10,000 (75% reduction)	10 in 10,000 (15%)	2981 in 10,000 (3%)	3 in 10,000 (5%)	97.7% (95.8 – 99.5%)
7	20 in 10,000 (50% reduction)	7 in 10,000 (10%)	497 in 10,000 (5%)	0 in 10,000 (0%)	98.7% (97.8-99.7%)
8	8 in 10,000 (80% reduction)	3 in 10,000 (5%)	994 in 10,000 (1%)	0 in 10,000 (0%)	99.7% (99.4-99.9%)
9	0 in 10,000 (100% reduction)	0 in 10,000 (0%)	0 in 10,000 (0%)	0 in 10,000 (0%)	99.9% (99.9 - 100%)
*Scenario reflects current screening performance based on UKCTOCS trial					

.

Table 10.6: Heteroscedastic logit model used to investigate scale heterogeneity between respondents based on numerical ability and self-reported task difficulty

	Scale terms only	Numeracy interaction model	Task difficulty interaction model
	Coefficient (95% CI)	Coefficient (95% CI)	Coefficient (95% CI)
Attributes			
Ovarian cancer deaths	**-0.07*** ** (-0.09 – [-0.06])	**-0.07*** ** (-0.08– [-0.06])	**-0.07*** ** (-0.08– [-0.06])
False negative results	**-0.02*** ** (-0.03 – [-0.01])	**-0.02*** ** (-0.04 – [-0.01])	**-0.03*** ** (-0.05 – [-0.02])
False positive results	**-2.71x10^{-4}*** ** (-3.56x10^{-4} – [-1.85x10^{-4}])	**-3.12x10^{-4}*** ** (-3.91x10^{-4} – [-2.45x10^{-4}])	**-4.10x10^{-4}*** ** (-5.04x10^{-4}– [-3.17x10^{-4}])
Overdiagnosed cancers	**-0.03*** ** (-0.04 – [-0.02])	**-0.04*** ** (-0.05 – [-0.03])	**-0.04*** ** (-0.05– [-0.02])
No screening	**0.35*** (-0.02 – 0.74)	**4.05x10^{-4}** (-0.48 – 0.48)	**-0.56**** (-1.08– [-0.03])
Interaction terms			
Ovarian cancer deaths × High numeracy		**0.05*** (-0.00 – [-0.09])	
False negative results × High numeracy		**0.02*** ** (0.01 – 0.04)	
False positive results × High numeracy		**2.41x10^{-4}*** ** (8.24x10^{-5} – 4.00x10^{-4})	
Overdiagnosed cancers × High numeracy		**0.03*** ** (0.02 – 0.05)	
No screening × High numeracy		**0.31** (-0.46 – 1.08)	
Ovarian cancer deaths × Task easy			**0.01** (-0.11 – 0.13)
False negative results × Task easy			**0.03*** ** (0.01 – [-0.05])
False positive results× Task easy			**3.91x10^{-4}*** ** (1.05x10^{-4} – 5.33x10^{-4})
Overdiagnosed cancers × Task easy			**0.02** (-0.02 – 0.05)
No screening × Task easy			**1.46** (-0.08 – [-0.06])
Ovarian cancer deaths × Task neutral			**0.02** (-0.06 – 0.11)
False negative results × Task neutral			**0.03**** (0.01 – 0.05)
False positive results × Task neutral			**2.74x10^{-4}*** (-9.38x10^{-6} – 5.58x10^{-4})
Overdiagnosed cancers × Task neutral			**0.02** (-0.01 – 0.05)
No screening × Task neutral			**0.93*** (-0.47 – 3.07)
Scale terms			
High numeracy	**0.17** (-0.06 – 0.39)	**1.32** (-0.43 – 3.07)	
Self-reported difficulty: very easy/easy	**0.11** (-0.16 – 0.38)		**0.74** (-1,26– 2.75)
Self-reported difficulty: neither easy or difficult	**0.09** (-0.18 – 0.36)		**0.83** (-1.35– 3.01)
Model fit statistics			
LL	2878.49	-2867.38	-2854.43
Observations	9,000	9,000	9,000
N	250	250	250
Key: ***significant at 99% confidence level; **significant at 95% confidence level; *significant at 90% confidence level			

10.4.3.5 Preference heterogeneity - Latent class model

Examination of model statistics indicated that a five-class model provided a good balance between model-fit and interpretability given the sample size (Appendix 10.5). Latent class logit results are shown in Table 10.7. The model included several sociodemographic characteristics which are interpreted relative to the fifth class. Mean attribute importance scores (Figure 10.2) and WTA estimates (Table 10.8) for each preference class were calculated based on model coefficients.

Class 1: Conscientious testers 15.6%

This group showed a significant preference for screening versus no screening and appeared willing to trade between all attributes as indicated by the significance of all model parameters. Decisions for this group appear to be most focused on the balance between the two most important attributes: false positive results (importance score= 0.51) and mortality (importance score=0.22). Despite being significant, false negatives (0.13) and overdiagnosed cancers (0.14) appeared less important to this class. Responses in this preference class were also less likely to be from individuals who know someone diagnosed with ovarian cancer and less likely to be university-educated.

Class 2: Risk-conscious traders (23.1%)

Class 2 was the second biggest preference group identified by the model. Similar to Class 1, risk of false positive results was the most important factor for this group (importance score=0.40). Overall, this class placed the lowest relative importance on ovarian cancer deaths (0.16) whilst also expressing a very strong preference for screening, as indicated by the large negative “no screening” ASC. This finding suggests undergoing screening may offer an additional intrinsic benefit beyond outcomes for this group, such as a psychological benefit or “peace of mind”. Responses in this group were significantly more likely to belong to individuals exhibiting lower numerical ability and have found the DCE tasks difficult—both of which may influence the ability to engage with choice tasks although sensitivity analysis suggested this was not a concern within this study.

The lower importance placed on mortality leads to surprising WTA results; respondents were willing to accept just 5 false negatives per 10,000 people screened and 7 overdiagnosed cancers per 10,000 people screened to avoid 10 ovarian cancer deaths over a period of 10 years.

Class 3: Mortality-focused non-testers (10.8%)

Class 3 was the smallest preference class identified within the model. Responses in this class were the most test-averse group. Ovarian cancer deaths was the only significant attribute within this model meaning WTA estimates for this group were also insignificant. There were no clear sociodemographic drivers of class membership although there was evidence on a weak relationship between decreased perceived risk of ovarian cancer and lower worry about ovarian cancer within this group.

Class 4: Screening-averse non-testers (19.6%)

This class displayed a strong preference for no screening. Preferences were primarily driven by the risk of ovarian cancer deaths and overdiagnosed cancers, with false positives and false negatives not significantly influencing utility at a 95% confidence level. The relative importance of overdiagnosed cancers was highest for this class overall (0.28) suggesting this class were more risk conscious than class 3 which was comparable in terms of preference for no screening. This class was significantly more likely to exhibit low worry about ovarian cancer and have low self-perceived risk of cancer. Individuals with responses in this class were also significantly less likely to regularly undergo cervical screening suggesting a general averseness to screening.

Class 5: Mortality minimisers (30.9%)

This was the largest class within the model. This class appeared to be indifferent towards testing as indicated by the non-significant no screening ASC. This group appeared to prioritise the outcomes and risks for individuals with ovarian cancer, placing greater importance on ovarian cancer deaths (0.57) whilst false positive results for those without cancer did not impact utility. Responses in this class exhibited the highest willingness to accept additional false negative results (2.89 per 10,000) and overdiagnosed cancers (2.22 per 10,000) in exchange for an additional live saved.

Table 10.7: Latent class analysis results

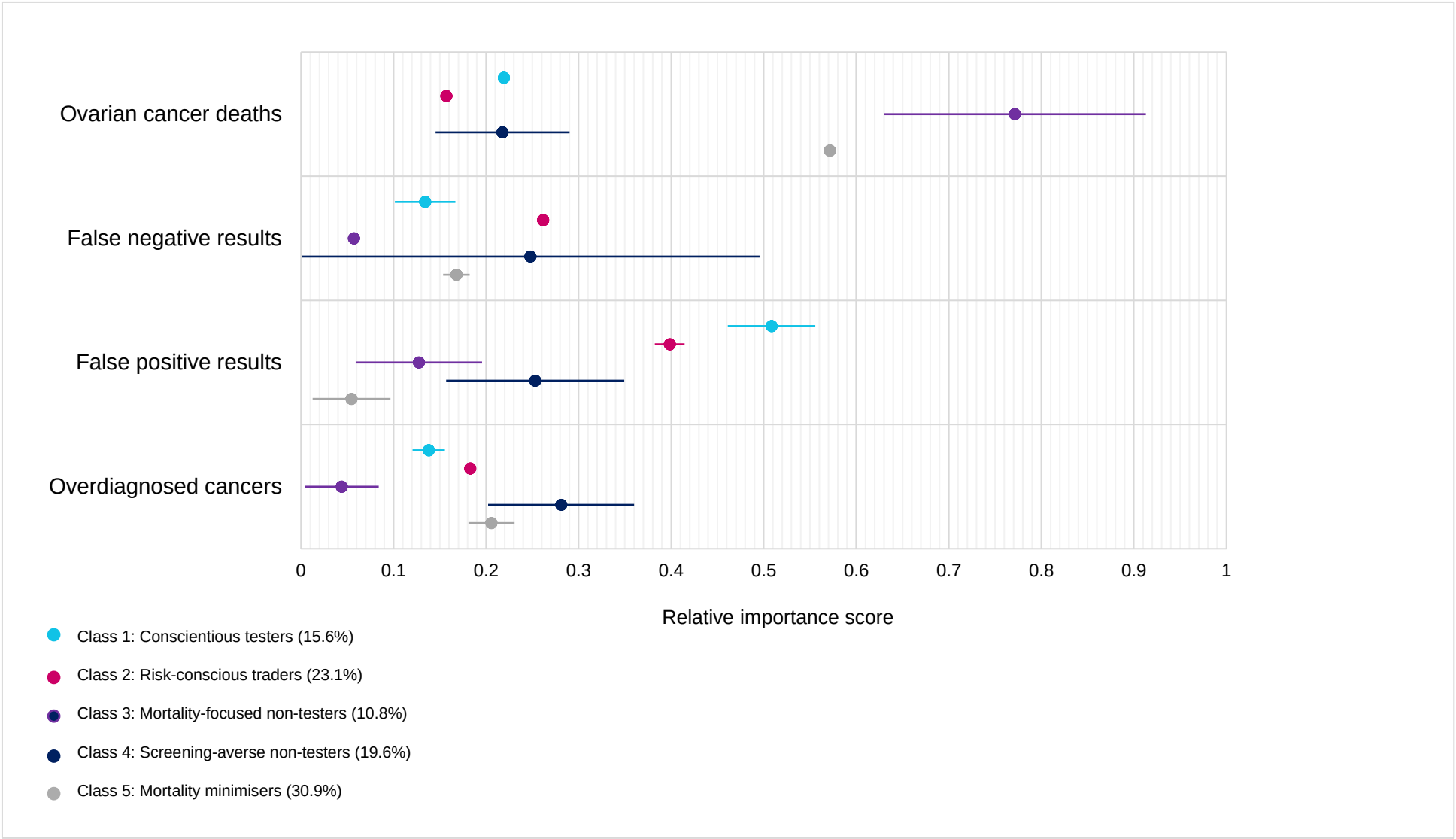
	Class 1	Class 2	Class 3	Class 4	Class 5
	Coefficient (95%CI)	**Coefficient (95%CI)**	**Coefficient (95%CI)**	**Coefficient (95%CI)**	**Coefficient (95%CI)**
Attribute utility weights					
Ovarian cancer deaths	**-0.07*** (-0.09 – [-0.05])	**-0.03*** (-0.05 – [-0.02])	**-0.19*** (-0.24 – [-0.15])	**-0.05**** (-0.09 – [-0.01])	**-0.34*** (-0.41 – [-0.26])
False negative results	**-0.05*** (-0.08 – [-0.02])	**-0.07*** (-0.09 –[-0.05])	**0.02** (-0.03 – 0.06)	**0.07*** (-0.00 – 0.14)	**-0.12*** (-0.15 – [-0.08])
False positive results	**-0.001*** (-0.001 – [-0.001])	**-5.80×10^{-4}*** (-7.21×10^{-4} – [-4.37×10^{-4}])	**-2.07×10^{-4}** (-4.83×10^{-4} – 7.01×10^{-5})	**-4.00×10^{-4}*** (-7.48×10^{-4} – [-5.23×10^{-5}])	**2.15×10^{4}** (-5.82×10^{-5}– 4.89×10^{4})
Overdiagnosed cancers	**-0.05*** (-0.08 – [-0.03])	**-0.05*** (-0.07 – [-0.03])	**-0.01** (-0.06 – 0.03)	**-0.08**** (-0.14 – [-0.02])	**-0.15*** (-0.21 – [-0.10])
No screening	**-1.35*** -2.23 – [-.0.47])	**-5.19*** (-6.09 – [-4.30])	**4.32*** (2.36– 6.28)	**3.53*** (1.88 – 5.18)	**1.03** (-0.27 – 2.31)
Class probability model					
Know someone diagnosed with OC	**-1.89**** (-3.57 – [-0.22])	**-1.00*** (-2.14– 0.15)	**0.60** (-0.50 – 1.70)	**-0.50** (-1.62 – 0.62)	-
Attended university	**-1.37**** (-2.41– [-0.33])	**-1.11*** (-1.95 – [-0.27])	**0.47** (-0.73 – 1.66)	**-0.65*** (-1.50 – 0.19)	-
Always attends cervical screening	**-0.95*** (-1.96 – [-0.07])	**0.10** (-0.80 – 0.99)	**0.19** (-1.00 – 1.38)	**-0.78**** (-1.61 – [-0.02])	-
Found DCE difficult/very difficult	**2.67*** (1.41 – 3.92)	**1.01**** (0.18 – 1.84)	**0.55** (-0.48 – 1.57)	**0.12** (-0.75 – 0.99)	-
Low OC worry	**-0.55** (-1.65 – 0.54)	**0.39** (-0.52 – 1.30)	**1.41*** (-0.07 – 2.90)	**1.47*** (0.36 – 2.58)	-
Low perceived OC risk	**1.89**** (0.50– 3.28)	**1.65*** (0.52 – 2.78)T	**1.21*** (-0.10 – 2.52)	**1.53*** (0.45 – 2.61)	-
Numeracy ability	**-0.08** (-0.54 – 0.37)	**-0.67*** (-1.05 – [-0.29])	**0.03** (-0.48 – 0.54)	**-0.31** (-0.72 – 0.09)	-
Low confidence in identifying OC symptoms	**0.79** (-0.76 – 2.34)	**0.10** (-1.08 – 0.89)	**0.58** (-1.79 – 0.63)	**-0.80**** (-1.71 – 0.11)	
Constant	**-1.07** (-3.67 – 1.52)	**1.91**** (0.05 – 3.77)	**-2.85**** (-5.63 – [-0.06])	**0.79** (-1.22 – 2.80)	-
Class probabilities					
	15.0%	23.2%	10.9%	20.0%	30.9%
Model fit statistics					
Log-likelihood	-1856.69				
AIC	3845.38				
BIC	4077.79				
CAIC	4025.98				

Key: ***significant at 99% confidence level; **significant at 95% confidence level; *significant at 90% confidence level

Table 10.8: Willingness to accept estimates for each preference class. WTA estimates represent additional harms (per 10,000 people screened) to avoid an additional ovarian cancer death over a 10-year period (95% CI).

	Class 1	Class 2	Class 3	Class 4	Class 5
Number of additional false negative results	**1.39*** ** (0.43 – 2.35)	**0.51*** ** (0.28 – 0.73)	**-11.44** (-39.62 – 16.73)	**-0.75** (-1.76 – 0.27)	**2.89*** ** (2.25 – 3.54)
Number of additional false positive results	**64.72*** ** (43.80– 85.63)	**59.12*** ** (36.47– 81.77)	**908.37** (-389.38 – 2209.11)	**129.12** (-40.12 – 298.40)	**-1569.70** (-3403.92 – 264.53)
Number of overdiagnosed cancers	**1.27*** ** (0.57– 1.97)	**0.69*** ** (0.38 – 1.00)	**14.03** (-28.93 – 56.99)	**0.62*** (0.09 – 1.33)	**2.22*** ** (1.66 – 2.79)

Figure 10.2: Relative importance scores for each preference class estimated in the latent class logit model



10.4.3.6 Opt-out behaviour

The no screening alternative was selected in 10% of all choice tasks (929/9000) and 109 (44%) respondents opted not to be tested in at least one instance. Logistic regression found individuals who considered themselves low risk (OR=1.48) or exhibit low levels of worry about ovarian cancer (OR=1.76) were significantly more likely to select the no screening alternative (Table 10.9). Risk-averse individuals were also more likely to opt-out (OR 1.45) Oppositely, individuals who regularly participate in cervical screening were less likely to opt for no screening (OR=0.63).

Serial opt-out selection

Twenty-seven (11%) of respondents opted not to undergo screening in all twelve choice tasks. Common reasons for this decision were limited benefits (i.e. reduction in mortality) from screening (n=9), potential harms were too high (n=9), specifically false-positives (n=7) and low self-perceived risks of ovarian cancer (n=6).

Table 10.9: Logit model results exploring the relationship between sociodemographic characteristics and selection of the "no screening" alternative

	Full model	Reduced model
	Odds ratio (95% CI)	Odds ratio (95% CI)
Age	**1.01** (0.99–1.03)	
Employed	**0.75** (0.53–1.07)	
Ethnicity- white	**0.77** (0.48–1.24)	
Number of children	**0.94** (0.81–1.10)	
Attended university	**1.08** (0.77–1.51)	
Know someone diagnosed with ovarian cancer	**0.77** (0.46–1.29)	
Always attends cervical screening	**0.66***** (0.48–0.92)	**0.65**** (0.47–0.90)
Found DCE difficult/very difficult	**0.90** (0.65–1.23)	
Low ovarian cancer worry	**1.68**** (1.11–2.53)	**1.77**** (1.17–2.69)
Low perceived ovarian cancer risk	**1.33*** (0.92–1.93)	**1.44**** (1.03–2.03)
Numerical ability	**1.00** (0.83–1.19)	
Self-reported health: Very good—good	**1.20** (0.81–1.76)	
Number of symptoms recognised	**0.99** (0.94–1.04)	
Risk averse	**1.52**** (1.06–2.18)	**1.46**** (1.05–2.04)
Low confidence in ability to recognise OC symptoms	**0.72*** (0.51–1.03)	**0.73**** (0.52–1.02)
Constant	**0.08** (0.02–0.30)	**0.09***** (0.05–0.14)
Model fit statistics		
LL	-2885.68	-2908.33
Pseudo R^2	0.03	0.03
N	250	250

10.5 Discussion

10.5.1 Key findings

This chapter quantifies preferences towards the benefits and harms of potential screening tests for ovarian cancer, with a particular focus on test performance characteristics. The results of this chapter provide a basis for understanding the minimum requirements for acceptability and expected uptake rates for potential future ovarian screening programmes.

Mixed logit results revealed the number of ovarian cancer deaths was considered the most important attribute overall (relative importance score of 0.42 (95% CI: 0.40-0.44), followed by the rate of false positive results (0.30, [95% CI: 0.30-0.30]). Overdiagnosed cancers and false negative results appeared to be of similar and of lower importance. Responses to the attribute ranking exercise revealed similar findings, although here overdiagnosis was ranked least importance behind false negative results. MRS calculations revealed respondents were willing to accept additional harms in exchange for the benefit of reductions in ovarian cancer mortality.

Demand for screening

Although results generally indicate a strong overall preference for screening, ,ixed logit estimates also demonstrated significant heterogeneity in preferences. Logistic regression revealed those who were employed and those who regularly attended cervical screening were less likely to forgo screened whereas those who considered themselves at low risk of ovarian cancer, those experiencing low levels of worry about ovarian cancer, and risk averse individuals were more likely to opt for no screening. However, the overall explanatory power of the model was low, suggesting opt-out decisions may be more intrinsically motivated.

Qualitative analysis of reasons for serial non-testers further confirmed low perceived risk of ovarian cancer as a key driver of choosing to forgo screening alongside attribute-driven reasons (e.g. unacceptable risk-benefit ratios).

Scenario analysis was used to further understand demand for testing. Intended participation in a screening program was generally high, with 78% (95% CI: 66-89%) of the population indicating they would opt for screening even under the current circumstances where screening provides no reduction in mortality and is associated with harms in the form of false positives and false negatives.

Preference heterogeneity

A five-class latent class logit model was used to further explore the heterogeneity highlighted by the ML model. Each class demonstrated a distinct set of preferences. Class membership was often associated with sociodemographic characteristics. For example, membership for classes with a preference for no screening (classes 3 and 4) was associated with reduced ovarian anxiety, lower perceived ovarian cancer risk and reduced cervical screening participation. Oppositely, members of classes with a strong preference for screening (classes 1 and 2) were more likely to consider themselves at higher risk of developing ovarian cancer, know someone diagnosed with ovarian cancer and were likely to regularly attend existing screening. Class 5 had the highest membership probability (30.9%). Similar to the full-population ML model, this group strongly prioritised ovarian cancer deaths. Interestingly, false positive results had no impact on utility for class 5, despite being the most important attribute within preference classes 1 and 2, together representing 39% of the population.

Factors affecting ability to complete choice tasks

Due to the nature of the DCE task and in particular, the inclusion of risk information in all attributes, it was important to investigate how personal factors such as numeracy skills or perceived task difficulty may systematically influence responses.

As expected, a large proportion of the sample reported finding the task difficult or very difficult (117/250; 47%). Nonetheless, completion rates remained high (just four people dropped out part way through the survey). Self-reported task-difficulty was found to significantly affect preferences within subgroup analysis and the latent class logit model. Respondents describing the task as difficult or very difficult were most likely to

belong to preference class 1 or 2. Both classes displayed a high preference for testing and placed highest importance on risk of false positive results. Members of this class appeared to consider all attributes and suggest willingness to trade between benefits and harms.

Numerical ability also appeared to be a source of preference heterogeneity during both subgroup and latent class analysis. Those displaying high numerical ability appeared more test-averse and placed increased importance on mortality and reduced importance on overdiagnosis.

Neither numerical ability nor task difficulty affected opt-out behaviour. Furthermore, a heteroscedastic logit confirmed that neither factor was a significant source of scale heterogeneity, indicating observed differences were due to genuine preference heterogeneity rather than increased error variance due to difficulties completing or understanding the choice tasks.

Finally, approximately 5% (12/250) of respondents failed the rationality check question by choosing the objectively inferior test option. Low occurrence meant further analysis of reasons for failure were not possible (e.g. association with task difficulty, numeracy etc.). Exclusion of these individuals had no significant impact on model estimates.

10.5.2 Key implications

In general, participants demonstrated a willingness to trade between the benefits and harms of ovarian cancer screening.

Potential benefits of ovarian cancer screening are limited by the low disease prevalence compared to other cancers where screening programmes are currently provisioned (e.g. breast and colorectal cancer) meaning a universal screening programme may never be feasible, particularly those utilising currently available modalities. The evidence from this study addresses key criteria considered by the UK National Screening Committee when assessing potential national screening programmes, by providing a valuable reference for assessment if and when candidate

tests emerge (UK National Screening Committee, 2015). Specifically, this study provides evidence on:

i. the acceptability of a potential screening programme
ii. the balance between benefit and harms from a public perspective.

Scenario analysis demonstrated that whilst for most participants screening decisions are dependent on test performance characteristics. for others screening decisions appeared to be determined on extrinsic factors. For instance, where a screening test provided a 100% reduction in mortality with zero potential harms, almost all participants would opt to undergo testing. However, even in the current situation where testing offers no benefits and results in exposure to potential harms such as false positive or false negative results, a significant proportion of the sample still wish to undergo testing.

Examination of reasons for opting out provides key insights into how screening uptake may be optimised by revealed key characteristics that may drive screening decisions beyond test characteristics. Low confidence in the ability to recognise symptoms and those experiencing higher levels of ovarian cancer anxiety or self-perceived cancer risks increased the willingness to be testing indicating increasing public awareness of symptoms and risk factors of ovarian cancer may help to empower patients and increase informed decision-making limiting potential harms of sub-optimal screening.

On the other hand, the relationship between reduced participation in cervical cancer screening and intentions to be screened for ovarian cancer implies for some, there may be a more fundamental attitude against screening in general.

Findings from this study also have important implications for current practice. Surveys of GPs have found that *ad hoc* screening of low-risk women is not uncommon with approximately 30% of GPs reporting ignoring guidelines by offering screening to asymptomatic women (Baldwin *et al.*, 2012; Ragland *et al.*, 2018). Importantly, results from this study indicate that this practice may be misaligned with the preferences of patients at present, as current screening modalities provide little-to-no benefit in terms of survival or stage of diagnosis. From a patient perspective, it is likely that consent to

undergo such testing is based on a misunderstanding or naivety about the balance between benefits and harms, as studies suggest almost all women believe ovarian cancer screening leads to reduced mortality (Fallowfield *et al.*, 2010). However, screening appears to offer additional benefits for a significant proportion of patients even when where there is no clinical benefit on average.

More generally, this study highlighted the ability of the public to engage with complex screening information. However, many respondents expressed surprise and dissatisfaction at the degree of potential harms described within the scenarios within the "additional comments" section of the survey implying a current lack of awareness around the magnitude of harms relating to not only ovarian cancer screening but screening tests more generally, given overdiagnosis levels were based on estimates relating to breast cancer screening. This finding highlights clear communication surrounding the benefits and harms of screening tests is needed, particularly if the aim is to encourage informed decision-making rather than persuaded coercive action.

Similarly, a large proportion of the population indicated they were unsure of their risk of ovarian cancer and did not feel confident in their ability to identify symptoms of ovarian cancer. Increasing education and awareness around ovarian cancer, encouraging help-seeking behaviour once symptoms arise and interventions to reduce mitigating lifestyle factors could provide a complementary or alternative strategies to improving ovarian cancer outcomes and are particularly important given a universal screening programme may never be achievable.

10.5.3 Comparison with existing literature

The importance of ovarian cancer deaths follows trends seen in the broader screening DCE literature as seen in Chapter 4. However, the importance placed on false positive results particularly by classes 1 and 2 within the latent class logit model is less common, with no studies included within the systematic review in Chapter 4 identifying specificity as the most important attribute. On average, individuals were willing to accept 205 (95% CI: 162 – 249) false positive results per 1 life saved per 10,000 people screened over a period of 10 years, despite the potential consequences (i.e. 3% of people undergoing unnecessary surgery). Ability to make comparisons with

findings from previous DCE studies is limited by the cancer site and the framing of the WTA calculations, with many studies opting to express MRS in terms of WTA increased deaths in exchange for improvements in risks. Sicsic *et al.* (2018) estimated women were willing to accept a lower rate of 47.8 (95% CI: 24.9-70.8) false positives per 1 breast cancer death avoided. Conversely, results from a cross-sectional survey directly eliciting the acceptability of false positive results in breast cancer found women were much more tolerant of such results, with 63% believing that false positives of 500 or more per life saved was reasonable, and 37% willing to endure false positive rates of 10,000 or more (Schwartz *et al.*, 2000) .

On the other hand, overdiagnosed cancers appeared to be less tolerated within this study in comparison to Sicsic *et al.* (2018) where on average 14.1 (95% CI:12.9-15.2) additional overdiagnosed cases were accepted in exchange for 1 breast cancer life saved.

The overall trend towards decreased importance of overdiagnosis relative to other attributes follows findings other studies, including DCEs (Howard *et al.*, 2011; Rozbroj *et al.*, 2021). Studies generally find individuals are sceptical or even hostile towards the concept of overdiagnosis, viewing the early detection of any cancer as a positive event (Hersch *et al.*, 2013; Waller *et al.*, 2013). Instead, studies typically find the concept of overtreatment to be of greater concern to participants (Hersch *et al.*, 2013; Park *et al.*, 2015). There is further evidence of a compounding “cancer effect” whereby participants are more willing to endure risks of overdiagnosis when facing a possible diagnosis of cancer compared to other potentially serious and life limiting non-cancer conditions such as aortic aneurysms (Hurley, 2018; Phillips *et al.*, 2016).

Several studies support the findings that decisions to undergo screening may be driven by factors external to test efficacy or beliefs about the curability of the disease (Salsman *et al.*, 2004). Specifically, this study follows Bennett *et al.* (2018) in finding low perceived cancer risk, low cancer anxiety and increased confidence in the ability to spot symptoms all increase the likelihood of forgoing screening. Similarly, examination of test acceptability of participation enrolled on an ovarian cancer screening trial found high rates of self-perceived risk (Holman *et al.*, 2014). More

recently, de Bekker-Grob *et al.* (2020) demonstrated similar findings within a DCE study, finding the 8-76% of non-participation behaviour in colorectal cancer screening was attributable to respondent characteristics respondent characteristics particularly the individual's attitude towards screening and previous screening behaviour, as opposed to the characteristics of specific tests.

10.5.4 Strengths and limitations

Careful steps were taken during the DCE development and piloting stages to facilitate respondent understanding as much as possible by following the latest risk presentation guidance (i.e. use of natural frequencies and consistent denominators). However, the communication of risk attributes was a clear challenge within this study. Despite following the best available guidance, a significant proportion of respondents still indicated the task was "very difficult" and over 20% of respondents indicated they had ignored attributes because "there were too many characteristics to look at" or attributes were unclear. The issue was compounded by the low prevalence of ovarian cancer, meaning risks were very small. It is possible presenting risks as natural frequencies with a denominator of 10,000 over a 10-year time horizon made it difficult for people to comprehend their individual risk and added to the cognitive burden of the choice tasks.

A further limitation within this study was the inability to stratify attribute levels according to age or underlying risk of cancer due to a current lack of data on test performance in specific populations. Studies of cervical and breast screening have demonstrated the effectiveness of screening at reducing mortality varies with age (Duffy *et al.*, 2020; Sasieni *et al.*, 2009; Wang *et al.*, 2017). Diagnostic testing of symptomatic women for ovarian cancer using CA125 tests has also been shown to be more effective for older patients perhaps due to the higher prevalence of cancer or the type of tumour (Funston *et al.*, 2020a). Similarly, screening of high-risk individuals is likely to be more effective due to higher incidence rates.

10.5.5 Future research

This study highlights some key areas for future research. Firstly, observed variations in preferences and willingness to trade across participants suggests any 'one-size-fits

all' screening programme would fail to fulfil the priorities of the population overall. Examination of opt-out behaviour indicated that perceived risk of cancer plays an important role in screening decisions. Risk-stratified screening program or limiting screening to high-risk individuals only based on genetic and/or lifestyle factors may be a potential solution, both in terms of acceptability and clinical efficacy. Risk-stratified screening has been considered in other areas, such as breast screening (Ghanouni *et al.*, 2020b). However, studies aiming to understand the acceptability of, and preferences for such strategies are crucial. Furthermore, ineffectiveness of ovarian screening even within high-risk populations means screening is not routinely offered by the NHS, therefore understanding the preferences of high-risk individuals is also needed, particularly in the context of alternative preventative strategies such as chemoprevention or preventative surgery.

Secondly, this study focused on the importance of test-performance characteristics based on the findings of the prioritisation BWS study described in Chapter 9. However, studies have demonstrated that public understanding of the test characteristics of screening programs is low, but screening is generally viewed favourably regardless (Gigerenzer *et al.*, 2009; Hoffmann & Del Mar, 2015). These findings suggest that, in reality, screening behaviours may instead be driven by service delivery factors that impact the convenience and overall experience of screening. Existing DCEs relating to cancer screening described in Chapter 4, provide an extensive evidence base to draw from when considering the influence of service delivery attributes on cancer screening: however, given the low prevalence of ovarian cancer, willingness to endure inconveniences and disruptions associated with screening may be less tolerated meaning an additional study may be of value, particularly once a viable screening modality emerges. Evidence from trial-based settings demonstrate the impact of test experience such as pain, embarrassment or inconvenience had very minimal (1-2%) impact on willingness on the acceptability and adherence of future screening (Holman *et al.*, 2014). It remains unclear if this finding is transferable to a general public setting. The presentation of risk within DCEs presents a key area for future research. The challenges of communicating health risks to public audiences are well-documented within the DCE literature and more broadly (Harrison *et al.*, 2014). In a recent publication, understanding how attribute presentation formats improve understanding

within choice tasks was highlighted as a methodological priority by a panel of preference elicitation experts (Smith *et al.*, 2021). However, studies examining alternative risk presentation methods in the context of DCEs are limited (Vass *et al.*, 2018a). A key area for future studies is the investigation if and how preferences and response quality differ based on risk formats and, in particular, the use of novel risk presentation formats within studies containing large amounts of risk information.

10.6 Chapter summary

This chapter quantified the trade-offs between the benefits and harms of ovarian cancer screening. Currently ovarian cancer screening is not recommended as available screening methods do not offer any benefits in terms of mortality reduction. The results of this study provide a useful resource for assessing the acceptability of future screening modalities which may become available in the future. Overall, results suggest the ability to reduce ovarian cancer deaths is the most importance test performance characteristic; however, there was significant heterogeneity across participants with almost 40% prioritising the risk of false positive results. Scenario analysis was used to estimate uptake of potential screening tests with results suggesting approximately one third of participant would decline screening even where there was a 100% reduction in mortality and no associated risks, indicating factors beyond test performance are important to screening participation.

11 Methodological extensions part 2: Attribute non-attendance

11.1 Chapter introduction

This is the second methodology-focused chapter of this thesis. This chapter examines the prevalence of stated attribute non-attendance (ANA) across the responses from the two applied DCEs which focus on preferences towards diagnostic testing and screening for ovarian cancer. The motivation for this methodological extension arose from early qualitative piloting results in Chapter 6 which suggested that counter to traditional decision theory, a sub-section of respondents may not consider all attributes when making decisions between choice alternatives. Due to the emergent nature of this research aim, the methods and results from each setting (diagnostic and screening) are described chronologically within this chapter.

11.2 Background

Under the standard DCE approach, respondents are assumed to consider all attributes presented within each alternative when making decisions during choice tasks. However, piloting (Chapter 6) suggested that in reality, some respondents appeared to make their choices based on a subset of attributes whilst seemingly disregarding others. This is known as attribute non-attendance (ANA).

The presence of ANA implies that individuals do not make the assumed trade-offs between attributes and attribute levels—a violation of the continuity axiom described in Chapter 3 (Gowdy & Mayumi, 2001; Rosenberger *et al.*, 2003). As a result, unidentified patterns of ANA may lead to biases in model estimation and increase the perception of preference heterogeneity (Alemu *et al.*, 2013). Addressing ANA is particularly important when estimating WTP or WTA. Since ANA represents non-compensatory behaviour, incorrectly assuming all attributes are equally attended may artificially inflate estimates (Campbell *et al.*, 2008; Lancsar & Louviere, 2006; Lockwood, 1996; Scarpa *et al.*, 2013).

Attribute non-attendance is a longstanding behavioural theory relating to the observation of choice behaviour (Lockwood, 1996). A range of methods have been

proposed for detecting and accommodating ANA in DCE studies. These methods fall into two broad approaches: stated ANA, wherein respondents are explicitly asked which attributes attended/ignored; and inferred ANA, wherein ANA can be empirically estimated based on the specification of the choice model (Collins, 2012).

Several studies have investigated ANA within health-focused experiments, however, findings relating to the impact of ANA are varied (Doherty *et al.*, 2021; Erdem *et al.*, 2015; Hole *et al.*, 2013; Hole *et al.*, 2016). Some studies find the accommodation of non-attendance has little or no impact on model estimates (Hole *et al.*, 2013). Whereas alternative studies find adjusting for ANA has a significant impact on choice modelling and can lead to improved model fit and changes in the marginal rates of substitution between attributes (Erdem *et al.*, 2015; Xiao *et al.*, 2022).

The purpose of this methodological extension was to measure the rate of self-reported ANA and investigate how this affected model estimates. The extension primarily focuses on the level and impact of stated ANA, as opposed to econometrically-inferred non-attendance. Analysis of self-reported responses within the diagnostic DCE revealed disparities between self-reported behaviour between ANA questions. This led to the addition of a further research question during the screening study aiming to understand how framing of attribute non-attendance questions may influence responses. The remainder of the chapter presents the investigations of attribute non-attendance relating to each study successively.

11.3 Attribute non-attendance behaviour in the diagnostic testing DCE

11.3.1 Aims

Investigations of attribute non-attendance in the diagnostic setting used survey versions 1-3 presented in Chapter 7.

The objectives of the research were:

i. To measure the rate of self-reported attribute non-attendance across respondents and explore reasons for non-attendance
ii. To understand how stated non-attendance may influence choice model estimates, namely MRS

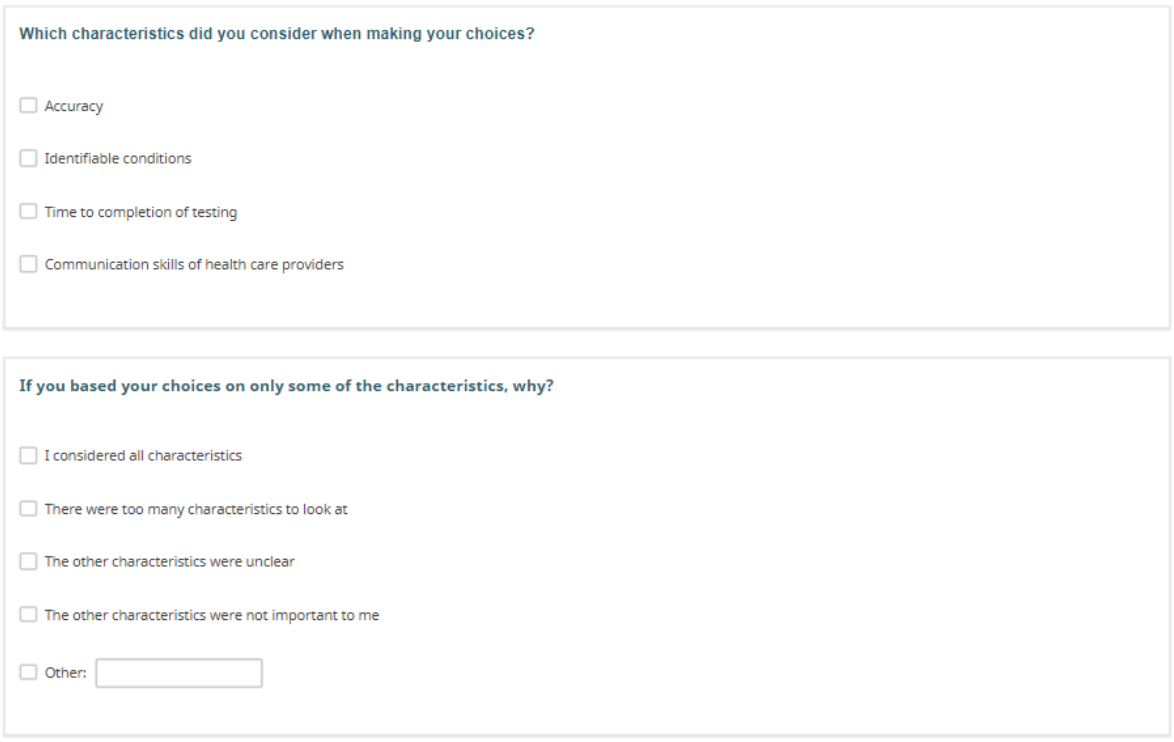
iii. To explore sociodemographic determinants of attribute non-attendance

11.3.2 Methods

11.3.2.1 Measuring the rate of self-reported attribute non-attendance

The survey included a follow up question after the final choice task asking respondents "*which attributes did you consider when making your choices?"* to investigate whether respondents ignored any of the attributes when making their choices (Figure 11.1). Any unselected attributes were considered to be non-attended by that respondent. If a respondent indicated not considering one or more attribute, an additional follow up question asked about the reasons for ANA. A list of reasons adapted from Alemu *et al.* (2013) was provided but respondents could also give their own explanation using a free-text "other" option.

Figure 11.1: Survey questions used to investigate stated ANA



11.3.2.2 Analysis of stated ANA responses

How to best incorporate stated non-attendance responses into choice models is debated. Early studies typically assume a non-attended attribute does not contribute

to the individual's utility function at all and therefore restrict non-attended attribute parameters to zero (Hensher *et al.*, 2005a). This approach implies observed choices essentially provide no information concerning the respondent's preferences for the ignored attribute. Constraining model coefficients to zero for respondents that indicate not attending an attribute relies on the assumption that stated ANA responses are completely accurate. However, several studies have found respondents demonstrate a non-zero sensitivity for attributes they indicated not attending (Carlsson *et al.*, 2010; Hess & Hensher, 2010; Nguyen *et al.*, 2015). This finding implies that respondents who claim to have ignored a given attribute may instead assign a lower (than the population average) weight to the attribute. In response, this chapter followed previous studies in using interaction terms to incorporate stated ANA responses within estimated models (Carlsson *et al.*, 2010; Nguyen *et al.*, 2015). The mean coefficient for each parameter represents the preferences of respondents who reported considering the attribute during decisions (α_0) similar to a typical choice model where full attendance is assumed. Interaction terms (α_1) between an attribute and a dummy variable then represent stated ANA (equal to 1 if the attribute is ignored and 0 otherwise) represent the difference in utility for those who reported not considering an attribute. The sum of these two terms represents the preferences of respondents who reported not considering (i.e. ignoring) a particular attribute. The model was estimated using a mixed logit model with 1000 Halton draws.

ANA-adjusted willingness to accept estimates are estimated following Nguyen *et al.* (2015). First aggregate mean ANA-adjusted coefficients were calculated for each attribute as the share of respondents who attended the attribute x α_0 + share of respondents who ignored the attribute x (α_0 + α_1). Next, ANA-adjusted WTA was calculated by dividing the adjusted mean coefficient of each attribute by the adjusted mean coefficient for the time to diagnosis attribute to give an adjusted willingness to wait estimate. Confidence intervals for the WTA estimates were estimated using the Delta method. Estimates between ANA-adjusted and full attendance model were compared using T-tests.

Finally, logistic regression was also used to explore the relationship between respondent characteristics and ANA for each attribute.

11.3.3 Results

Table 11.1 shows the self-reported ANA rates of respondents. Just 30% of respondents reported basing their decisions on all four attributes. Most commonly, participants indicated basing decisions on accuracy only (87/450; 19%) or timing and accuracy combined (60/450; 13%). These results were verified by a manual inspection of non-trading behaviour which found 61 (13.6%) respondents always chose the alternative with the highest level of accuracy, whereas choices based on the dominance of time to diagnosis (3/450; 1%), identifiable conditions (2/450; <1%) and communication (2/450; <1%) were rare. The most commonly reported reason for not considering some attributes was due to lack of importance (58/450; 13%), however, when asked about the reasons for attribute non-attendance, 79% (357/450) of respondents selected the option “I considered all characteristics”, which appears to be a direct contradiction to the previous question (Table 11.2).

Table *11.1*: Self-reported attribute non-attendance-combinations

Attribute combinations reported as not considered	N (%)
Considered all	135 (30%)
Accuracy	1 (<1%)
Timing	5 (<1%)
Communication	43 (10%)
Conditions	15 (3%)
Accuracy-timing	2 (<1%)
Accuracy-timing-conditions	3 (1%)
Accuracy-timing-communication	15 (3%)
Accuracy-conditions-communication	28 (6%)
Accuracy-conditions	1 (<1%)
Accuracy-communication	3 (1%)
Timing-communication	41 (9%)
Timing-conditions	11 (2%)
Timing-conditions- communication	87 (19%)
Conditions-communication	60 (13%)

Table 11.2: Reasons for non-attendance

Reason	N (%)
“I considered all characteristics”	357 (79.3%)
“The other characteristics were not important to me”	58 (12.9%)
“The other characteristics were unclear”	1 (0.2%)
“There were too many characteristics”	20 (4.4%)
Other	18 (4.0%)

11.3.3.1 Interaction model

Results of the interaction model used to adjust for self-reported ANA is shown in Table 11.3. A likelihood ratio test demonstrated a significant improvement in model fit when adjusting for ANA compared to the mixed logit model which assumed full-attendance[10].

[10] LL= -3432.17 for the pooled mixed logit model of DCE versions 1-3 assuming full attribute attendance presented in Chapter 7

All interaction terms were significant indicating there are differences in preferences between the respondents who reported considering and those who reported not considering a given attribute. Each interaction term had the expected, opposite sign to the mean attribute coefficients demonstrating those who reported ignoring an attribute were more likely to place less emphasis on that attribute during deliberations.

Table *11.3*: Mixed logit results incorporating ANA using interaction terms

	Mean (95% CI)	**SD (95% CI)**
Attributes		
Accuracy	**0.21*** ** (0.19 – 0.23)	**0.11*** ** (0.10– 0.13)
Time to diagnosis	**-0.90*** ** (-0.99 – [-0.80])	**0.66*** ** (0.48 – 0.64)
Identifiable conditions: cancer plus related conditions	**1.59*** ** (1.40 – 1.78)	**0.91*** ** (0.75 – 1.08)
Communication: fair	**1.58*** ** (1.36 – 1.80)	**0.23*** (-0.04 – 0.50)
Communication: good	**2.09*** ** (1.83 – 2.35)	**0.72*** ** (0.52 – 0.91)
Neither test ASC	**-3.46*** ** (-2.98 – [-1.64])	**4.36** (3.69– 5.03)
Interaction terms		
Accuracy x accuracy not considered dummy	**-0.12*** ** (-0.15 – [-0.08])	**0.01** (-0.02 – 0.04)
Time to diagnosis x not considered dummy	**0.51*** ** (0.38 – 0.64)	**0.08** (-0.09 – 0.26)
Identifiable conditions: cancer plus related conditions x not considered dummy	**-1.11*** ** (-1.41 – [-0.80])	**0.25** (-0.11 – 0.60)
Communication: fair x not considered dummy	**-0.68*** ** (-0.94 – [-0.42])	**0.06** (-0.42 – 0.54)
Communication: good x not considered dummy	**-1.00*** ** (-1.26 – [-0.74])	**0.13** (-0.32 – 0.58)
Model fit statistics		
LL	-3,329.2	
Observations	20,598	
N	450	

Key: ***significant at 99% confidence level; **significant at 95% confidence level; *significant at 90% confidence level

11.3.3.2 Willingness to wait estimates

Willingness to wait estimates after adjusting for ANA are reported in Table 11.4. Comparisons of WTA estimates did not identify any significant differences between the adjusted and non-adjusted model estimates. Separate WTA estimates were also calculated for each attribute according to reported attribute attendance patterns. These estimates demonstrated significant differences between individuals who reported attending or not attending each attribute (Appendix 11.1). As expected, non-

attendance of time to diagnosis resulted in the largest change in willingness to wait estimates for all attributes.

Table 11.4: Willingness to accept estimates from the ANA-adjusted model

	ANA-adjusted estimates	Full attendance estimates	p-value*
Accuracy			
Per 1%	0.27 (0.24–0.30)	0.28 (0.25–0.31)	0.67
Identifiable conditions			
Cancer only	-	-	-
Cancer plus additional related conditions	1.60 (1.37–1.83)	1.70 (1.45–1.96)	0.57
Communication			
Poor	-	-	-
Fair	1.64 (1.42–1.86)	1.67 (1.44–1.90)	0.86
Good	1.98 (1.72–2.25)	2.02 (1.74–2.30)	0.84
*p-value from t-tests to test for differences in marginal rates of substitution between the ANA-adjusted model and the mixed logit model assuming full attendance			

11.3.3.3 Determinants of non-attendance – logit model

Logistic regression results investigating associations between respondent characteristics and of self-reported ANA are shown in Table 11.5. In general, relationships were limited and inconsistent across different attributes. Completion of university education was consistently associated with an approximately 50% reduction in the likelihood of ANA across all attributes. Aside from education, no other sociodemographic associations were found with non-attendance of the time to diagnosis attribute. On the other hand, non-attendance of the accuracy attribute was linked to the most respondent characteristics overall. Knowing someone diagnosed with ovarian cancer and failing the rationality check more than doubled the chance of ignoring accuracy. Oppositely, experiencing low worry about ovarian cancer and increased selection of the opt-out alternative reduced the likelihood of ANA for this attribute. Desiring an active role in medical decision-making was linked to a reduction in the chance of ignoring identifiable conditions. As was increased help-seeking behaviour. Alternatively, increased age was associated with a moderate increase in the overlooking of this attribute. Finally, the likelihood of considering communication was linked to a reduction in help-seeking behaviour.

Table *11.5*: Logit model investigating associations between respondent characteristics and ANA for each attribute. Coefficients are expressed as odds-ratios (95% CI)

	Accuracy	Time to diagnosis	Identifiable conditions	Communication
Age	**1.00** (0.96–1.04)	**1.00** (0.97–1.02)	**1.03**** (1.01–1.06)	**1.01** (0.98–1.03)
Attended university	**0.49**** (0.26–0.90)	**0.54***** (0.36–0.81)	**0.52***** (0.35–0.80)	**0.50***** (0.32–0.77)
Ethnicity-white	**0.41** (0.14–1.19)	**1.38** (0.54–3.53)	**0.55** (0.23–1.33)	**0.62** (0.25–1.51)
Previously tested for ovarian cancer	**1.02** (0.80–1.31)	**1.04** (0.90–1.21)	**0.91** (0.76–1.08)	**1.20** (0.86–1.68)
Know someone diagnosed with ovarian cancer	**2.35**** (1.15–4.81)	**0.94** (0.55–1.62)	**0.96** (0.56–1.64)	**1.09** (0.64–1.85)
Low ovarian cancer worry	**0.49**** (0.24–0.97)	**0.74** (0.44–1.24)	**0.64** (0.38–1.09)	**0.81** (0.48–1.36)
Number of symptoms recognised	**0.94** (0.85–1.04)	**0.98** (0.92–1.05)	**0.96** (0.90–1.03)	**0.99** (0.93–1.06)
Confident in ability to recognise OC symptoms	**0.70** (0.25–1.94)	**1.30** (0.69–2.45)	**1.40** (0.73–2.68)	**1.16** (0.60–2.25)
Self-reported health: Very good—good	**1.47** (0.73–2.99)	**1.37** (0.89–2.11)	**1.00** (0.65–1.55)	**1.30** (0.84–2.02)
Risk averse	**1.03** (0.45–2.33)	**1.05** (0.65–1.69)	**1.23** (0.76–1.99)	**1.01** (0.63–1.60)
Helpseeking (would seek GP advice after 1 week of OC symptom onset)	**0.94** (0.48–1.84)	**0.91** (0.58–1.42)	**1.09** (0.70–1.71)	**0.61**** (0.39–0.94)
Would like an active role in decision-making	**0.67** (0.34–1.33)	**1.01** (0.62–1.67)	**0.44***** (0.27–0.71)	**0.73** (0.44–1.20)
Currently has an active role in decision-making	**1.21** (0.57–2.55)	**1.35** (0.84–2.16)	**1.03** (0.64–1.66)	**0.83** (0.53–1.30)
Found DCE difficult/very difficult	**1.81*** (0.96–3.43)	**1.09** (0.72–1.65)	**0.63**** (0.41–0.95)	**0.93** (0.61–1.42)
Failed rationality check	**2.60***** (1.36–4.98)	**0.70** (0.43–1.13)	**0.94** (0.59–1.51)	**0.93** (0.58–1.50)
Total number of opt-out selections	**0.81**** (0.26–0.90)	**0.99** (0.90–1.10)	**1.05** (0.96–1.15)	**1.01** (0.92–1.11)
Total number of indifference selections	**1.07** (0.86–1.32)	**0.96** (0.84–1.09)	**0.94** (0.83–1.06)	**0.93** (0.82–1.05)
Completed the survey in under 10 minutes	**1.21** (0.27–5.34)	**0.88** (0.29–2.61)	**1.61** (0.57–4.61)	**0.74** (0.26–2.07)
Cancer risk level (2% dummy)	**0.89** (0.43–1.84)	**1.30** (0.80–2.11)	**1.08** (0.67–1.75)	**0.99** (0.61–1.61)
Cancer risk level (3% dummy)	**0.66** (0.31–1.44)	**0.90** (0.55–1.49)	**0.95** (0.57–1.56)	**1.02** (0.61–1.69)
Constant	**0.58** (0.04–9.52)	**0.79** (0.15–4.11)	**1.40** (0.27–7.38)	**4.23*** (0.80–22.24)
Model fit statistics				
LL	-9.115	-18,239	-18,496	-18,287
Pseudo R^2	0.13	0.03	0.07	0.05
Significance key: * 90%, **95%, ***99%				

11.3.4 Discussion of findings on ANA during the diagnostic DCE

Self-reported attribute non-attendance appeared to be high within this study. Overall just 30% of respondents indicated they considered all the attributes when making choices during the DCE. Self-reported ANA was highest for the communication attribute with 62% of respondents indicating they had not considered this attribute during deliberations. Adjustments for ANA suggested non-attending respondents assigned a significantly lower, but non-zero value to attributes implying non-attended attributes were valued to a lesser extent rather than completely ignored. Adjusting for stated ANA resulted in improved model fit. However, despite the high levels of self-reported ANA, no differences were found between WTA estimates between the ANA-adjusted model and standard mixed logit model assuming full attendance. Notably, there appeared to be discordance in estimated stated ANA between questions, with 79% of respondents indicating they "considered all attributes" during a follow up question exploring for reasons for non-attendance.

Piloting results highlighted ANA as a potential problem for a sub-section of respondents, however, the high rates on stated ANA experienced during this study were surprising given the rigorous attribute selection process described in Chapter 6. The discontinuity between self-reported non-attendance and reasons for ANA means the true rates on non-attendance within this survey are unclear. However, the large proportion of respondents indicating they considered all attributes supports the implication "non-attended" attributes are valued to a lesser extent rather than disregarded completely. However, upon reflection disparities in stated ANA may also be due to ambiguity in question wording. Respondents were asked to indicate which attributes they "considered" rather than explicitly asked which attributes were "ignored".

11.4 Attribute non-attendance behaviour in the ovarian cancer screening DCE

11.4.1 Aims

The investigation of attribute non-attendance behaviour within the screening DCE aimed to build on findings from the earlier diagnostic study. In addition to addressing objectives i.-iii. as outlined in Section 11.3.1. This later study also aimed to understand the impact of question framing on self-reported ANA rates. This additional aim was motivated by the unexpectedly high rates of self-reported ANA in the diagnostic study and the disparities between ANA rates and the follow-up question eliciting reasons for non-attendance. It was hypothesised this disparity was caused by ambiguity in question wording (i.e. due to asking respondents which attributes they "considered" rather than which attributes they "ignored").

11.4.2 Methods

11.4.2.1 Measuring the rate of self-reported ANA: investigation of framing effects

To understand the contribution of framing effects on self-reported ANA, the DCE eliciting preferences for ovarian cancer screening included two follow-up questions placed shortly after the completion of all choice tasks. The first question used identical wording to the diagnostic DCE, asking respondents to indicate which attributes they "considered" (i.e. "considered" framing). As before, unselected attributes were assumed non-attended. The second question explicitly asked respondents to identify any attributes which they had ignored during the choice task (i.e. "ignored" framing). Respondents completed both questions which were separated by a few other debriefing questions (e.g. task difficulty). the order the questions appeared was randomised between respondents to control for any position effects.

As before, an additional question asking about the reasons for ANA was asked to those who selected one or more attribute within the "ignored" framing.

11.4.2.2 Analysis of stated ANA responses

Stated ANA responses were incorporated into models using the interaction approach described in Section 11.3.2.2 to account for non-zero sensitivities in "non-attended" attributes (Carlsson *et al.*, 2010; Hess & Hensher, 2010; Nguyen *et al.*, 2015). Separate models were estimated based on the alternative stated ANA question frames. All models were estimated using mixed logit modelling with 1000 Halton draws. Models were compared based on adjusted-WTA estimates.

11.4.3 Results

11.4.3.1 Self-reported non-attendance

Responses to the self-reported attribute non-attendance debriefing questions are summarised in Table 11.6. Stated attribute non-attendance was high within this survey, with just 12.8% (considered framing) – 26.8% (ignored framing) of people indicating they considered all attributes during decision-making. Results suggested that overdiagnosed cancers were most likely to be ignored by respondents and ovarian cancer deaths was most likely to be taken into account. However, responses varied substantially when framed as "attributes ignored" and "attributes considered". Rates of non-attendance were lower when respondents were explicitly asked to indicate the attributes they ignored.

Reasons for non-attendance are summarised in Table 11.8. A perceived lack of importance was the primary reason for non-attendance.

11.4.3.2 Adjusting for attribute non-attendance

Results of the interaction model are shown in Table 11.9. The sum of the attribute and corresponding interaction coefficients represent the preference of respondents who stated they ignored the attribute. All interaction terms in both question frames were significant with the exception of the parameter associated with non-attendance and false negative results. This finding indicates that there were differences in preference between respondents who reported attending an attribute and those who did not. Interaction terms were positive meaning attributes were of reduced importance to

those who indicated they did not attend the attribute. Models adjusting for ANA following both question frames demonstrated a small but significant improvement in model fit according to LR tests. The model adjusting for ANA based on responses to the “considered” framing provided the most improvement.

Table *11.6*: Rates of self-reported attribute non-attendance using two different question frames

Attribute	“Considered” framing	“Ignored” framing
Ovarian cancer deaths	66 (26.4%)	26 (10.4%)
False negative results	126 (50.4%)	27 (10.8%)
False positive results	121 (48.4%)	45 (18.0%)
Overdiagnosed cancers	193 (77.2%)	117 (46.8%)
No attributes ignored	28 (11.2%)	67 (26.8%)

Table *11.7*: Reasons for self-reported attribute non-attendance (asked to those who reported "ignoring" one or more attributes)

Reason for non-attendance	Total respondents (%)
There were too many characteristics to look at	49 (19.6%)
The other attributes were unclear	9 (3.6%)
The other characteristics were not important to me	102 (40.8%)
The characteristics did not vary that much	5 (2.0%)
I’m not sure	3 (1.2%)
Other	21 (8.4%)

11.4.3.3 ANA-adjusted willingness to accept

Population-level WTA estimates adjusted for stated non-attendance are shown in Figures 11.2.1-11.2.3. At an aggregate-level, the incorporation of stated ANA had no significant impact on the willingness to accept increased harms in exchange for a reduction in ovarian cancer mortality irrespective of question framing in comparison to estimates derived from the full-attendance model. Similar to the diagnostic setting, reduced importance of non-attended attributes was reflected in differences in WTA

estimates between respondents based on attendance patterns. WTA estimates were significantly higher for respondents if an attribute was considered “non-attended” regardless of the question frame with exception of individuals who indicated they did not attend overdiagnosed cancers within the “ignored” framing (Appendix 2). Notably, the framing of the non-attendance question did not lead to significant differences in WTA for any attribute, however, uncertainty surrounding estimates under the “ignored” framing was much larger implying greater heterogeneity in this subgroup.

11.4.3.4 Sociodemographic drivers of attribute non-attendance

Logistic regression results investigating sociodemographic drivers of self-reported ANA are shown in Table 11.10. Two models were estimated for each attribute to account for the two alternative question framings of the non-attendance question; attributes not selected when asked which attributes were considered during decision-making and attributes selected when asked which attributes were ignored during decision-making. Colour-coding was used to aid comparisons across attributes and question frames.

Overall, few strong associations between sociodemographic drivers were identified. Increasing age was found to significantly increase the chance of ignoring the ovarian cancer deaths attribute and had a borderline effect on reducing the chance of ignoring overdiagnosed cancers. Increased numeracy skills were associated with a reduction in the chance of ignoring both false positive and false negative attributes. Interestingly, the most impactful factor was the number of times the “no screening” alternative was selected. For both ovarian cancer deaths and false negative results, increased selection of the opt-out alternative was associated with an increased chance of non-attendance. However, for false positives increased selection of the opt-out alternative was associated with a reduction in the chance of attribute non-attendance. Corroboration between the two non-attendance questions for each attribute was inconsistent. Sociodemographic factors influencing non-attendance for ovarian cancer deaths and false positive results appeared similar for both question frames, however, sociodemographic associations with false negative results and overdiagnosed cancers appeared very different between questions.

Table *11.8*: Mixed logit model including interaction terms to account for stated ANA

	"Considered" framing		"Ignored" framing	
	Mean (95% CI)	**SD (95% CI)**	**Mean (95% CI)**	**SD (95% CI)**
Attributes				
Ovarian cancer deaths	**-0.16*** (-0.18 – [-0.14])	**0.08*** (0.06– 0.10)	**-0.14*** (-0.16 – [-0.1***3])	**0.09*** (0.08– 0.11)
False negative results	**-0.08*** (-0.10 – [-0.06])	**0.04*** (0.02 – 0.06)	**-0.06*** (-0.07 – [-0.04])	**0.05*** (0.03 – 0.07)
False positive results	**-0.00*** (-0.00 – [-0.00])	**0.00*** (0.00 – 0.00)	**0.00*** (-0.00 – [-0.00])	**0.00*** (0.00 – 0.00)
Overdiagnosed cancers	**-0.11*** (-0.28 – [-0.03])	**0.03**** (0.01 – 0.06)	**-0.08*** (-0.10 – [-0.06])	**0.03**** (0.01 – 0.06)
No screening	**-1.66*** (-2.36 – [-0.96])	**4.25** (3.50 – 5.00)	**-1.79*** (-2.54 – [-1.05])	**4.90** (4.12 – 5.68)
Interaction terms				
Ovarian cancer deaths × OC deaths not considered	**0.14*** (0.10 – 0.17)	**0.00** (-0.06 – 0.07)		
False negative results x False negative results not considered	**0.00*** (0.00 – 0.00)	**0.00** (-0.00 – 0.00)		
False positive results x False positive results not considered	**0.06**** (0.03 – 0.09)	**0.02** (-0.02 – 0.06)		
Overdiagnosed cancers × Overdiagnosed cancers not considered	**0.06**** (0.03 – 0.09)	**0.02** (-0.02 – 0.05)		
Ovarian cancer deaths × OC deaths ignored			**0.12*** (0.08 – 0.17)	**0.00** (-0.05 – 0.04)
False negative results x False negative results ignored			**0.01** (0.04 – 0.06)	**0.01** (-0.08 – 0.09)
False positive results x False positive results ignored			**0.00*** (0.00 – 0.00)	**0.00** (-0.00 – 0.00)
Overdiagnosed cancers × Overdiagnosed cancers ignored			**0.04*** (0.01 – 0.07)	**0.01** (-0.03 – 0.05)
Model fit statistics				
LL	-1841.69		-1888.49	
LR test (ANA model vs full attendance mixed logit model)	139.2***			
LR test ("considered" framing vs "ignored" framing)	93.6***			
Observations	9,000		9,000	
N	250		250	

Key: ***significant at 99% confidence level; **significant at 95% confidence level; *significant at 90% confidence level

Figure 11.2: Comparison of willingness to accept estimates adjusted for self-reported attribute non-attendance

Figure 11.3.1: Willingness to accept increases in the number of false positive results in exchange for 1 less cancer death per 10,000 people screened. One-way ANOVA p-value= 0.678

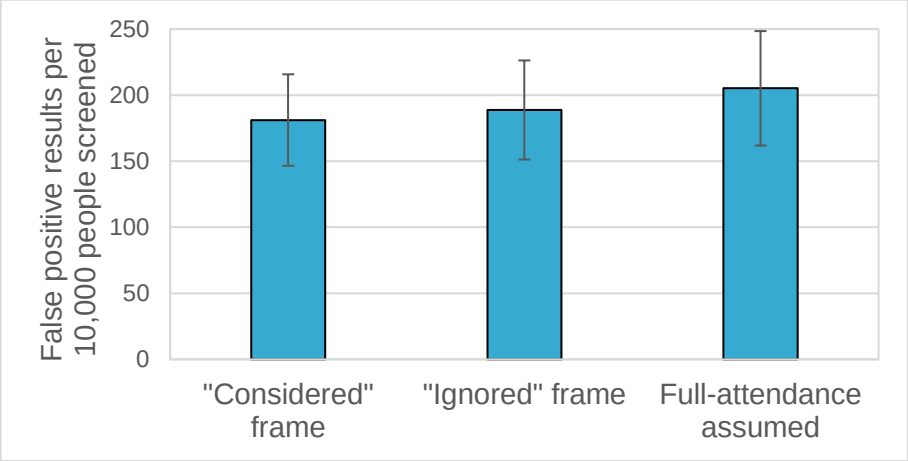


Figure 11.3.2: Willingness to accept increases in the number of false negative results in exchange for 1 less cancer death per 10,000 people screened. One-way ANOVA p-value= 0.884

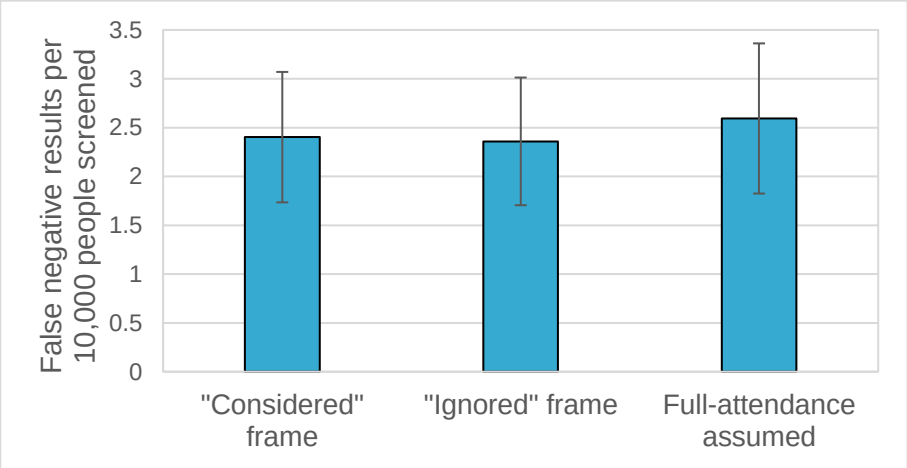


Figure 11.3.3: Willingness to accept increases in the number of overdiagnosed cancers in exchange for 1 less cancer death per 10,000 people screened. One-way ANOVA p-value= 0.797

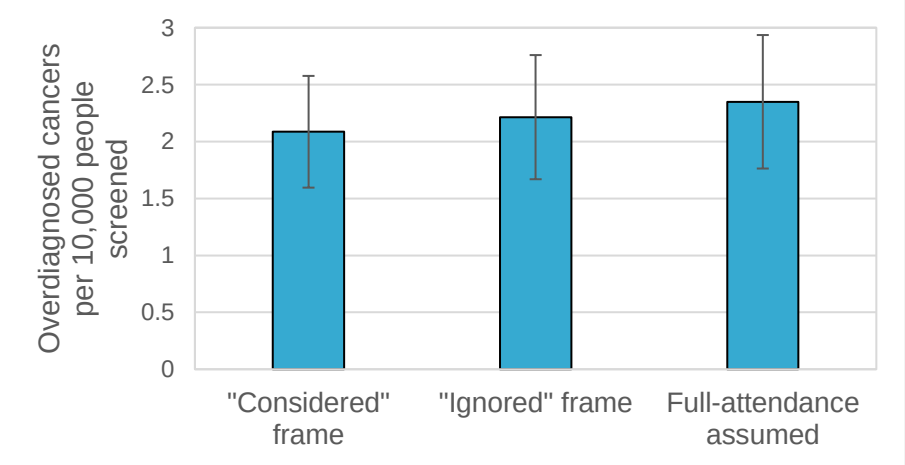


Table 11.9: Results from logit models investigating the relationships between attribute non-attendance and respondent characteristics

	Ovarian cancer deaths non-attendance		False negatives non-attendance		False positives non-attendance		Overdiagnosed cancers non-attendance	
	"Considered" framing	"Ignored" framing	"Considered" framing	"Ignored" framing	"Considered" framing	"Ignored" framing	"Considered" framing	"Ignored" framing
Age	**1.05**** (1.01–1.10)	**1.07**** (1.01–1.13)	**1.01** (0.99–1.03)	**1.00** (0.96–1.04)	**0.99** (0.97–1.01)	**1.00** (0.96–1.04)	**1.00** (0.99–1.01)	**0.97*** (0.93–1.00)
Employed	**1.70** (0.74–3.90)	**2.49*** (0.88–7.07)	**0.98** (0.62–1.55)	**0.37**** (0.15–0.90)	**0.71** (0.45–1.13)	**0.71** (0.33–1.50)	**1.13** (0.84–1.53)	**1.22** (0.65–2.31)
Know someone diagnosed with ovarian cancer	**0.67** (0.26–1.73)	**0.90** (0.28–2.93)	**0.76** (0.43–1.35)	**1.15** (0.43–3.02)	**0.84** (0.51–1.36)	**1.19** (0.53–2.67)	**0.91** (0.64–1.29)	**0.85** (0.40–1.82)
Attended university	**0.35***** (0.17–0.71)	**0.47*** (0.20–1.10)	**1.18** (0.79–1.79)	**2.37*** (1.00–5.61)	**1.34** (0.88–2.03)	**1.74** (0.89–3.42)	**0.89** (0.69–1.15)	**1.05** (0.59–1.87)
Always attends cervical screening	**0.50*** (0.25–1.00)	**0.90** (0.39–2.09)	**0.79** (0.53–1.17)	**0.93** (0.44–1.96)	**1.13** (0.73–1.74)	**0.83** (0.42–1.65)	**0.79*** (0.62–1.01)	**1.29** (0.71–2.33)
Found DCE difficult/very difficult	**1.06** (0.56–2.00)	**0.82** (0.37–1.81)	**0.72*** (0.49–1.05)	**1.32** (0.64–2.75)	**0.73** (0.49–1.09)	**1.04** (0.56–1.96)	**0.85** (0.66–1.08)	**0.74** (0.43–1.27)
Low ovarian cancer worry	**1.29** (0.54–3.09)	**1.20** (0.41–3.50)	**1.12** (0.70–1.79)	**1.38** (0.59–3.26)	**1.04** (0.67–1.62)	**0.58** (0.29–1.51)	**0.97** (0.74–1.28)	**1.24** (0.65–2.38)
Low perceived ovarian cancer risk	**1.45** (0.66–3.18)	**3.99***** (1.66–9.58)	**0.97** (0.60–1.58)	**0.96** (0.41–2.24)	**0.95** (0.56–1.60)	**1.18** (0.54–2.59)	**0.91** (0.68–1.23)	**0.55*** (0.28–1.08)
Numerical ability	**0.85** (0.60–1.19)	**0.77** (0.52–1.14)	**0.81**** (0.68–0.96)	**1.02** (0.71–1.48)	**0.82**** (0.67–0.99)	**0.73**** (0.54–0.98)	**1.02** (0.90–1.17)	**1.12** (0.86–1.46)
Self-reported health: Very good—good	**1.45** (0.71–3.00)	**0.91** (0.37–2.22)	**0.94** (0.62–1.41)	**2.11*** (0.95–4.66)	**1.12** (0.73–1.71)	**1.22** (0.61–2.44)	**1.03** (0.80–1.32)	**0.82** (0.45–1.50)
Number of children	**1.04** (0.80–1.35)	**0.84** (0.59–1.19)	**0.94** (0.80–1.11)	**0.93** (0.69–1.25)	**1.13** (0.95–1.33)	**0.92** (0.69–1.22)	**1.01** (0.90–1.14)	**1.01** (0.81–1.27)
Number of symptoms recognised	**0.92*** (0.84–1.01)	**0.91** (0.80–1.03)	**1.00** (0.94–1.05)	**1.03** (0.93–1.14)	**1.01** (0.96–1.07)	**1.07** (0.97–1.17)	**1.01** (0.97–1.04)	**1.00** (0.93–1.08)
Ethnicity- white	**0.81** (0.35–1.90)	**0.48** (0.19–1.21)	**0.95** (0.59–1.53)	**0.83** (0.32–2.18)	**1.00** (0.59–1.69)	**0.84** (0.39–1.81)	**1.08** (0.78–1.49)	**1.80** (0.88–3.67)
Total number of opt-out selections	**1.20***** (1.11–1.29)	**1.13***** (1.04–1.22)	**1.02** (0.98–1.07)	**1.10***** (1.02–1.19)	**0.82***** (0.77–0.88)	**0.86***** (0.79–0.94)	**0.98** (0.95–1.01)	**0.96** (0.91–1.03)
Constant	**0.04** (0.00–0.87)	**0.01**** (0.00–0.50)	**1.00** (0.21–4.75)	**0.07** (0.00–1.66)	**2.51** (0.48–13.17)	**2.21** (0.15–33.48)	**0.84** (0.28–2.53)	**3.68** (0.36–38.22)
Model fit statistics								
LL	-4146.3	-2857.8	-5625.6	-3653.9	-5049.6	-4572.4	-6098.2	-5732.8
Pseudo R^2	0.21	0.19	0.02	0.10	0.11	0.11	0.01	0.06
Significance key: * 90%, **95%, ***99%								

11.4.4 Discussion of findings on ANA during the screening DCE

Similar to the diagnostic setting, self-reported ANA rates were high within the DCE eliciting preferences towards ovarian cancer screening. The exploration of framing effects provides a small, but to my knowledge unique, aspect to the existing ANA literature. Rates of non-attendance varied significantly according to the question frame. Self-reported ANA rates were higher under the “considered” question frame, with just 11% of respondents indicating they considered all attributes. This rose to 27% when respondents were asked to indicate which attributes they had ignored. From the model results it is unclear which question frame more accurately captures true behaviour although asking respondents to indicate the attributes they “considered” rather than attributes they “ignored” appeared to lead to the biggest improvement in model fit.

Despite differences in the incidence of ANA, estimates of willingness to accept did not significantly according to the framing of the stated ANA question. Furthermore, WTA estimates also did not differ from the non-adjusted model estimates that assumed full attribute attendance at a population-level. However, significant differences were consistently identified between respondents based on self-reported attendance patterns within both question frames. Individuals who reported attending attributes experienced a significantly lower (but non-zero) willingness to accept harms in exchange for improvements in ovarian cancer mortality compared to those who reported non-attending particular attributes.

Several personal characteristics were associated with the probability of self-reported ANA. However, differences demonstrated limited consistency between question frames. The majority of relationships appeared to be factors relating to the ability to complete choice tasks rather than sociodemographic factors. For instance, increased numerical ability reduced the likelihood of not attending false positive and false negative attributes and university attendance decreased the likelihood of non-attendance to ovarian cancer deaths.

Forty-one percent (102/250) of respondents indicated ignoring attributes due to a lack of importance, suggesting the majority of ANA can be attributed to taste variations.

However, 20% of respondents who ignored one or more attributes indicated doing so due to a lack of clarity or the presence of too many attributes suggesting ANA may also be a simplifying heuristic for some. This finding is further supported by the prevalence of interactions between respondent factors affecting ability to complete DCE tasks (e.g. lower educational attainment, lower numerical ability, self-reported task difficulty) and the increased likelihood of ANA. This finding further highlights the importance of exploring alternative methods for communicating complex attribute information within DCEs.

11.5 Chapter summary

This chapter explored the frequency and implications of attribute non-attendance within the applied DCEs presented within this thesis using a stated attendance approach. The incidence of self-reported non-attendance behaviour was high in both a diagnostic and screening setting. Adjusting for non-attendance behaviour appeared to improve the fit of estimated models.

No differences in WTA between adjusted and unadjusted models at a population-level were identified in either case study, however differences in WTA estimates according to attendance patterns were consistently observed in both settings. This finding implies understanding and exploring ANA can add to the richness and application of DCE findings. Differences based on non-attendance patterns provide useful insights into the preference heterogeneity observed during the experiment and may be particularly important when designing policies to increase screening adherence across subgroups. If a group of individuals is more likely to ignore a specific attribute, then improvements in this dimension are less likely to improve uptake or satisfaction for this group despite what aggregate estimates may suggest. For example, within the screening DCE the finding that the increased selection of the opt-out alternative was associated with decreased non-attendance of the false positive attribute, suggests uptake could be enhanced by focusing improvements in this dimension whereas the increased likelihood of non-attendance to mortality attribute suggest improvements in this aspect may be less effective in increasing uptake.

Variability in stated ANA estimates according to framing highlights a well-acknowledged criticism of the approach. Reliance on respondents' ability to not only recall their decision-making process but also accurately evaluate their behaviour in the first instance is a further weakness of the approach. As such, there is a trend towards inferred methods to evaluate and accommodate ANA in recent years (Arora *et al.*, 2022; Doherty *et al.*, 2021; Heidenreich *et al.*, 2018; Xiao *et al.*, 2022). However, Hensher and Greene (2010) importantly note that researchers do not know whether stated or analytical methods are closer to the 'truth' in terms of determining ANA. Although, inferred methods have been shown to suitably estimate ANA when compared to observational methods such as visual non-attendance based on eye tracking methods (Yegoryan *et al.*, 2020) (visual attendance does not automatically dictate the influence of attribute during decisions).

Nonetheless, significant non-attendance interaction terms demonstrate respondents report non-attendance appear to signify a distinguishable difference in preferences and demonstrate the ability of respondents to identify differences in processing between attributes (although this is not strictly "non-attendance") (Nguyen *et al.*, 2015). Furthermore, studies typically find that stated and interred ANA, whilst not identical, are comparable and stated non-attendance methods provide useful insights into attribute processing patterns (Hensher & Greene, 2010; Hole *et al.*, 2013). More recently studies have attempted to use more experimental and observational approaches to understand the behavioural processes behind attribute non-attendance such as think aloud interviews and eye tracking (Alemu *et al.*, 2013; Heidenreich *et al.*, 2018; Vass *et al.*, 2018a; Yegoryan *et al.*, 2020). Whilst these studies provide a greater depth of information, they are resource intensive and the process of data collection may influence responses.

The overall finding of no difference in aggregate WTA observed in this study is in keeping with several previous studies (Carlsson *et al.*, 2010; Hole *et al.*, 2013; Nguyen *et al.*, 2015), but is in contrast with other studies (Campbell *et al.*, 2008; Hensher *et al.*, 2005a) where differences were observed. In a survey, 30% of authors of DCEs in healthcare reported including debriefing question(s) eliciting attribute non-attendance (Pearce *et al.*, 2020). However, reporting of stated ANA findings appears to be limited

within applied studies, perhaps greater transparency in reporting is the first step to better understanding decision-processing within DCEs and the implications for the interpretation of findings.

12 Discussion

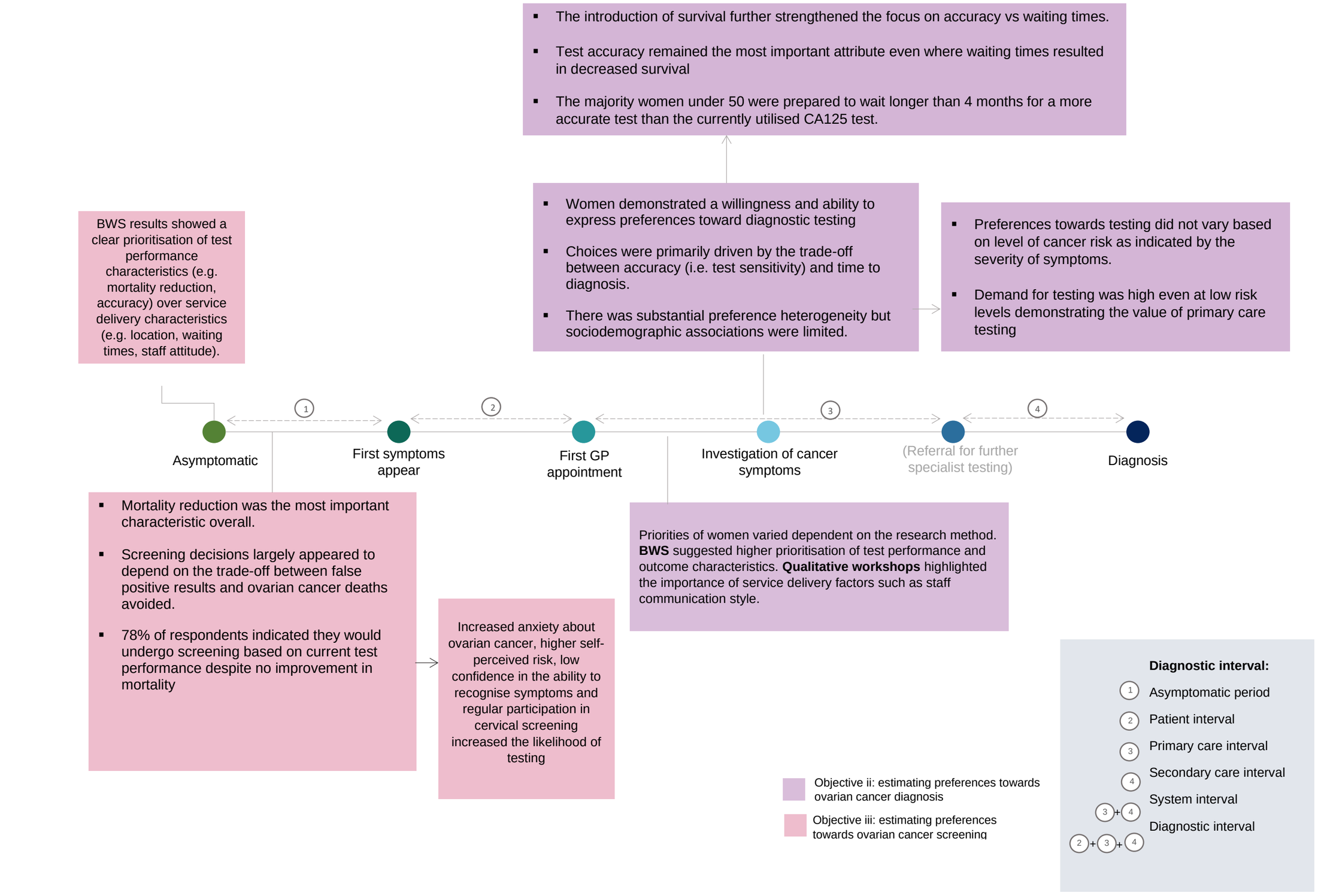
12.1 Introduction

This chapter discusses the findings and implications of the research conducted throughout this thesis. The empirical chapters within this thesis provided an in-depth discussion of results including policy implications, strengths and weaknesses and areas for future research. Therefore, this chapter takes a more holistic view by discussing the broader implications of thesis in relation to both policy making and future applications of DCEs to cancer testing.

12.2 Key findings and implications for ovarian cancer testing

A summary of the key findings relating to the policy-related research questions first introduced in Chapter 1 is provided in Figure 12.1.

Figure 12.1: Summary of the policy related findings throughout the thesis



12.2.1 Diagnostic testing for ovarian cancer

This thesis developed a DCE with several sub-versions to explore preferences towards diagnostic testing for ovarian cancer when facing different risks of cancer. Overall, demand for testing was high, even when the risk of ovarian cancer was low (1%) and attribute importance did not differ according to risk level. At an aggregate level, testing decisions were primarily driven by the trade-off between accuracy (i.e. test sensitivity) and time to diagnosis. Remaining attributes (communication during the testing process and identifiable conditions) were valued by participants - but to a lesser extent.

Mixed logit models revealed substantial preference heterogeneity across the sample, particularly in relation to the opt-out alternative and time to diagnosis attribute. Latent class modelling was used to further investigate heterogeneity. Five distinct preference classes were established. Membership was associated with a limited number of identifiable sociodemographic drivers. White respondents were more likely to place higher importance on test accuracy (Class 1) than respondents with other ethnic backgrounds. Women who reported being in less-than-good health and those who desired a more passive role in medical decisions were more likely to belong to exhibit strong test aversion (Class 2). Overall, the limited significance of sociodemographic factors indicated test decisions may be more specific to individuals based on unidentifiable factors.

A further DCE sub-version introduced an age-stratified relationship between time to diagnosis and survival. The prominence of the accuracy-waiting time trade-off observed in the earlier versions was strengthened further. Accuracy remained most important overall and time to diagnosis became increasingly important, largely at the expense of further reduction in the prioritisation of alternative identifiable conditions and communication attributes.

Choice share scenario analysis revealed an alternative testing strategy to the quick but less accurate CA125 blood test was preferred by a large proportion of people. This finding was particularly strong for women under 50 years old, where the majority of respondents were willing to wait in excess of 4 months for a test that offered greater

accuracy even when this had a significant impact on the chance of survival for those with ovarian cancer.

Results have several implications for both clinical practice and future policy decisions. Firstly, high demand for testing even where cancer risks are low and the elevated importance of accuracy and time to diagnosis suggest lowering the urgent referral threshold would be welcomed by the public (Banks *et al.*, 2014). Since this may not be feasible from an NHS perspective, the results also demonstrate the value of primary care testing for individuals experiencing low risk symptoms. Current waiting times for alternative tests such as TVUS imply that using CA125 testing to triage and expedite high risk patients is likely to be the most acceptable test strategy for many patients. However, preference heterogeneity and variations in test performance indicate TVUS may be a preferred first-line option for younger patients if available (assuming this provides higher test sensitivity). More generally, this finding suggests universal guidance on test pathways may not be appropriate. Finally, results demonstrate women's willingness and ability to convey preferences towards diagnostic investigations. Most respondents also expressed a desire for increased input in their medical care during primary care consultations highlighting the importance of individualised and collaborative decision-making in the field of diagnosis as well as with treatment decisions.

12.2.2 Preferences towards an ovarian cancer screening programme

At the population level, screening decisions largely appeared to depend on the trade-off between false positive results and ovarian cancer deaths avoided. False negative results and overdiagnosed cancers significantly impacted utility but were less prioritised. However, further analysis revealed significant preference heterogeneity across all attributes, which was particularly strong for the overdiagnosed cancers attribute and the no screening alternative.

Latent class analysis revealed five distinct preference classes. The importance and significance of attributes varied substantially across all classes. Interestingly, the class with the highest membership probability (class 5) appeared to be neutral towards screening versus no screening (opt-out ASC was not significant) and false positive

results also appeared not to influence decisions. In contrast to the diagnostic setting, differences in screening preferences appeared to have stronger associations with observable sociodemographic factors. Level of education, previous screening behaviour, confidence in the ability to spot symptoms, ovarian cancer worry and self-perceived risk all influenced preferences (i.e. class membership probabilities).

The low incidence of ovarian cancer means the scope for benefits (i.e. lives saved) will always appear relatively small in comparison to other existing interventions. Nonetheless, scenario analysis suggested than uptake for screening, even where there is no proven benefit to screening was 77.5%. This figure is comparable to uptake rates of existing screening programmes such as breast and cervical screening, both of which achieve approximately 70% uptake (NHS Digital, 2019; NHS England, 2022). Increased likelihood of screening uptake was related to several sociodemographic characteristics, including increase anxiety about ovarian cancer, higher self-perceived risk, low confidence in the ability to recognise symptoms and regular participation in cervical screening.

Results provide guidance on the potential acceptability of future tests as and when they are developed. Variations in preferences reiterate the need for a more individualised approach to testing decisions and information sharing during discussions surrounding testing between doctors and their patients. Finally, since ovarian cancer remains ineffective, there remains an important role for ongoing awareness campaigns to enable early recognition of cancer symptoms and prompt help-seeking once symptoms arise.

12.2.3 Comparison of preferences in screening and diagnostic settings

Willingness to undergo testing

The lack of overlap between attributes included within the diagnostic and screening DCEs in this thesis means the ability to make comparisons of preferences at different stages along cancer pathway is limited. However, willingness to be tested was one clear difference between stated preferences in diagnostic and screening settings. Unsurprisingly, demand for testing was much higher in the diagnostic context where

respondents were asked to imagine experiencing symptoms associated with a distinct chance of ovarian cancer. This finding follows previous studies where people expressed high levels of willingness to test in the presence of any cancer risk, even where very low (Banks *et al.*, 2014). However, the difference in included attributes cannot be ignored and limits the ability to meaningfully compare DCE findings, including estimated demand. In particular, there was a comparatively heavy presence of potential harms in the screening setting (although this also reflects differences in test performance in the two settings).

Prioritisation of attributes

Best-worst scaling studies conducted during the attribute selection process provide a clearer opportunity to compare preferences and test priorities across settings. In total, the diagnostic and screening BWS shared ten attributes in common.

The ability to reduce ovarian cancer deaths was the most important factor in both contexts, followed by test sensitivity (i.e. chance of false negative results). In general, service delivery and process attributes appeared to be less prioritised in both settings in comparison to outcomes and test performance characteristics. However, this distinction was much clearer in the screening context. The prioritisation was less clear cut within the diagnostic setting, though this may be due to the higher number of attributes within the study and the efficiency of the experimental design. It also seems intuitive that attributes such as waiting times appeared more important when experiencing symptoms.

When considering service delivery ("process") factors there also appeared to be a prioritisation of experience-based factors (e.g. communication, waiting times, support and staff attitude) over convenience-based attributes (e.g. location, travel time, appointment times). This finding was particularly clear in diagnostic studies but was similarly reflected in the results of the screening study where attributes such as appointment scheduling or location were lower in importance than issues such as waiting times for tests or results.

Sociodemographic variations in preferences

For screening, preferences were more likely to be associated with observed sociodemographic factors such as age, previous screening behaviour, numeracy skills, anxiety around ovarian cancer. Despite the presence of preference heterogeneity in the diagnostic setting, the identification of distinct sociodemographic drivers was more limited. This seemingly intrinsic nature of decision-making further highlights the need for individualised care and understanding of patient preferences in diagnostic settings. Despite this finding, a common theme across both settings was an association between increased confidence in the ability to recognise ovarian cancer symptoms and a preference towards no testing. This is an important finding given the limited association between confidence in symptom knowledge and actual ability to recognise symptoms within both DCE surveys. This finding further reinforces the need for ongoing public education on symptoms to avoid late presentation.

Implications

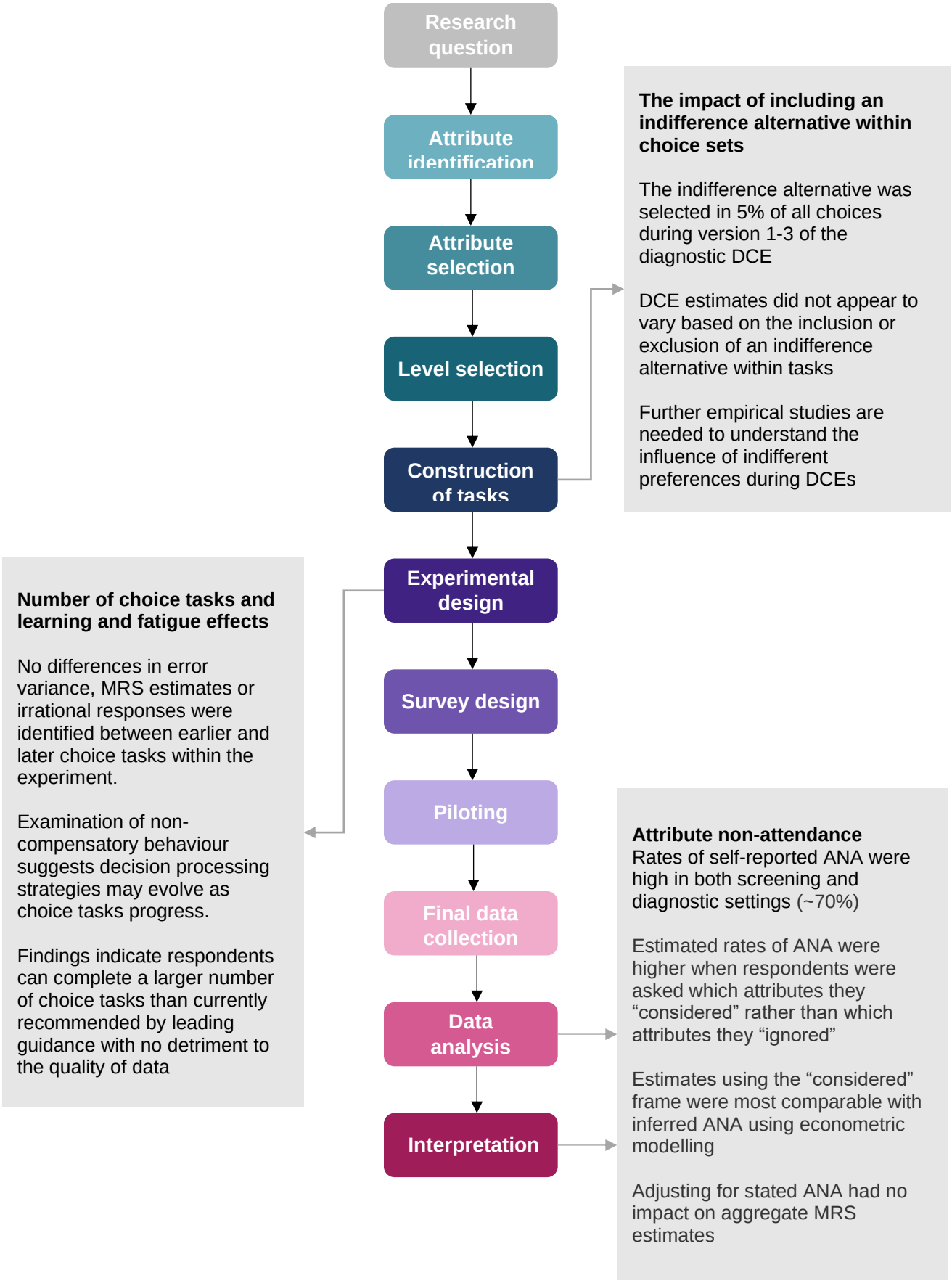
Overall, observed differences as well as the inability to make full comparisons between screening and diagnostic DCEs highlights the importance of future DCEs explicitly eliciting preferences in diagnostic settings. Preferences in screening settings should not automatically be assumed to be transferable. The prioritisation of test attributes appears to be different, even at the selection stage, meaning existing screening studies may exclude important and influential attributes to this setting (e.g. waiting times) increasing the likelihood of omitted variable bias. Furthermore, test performance characteristics vary between symptomatic and asymptomatic patients. This means attribute levels used in screening settings may not be relevant to diagnostic settings limiting the comparability of results, especially since the level range of attributes appears to influence preferences (Hall *et al.*, 2021).

12.3 Key findings and implications of methodological investigations

Despite the continued increase in published studies, the results of the systematic review showed the application of cancer DCE findings in policymaking contexts appears to be limited to date. Alongside the applied element of research, this thesis sought to address a few methodological uncertainties that may help to improve the

generalisability and practical applications of DCE results. A summary of the methodological extensions and where they lie along the DCE development process in shown in Figure 12.2.

Figure 12.2: Summary of the findings from the methodological extensions conducted throughout the thesis



12.3.1 The impact of including indifference alternatives within choice sets

Expression of indifferent preferences were frequent within the diagnostic testing DCE, accounting for about 5% of all responses. The omission of an indifference alternative forces respondents to artificially indicate a preference in instances when they value alternatives equally. Previous studies have shown this can lead to increased error variance and have a detrimental impact on model performance, although existing evidence is primarily reserved to the field of transportation (Bahamonde-Birke *et al.*, 2017; Cantillo *et al.*, 2010; Pan & Zuo, 2020).

Chapter 8 investigated the implications of including or excluding an indifference alternative in a healthcare context. Examination of rationality failures, stated attribute non-attendance, error variance across responses and willingness to wait estimates found no significant differences between DCEs with and without an indifference alternative. However, further empirical investigations in the field of healthcare are needed to draw an overall conclusion on the presence and implications of indifference towards healthcare decisions.

The results from the diagnostic DCE highlight the prominence of indifferent preferences towards diagnostic testing for ovarian cancer, with 28% of respondents choosing the indifference alternative on at least one occasion. Exclusion of an indifference alternative may introduce bias and noise into responses by forcing respondents to arbitrarily choose between alternatives. For this reason, leading guidance recommends allowing respondents to express indifference within choice tasks (Bridges *et al.*, 2011). However, failure to adequately model indifferent responses results in reduced efficiency in the experimental design which may have a detrimental impact identification and precision of model parameters (Johnson *et al.*, 2013; Louviere & Lancsar, 2009).

12.3.2 Number of choice tasks and learning and fatigue effects

Observations during the qualitative DCE piloting stage in Chapter 5 motivated an investigation of learning and fatigue effects throughout the diagnostic DCE survey. During think-aloud interviews some respondents exhibited learning behaviour as they progressed through the choice tasks by answering choice tasks faster over time and appearing to revaluate attribute meanings and importance between earlier and later

tasks. Alternatively, other respondents appeared to experience fatigue as tasks progressed. This methodological extension added to an established body of evidence with inconclusive findings and added a unique element by investigating the rationality of choices at different stages across the survey (Bech *et al.*, 2011; Hess *et al.*, 2012). Examination of non-compensatory behaviour indicated some individuals may experience evolving decision processing strategies as tasks progress, however, at an aggregate level results found little-to-no evidence of learning or fatigue effects across multiple dimensions: rationality of choices, model scale and error variance, and estimates of marginal rates of substitution between attributes. Instead, findings suggest that variations in scale between choice tasks were primarily due to differences in utility balance between alternatives within a choice tasks and choice confidence rather than the progressive number of questions completed. An observation that aligns with previous studies that find differences in error variance based on order or position effects of choice tasks within the experimental design (Campbell *et al.*, 2015; Day *et al.*, 2012).

Overall, failure to find evidence of either learning or fatigue effects according to the number of choice tasks completed indicate leading guidance may currently underestimate the number of choice tasks that can be completed by respondents without detriment to the quality of data (Bridges *et al.*, 2011).

12.3.3 Attribute non-attendance during the completion of choice tasks

The DCEs in this thesis explored the implications of self-reported attribute non-attendance (ANA) on model estimates and in particular, estimates of the marginal rates of substitution between attributes. Self-reported rates of ANA in both DCEs were high (~70% of all respondents indicated they did not attend least one attribute), however, comparable to findings in previous publications (Hole *et al.*, 2013; Scarpa *et al.*, 2013). Stated ANA methods have been critiqued due to the potential for variations due to procedural factors (e.g. How and when to ask the question? How is the question interpreted? How well can respondents recall and evaluate their behaviour?) (Scarpa *et al.*, 2013). In response, this thesis added a small but to my best knowledge unique contribution to the interdisciplinary body of research on ANA by considering the impact of question framing on DCE estimates. Results indicate that question framing leads to

differences in reported rates of ANA. Although adjustments for non-attendance regardless of elicitation method had little-to-no impact on WTA estimates. There is a rich and conflicting body of research relating to ANA in DCEs, however, engagement with this research in the field of health economics remains fairly limited and peripheral to applied studies (Arora *et al.*, 2022; Erdem *et al.*, 2015; Jonker *et al.*, 2018). Results of this thesis and previous studies (Arora *et al.*, 2022; Hess *et al.*, 2013) suggest considering ANA can be useful in understanding preference heterogeneity and can be particularly valuable when considering policies relating to increasing uptake. As a first step, greater transparency in rates of ANA appears necessary. Pearce *et al.* (2020) found that whilst many authors of DCEs (31%) reported including debriefing questions relating to ANA, very few actually report such data. This may be due to a lack of precedent on ANA in health-based research. Questions such as acceptable levels of ANA, how to measure ANA and how to accommodate non-attendance into modelling remain unresolved. This appears to be area in need of future development.

12.4 Learnings/reflections on eliciting preferences towards diagnostic testing

An overarching objective of the thesis was to provide an example of how discrete choice experiments can be used to elicit preferences towards diagnostic testing for cancer- an area where medical decision-making remains overly paternalistic, despite the general trend towards shared decision-making and increased patient involvement in broader healthcare settings. The application to ovarian cancer demonstrated the willingness and ability of patients to form and express preferences in this context. Additionally, the application towards ovarian cancer provided ways in which preferences can be incorporated into policymaking and clinical care with the purpose of improving diagnostic outcomes and patient satisfaction. The findings of the thesis and the process of developing DCEs in this context identified several issues future researchers may wish to consider when extending the methodology into the field of cancer diagnosis. This section is not exhaustive and instead focuses on some key learnings that have been discussed to a lesser degree through the chapter discussions.

12.4.1 Formulating a relevant research question

The systematic review in Chapter 4 highlights the current underutilisation of findings from cancer DCEs within decision-making settings despite the continual increase in DCE publications, most of which have policy-based research questions as the motivation for the study. Alongside methodological uncertainties, Chapter 4 also highlighted disparities between research questions, chosen attributes and reported outcomes as a potential source of moderator of application to policymaking.

This thesis aimed to provide an exemplar of how DCEs can be applied to diagnostic setting for the purpose improving diagnostic outcomes and satisfaction with care and shared decision-making in diagnostic settings. The diagnostic DCE in Chapter 8 provided insights into priorities around ovarian cancer which can be useful to future guidance revisions. However, the explorative and adaptive nature of PhD research meant final results from the diagnostic DCE were somewhat more removed from the policy context relating to ovarian cancer than first hoped. For example, direct comparisons of alternative test pathways were not possible (e.g. CA125 vs CA125 + TVUS, or TVUS only).

The formulation of the research question will influence the design and implementation of a preference elicitation study and ultimately determine value of results. As such, the importance of specifying a clear and relevant research question, motivated by clinical practice should not be overlooked. Whilst PPIE input was sought during the development of the DCEs within this thesis, engagement with clinical experts during the conceptualisation process may have helped to ensure DCEs are better matched and relevant to clinical practice and policy related issues. Greater collaboration with clinical and policy-making experts should be an important consideration for future researchers.

12.4.2 Attribute selection- combining methods, importance of input from target audience

The results of this thesis demonstrate the value of combining multiple methods in order to gain different perspectives. Reliance on a single method (e.g. systematic review)

during attribute development may lead to inconclusive or suboptimal evidence on the which attributes to include. However, combining evidence from multiple sources was a key challenge throughout the thesis and particularly within the diagnostic DCE, due the need to reconcile contrasting evidence between in-depth qualitative workshops with a small sample and the more high-level prioritisation BWS survey with a larger sample of the target population. Over-weighting of attributes prioritised during the workshops (e.g. communication) may have been a contributing factor to the high levels of stated attribute non-attendance observed in the final DCE study.

There is an increasing number of publications offering guidance on attribute development (Helter & Boehler, 2016). Notably, since the completion of attribute selection within this thesis, Webb *et al.* (2021) published a framework for the inclusion of BWS during attribute development. Importantly guidance is typically based on a single case-study and comparisons of alternative selection methods and their outcomes are limited. Existing examples comparing different methods and/or different combinations of methods demonstrate alternative methods result in differences in the final attributes and/or the number of attributes selected (Helter & Boehler, 2016; Timmis, 2020). To my knowledge no studies have gone beyond this point and considered how attributes generated using different methods affect DCE estimates and in particular attribute non-attendance (i.e. relevance of attributes to the final sample).

There is a need for pragmatic guidance on merging evidence from multiple sources and finalising attributes during attribute selection—a stage which unavoidably relies on the subjectivity and judgement of researchers introducing the potential for bias. Reflections from the research during this thesis provide a starting point but further research is needed.

Advice for future studies in cancer testing when determining methods for attribute selection:

- **Begin with a clear research question:** What do you fundamentally want to know? Are there any attributes that are essential to achieve the research objectives

regardless of the outcomes of attribute selection methods (e.g. price attribute if WTP calculations are needed)?

- **Identify any existing evidence relating to your study and any clear evidence gaps:** what existing evidence in the area exists? Is there already substantial qualitative work or previous BWS or DCE studies? What would be most beneficial given any research constraints (e.g. time, financial, recruitment)?

- **Have clear inclusion criteria and development pathway from the outset:** Limit *ad hoc* decisions by outlining a framework for finalising attribute selection. This will help to mitigate the influence of researcher bias and allow for a systematic approach during stages of the process where greater subjectivity is required.

- **Involve PPIE throughout the selection process:** The inclusion of PPIE representatives will allows for the target population perspective to be consistently present throughout the development process. This engagement may be particularly valuable during stages involving researcher judgments (e.g. final selection).

- **Consider the order of attribute development methods when using a multi-method approach:** Alternative methods may be more useful for particular stages of attribute development. Qualitative research is typically more exploratory in nature and may be useful to identify a longlist of potential attributes whereas best-worst scaling is more reductive in nature, helping to narrow down candidate attributes. For this reason, upon reflection it may have been preferable to reverse the order of methods utilised within Chapter 5. Where resources allow an iterative approach between qualitative, quantitative and researcher-based stages is likely to be best.

- **Consider the skills of the research team:** Poor research practice due to inexperience can affect the validity of results. It is therefore important to consider which methods most suit the skills of the researchers and to access expert guidance or training where necessary. Arguably, a limitation of attribute selection within this thesis was that all qualitative aspects were performed by typically quantitative researchers (although training and expert guidance were obtained).

12.4.3 Importance and influence of level assignment

Level assignment is a crucial step of attribute development that influences the importance of attributes within the final DCE and the application of results to broader contexts. By contrast, level assignment receives comparatively less attention in than attribute development as highlighted by the systematic review in Chapter 4. In hindsight there are two key avenues within the thesis research where level selection may have influenced the outcomes of the study and contributed to ANA.

The first avenue relates to the inclusion of attribute levels during attribute development research involving the target audience. Following convention, attribute levels were not described to participants when choosing “most” and “least” important attributes during BWS tasks (i.e. Case 1 BWS). This approach limits the cognitive burden of tasks and is therefore particularly useful in studies with many attributes as seen in the diagnostic study (n=25). During the diagnostic attribute development, potential attribute levels were introduced to workshop participants during the follow up questionnaire. This decision was based on a general-to-specific approach which introduced more complexity as engagement and understanding with the study aims and context increased. Questionnaire results demonstrated that the introduction of attribute levels influenced the importance and acceptability of potential attributes and was a crucial stage of the selection process. This finding highlights the importance of including attribute levels during the development stage. No introduction at all may lead to inclusion to irrelevant attributes once the contexts/constraints of the decision context are introduced in the final DCE. For example, when considering colorectal cancer testing, modality may be important to the target population without context but if both tests involve faecal sampling, they are likely to be similar in terms of acceptability. As opposed to comparison of a faecal test versus colonoscopy. Case 2 BWS, wherein participants indicate the importance of attribute levels, may be preferable where there are fewer attributes under consideration/the complexity of attributes is limited. This is particularly important if this is the primary method for attribute selection.

The second avenue relates to the combination of linguistically and numerically expressed levels within a single DCE. This study follows previous literature in finding

self-reported ANA was much higher for attributes with linguistically expressed levels (e.g. communication and identifiable conditions) compared to numeric levels (i.e. time to diagnosis, accuracy). Evidence from the field of marketing demonstrates numeric levels are considered to be more “concrete”, meaning information is specific, tangible and presented in an easily processible form (Viswanathan & Childers, 1996). Linguistic attributes are typically more “abstract”; the meaning is vaguer and evaluation requires further processing. Research shows concrete attributes are easier to understand, process and directly compare (Stone & Schkade, 1991). Abstract attributes require greater cognitive effort leading to selective attribute processing and increased ANA (Horsky *et al.*, 2004; Huber, 1980; Jiang & Punj, 2010; Nisbett & Ross, 1980). Mixing qualitatively and numerically described attributes may often be unavoidable but attention should be paid to the clarity and subjectivity of language compared to numerically described attributes. In the diagnostic DCE, attribute levels for “communication” were very subjective, whereby respondents were explicitly asked to assign their own definition to the levels. If replicated, this is an aspect of the DCE which could be further refined.

12.4.4 Establishing the choice context: balancing difficulty and engagement

Balancing the level of information to ensure respondents are appropriately motivated but not overwhelmed is a clear challenge when considering preferences towards cancer testing given the complexity and nuance of the topic. Participants may be required to formulate preferences towards issues they had never considered, using unfamiliar language or requiring consideration of risky scenarios.

During the diagnostic DCE, 38% of respondents reported finding the tasks difficult or very difficult. This figure rose to 47% in the screening DCE where attributes all involved risk. Increased perceived difficulty may indicate genuine engagement with the task, particularly when using an efficient experimental design since alternatives are paired based on utility balance. However, highly difficult tasks can also affect a respondent’s ability or willingness to meaningfully engage with the tasks.

The challenge of obtaining authentic engagement is exacerbated in online settings. Attention check questions were utilised within the surveys of this thesis to encourage

participants to thoroughly read and consider questions. A caveat of this method is online participants are typically experienced in spotting these attention-type questions and such questions do not allow understanding of the choice task to be gauged. Furthermore, engagement does not necessarily ensure understanding.

The use of alternative risk displays was discussed in depth throughout previous chapters as a means to increase participant understanding. There are other ways in which the communication of information may have been approached differently and to provide insights for future studies.

Training materials

Good practice guidance emphasises the need for an introductory section explaining the context of the DCE and the content of the survey including attributes and levels prior to the presentation of the DCE choice tasks. These training materials require a balance between providing enough information to motivate and facilitate genuine engagement with the choice tasks versus overwhelming respondents with excessive information.

A recent area of development is the use of interactive tools or videos rather than written information to introduce the information relevant to the DCE to improve engagement and understanding. The effectiveness of alternative methods is an emerging area of research. Results to date demonstrate interactive tools improve the choice consistency (Vass *et al.*, 2020) but effects on preferences appear limited (Lim *et al.*, 2020; Vass *et al.*, 2020) although Bywall *et al.* (2021) found differences in attribute importance across different educational tools. Further research is needed; however, interactive tools may help to improve engagement particularly with the continued move towards online survey administration.

Comprehension questions

Vass and Payne (2017) suggest the inclusion of comprehension questions may help to assess respondent understanding of training materials. Currently, empirical evidence on the effects of comprehension questions is currently lacking. Inclusion of such questions was considered during the early stages of DCE development within

this thesis (e.g. “True or false: Some tests used to identify ovarian cancer might also unintentionally discover an unrelated condition, known as an incidental finding?”). No comprehension questions were included within the final survey due to negative feedback from PPIE and peer-reviewers during the ethics application process.

12.5 Limitations

This thesis provides novelty in being an early example of the importance of evaluating preferences in diagnostic settings and the only application of stated preference techniques to ovarian cancer testing (to my best knowledge). A key strength of this research is the rigorous and transparent approach to DCE methods following best practice guidance.

Limitations of the empirical aspects of this thesis are discussed in each chapter. This section provides a summary of some of the overarching limitations of the methods undertaken within this PhD.

Firstly, as with all stated preference studies, the hypothetical nature of the DCEs means establishing the external validity of findings is a challenge. The choice to focus on the preferences of women from the general public rather than a patient populationalso means that most respondents were required to imagine scenarios which they had no prior experience (although it is likely all respondents had received some form of medical testing throughout their life).

Next, the representativeness of sample was a challenge highlighted throughout the empirical chapters of this thesis. Steps were taken to improve the representativeness of the sample in terms of ethnicity in later studies but the choice of recruitment combined with the limits to the sample (i.e. women over 40) meant a fully representative sample was not possible. This may have an impact on the generalisability of results to the broader population, although results suggested differences in preferences did not generally appear to be associated with typical sociodemographic factors (e.g. ethnicity, age, employment status etc). Similarly, the online-only approach to recruitment introduces the possibility of selection bias, as only

those who had access to a computer and were computer literate were able to take part.

Finally, it is important to recognise the limitations of findings for the purpose of improving diagnostic outcomes through policy recommendations. Preferences play an important role in satisfaction and adherence to care plans. However, during the course of the thesis it became apparent that delays in diagnosis are largely driven by knowledge gaps in symptom awareness from both a patient and provider perspective. For context, Lyratzopoulos *et al.* (2015) found the mean length of time from symptom onset to first consultation (39 days) was around twice as long as the time from first consultation to specialist referral (21 days) for ovarian cancer. This highlights the importance of continued public health interventions aimed at shortening patient intervals.

12.6 Future Research

12.6.1 Optimising ovarian cancer diagnosis

The importance placed on test accuracy and timeliness in diagnostic settings suggests current testing recommendations may not be optimised to meet women's preferences, particularly for patients under 50 years old where CA125 testing significantly reduces in accuracy (Funston *et al.*, 2020a). However, whilst the results of this thesis provide a useful reference of women's preferences, there are several gaps in the current evidence base which need to be addressed in order to understand how preferences may be implemented into clinical practice and before current guidelines to be reconsidered.

Key areas for future research to address these uncertainties and facilitate the application of preferences towards ovarian cancer diagnosis with the aim of improving diagnostic outcomes and patient satisfaction are:

i. Understanding the diagnostic performance characteristics of TVUS for the diagnosis ovarian cancer in primary care
ii. Understanding the efficacy of alternative testing pathways e.g. combined testing versus sequential testing versus single modality testing

iii. Investigations of the cost-effectiveness and feasibility of alternative tests and combinations of testing from a capacity aspect is also needed.

This thesis also demonstrated demand for testing even at very low risks of cancer, suggesting an expansion of CA125 testing in primary care (i.e. increasing the symptoms that trigger primary care testing) would be welcomed; although, again it is important to consider the cost effectiveness and capacity impacts of such a change.

12.6.2 Preferences towards cancer testing

The research presented in this thesis indicates that people are willing and able to express preferences towards diagnostic testing and wish to play a greater role during primary care consultations.

Given the current lack of evidence relating to preferences for diagnostic testing further DCEs studies eliciting preferences towards diagnostic testing to facilitate shared decision-making and increase patient input, particularly for cancer sites where there are new emerging diagnostic modalities or where evidence or guidance reviews are expected. The best-worst scaling studies presented in Chapters 5 and 9 as part of attribute selection provide useful starting points for future other studies in the field of cancer testing as the use of Case 1 BWS results are likely to be generalisable to cancer testing as site-specific levels were not included.

As mentioned throughout the thesis, the research was developed in collaboration with cisgender women and is likely to have collected responses from this population too. Future preference-based studies should aim to measure the preferences of transgender and non-binary people at risk of ovarian cancer. This research should be developed in collaboration with members of this population rather than repeating experiments such as the ones in this thesis in a new population.

Finally, given the collaborative nature of shared decision-making, future studies investigating doctors' preferences would also be of value. Based on the current dominant role of GPs in diagnostic decisions, studies may investigate doctors' own preferences towards diagnostic testing and/or doctors' perceptions of patient preferences. Existing research has demonstrated that GP recommendations are

valued by patients and influence decision-making (Peterson *et al.*, 2016) therefore, future studies may also focus on understanding collaborative decision-making by applying multi-agent DCE methods such as Interactive Agency Choice Experiments developed in other fields such as transport (Brewer & Hensher, 2000; Hensher *et al.*, 2008; Rose & Hensher, 2004).

12.6.3 Discrete choice experiments methods

Several avenues for further research in the application of DCEs to cancer testing (and healthcare more broadly) were suggested throughout the thesis, a few of which are reiterated here.

Firstly, as outlined in section 12.4.2, further research into the attribute selection process is a potential area for future research. In particular, comparisons of alternative methodological approaches with the ultimate aim of developing a framework for combining multiple streams of evidence for future authors. It is hoped that future research will be helpful in understanding and addressing discrepancies between attribute development and final DCE findings as seen in the diagnostic DCE conducted in this thesis as well as previous studies (Timmis, 2020).

Secondly, the high levels of self-reported difficulty within the DCEs in this thesis indicate further research is needed to explore alternative methods for communicating contextual information and attributes within choice tasks. This is particularly important where there is a heavy presence of risk-based attributes which add to overall complexity of tasks as seen in Chapter 10. There exists a small body of evidence comparing alternative approaches to risk communication within DCEs (Vass *et al.*, 2018a) however, applications of best evidence from the field of risk communication remain limited (Harrison *et al.*, 2014). This area of research is also highlighted as a methodological priority in a recent publication by a panel of preference elicitation experts (Smith *et al.*, 2021).

Finally, the development of transparent and readily available approaches allowing the incorporation of indifference preferences into choice models is a further avenue for future research. In particular, the extension and adaptation of existing to accommodate

the presence of no choice alternatives as seen in Chapter 8 of this thesis (Bahamonde-Birke *et al.*, 2017; Cantillo *et al.*, 2010).

12.7 Conclusion

There is an ever-increasing body of literature eliciting preferences towards cancer testing with the purpose of increasing patient satisfaction, uptake and treatment outcomes. To date, studies almost exclusively focus on preferences towards screening testing in asymptomatic populations. The research in this thesis provides an example of how discrete choice experiment methods can be applied to diagnostic settings, where medical decision-making remains largely paternalistic despite the increasing emphasis on shared decision-making between doctors and patients within the NHS. Further novelty was introduced through the focus on preferences towards ovarian cancer, a site where no quantitative preference studies were identified and several uncertainties surrounding diagnostic testing remain.

To ensure rigorous research methods, an extensive systematic review of DCEs on cancer testing provided the basis on discrete choice development which also combined multiple evidence streams including best-worst scaling surveys and online workshops with the general public. This thesis provides evidence that discrete choice methods can be successfully applied to diagnostic settings and respondents are willing and capable of expressing rational preferences towards hypothetical diagnostic scenarios. The results of this thesis provide useful information for policymakers if and when clinical diagnostic guidance is reconsidered and provides useful insights for researchers aiming to undertake future preference elicitation studies relating to diagnostic testing.

Lack of symptom awareness from both a patient and provider perspective was highlighted as a key barrier to earlier diagnosis throughout the thesis. The development of an ovarian cancer screening program is widely considered to be a solution to the high rates of late diagnoses currently experienced. Therefore, the extension into ovarian cancer screening also provides evidence on the acceptability and demand for potential screening programmes which continues to be a prevalent area of ovarian cancer research.

13 Bibliography

Adamowicz, W., Boxall, P., Williams, M., & Louviere, J. (1998). Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. *American Journal of Agricultural Economics, 80*(1), 64-75.

Adamowicz, W., Louviere, J., & Williams, M. (1994). Combining revealed and stated preference methods for valuing environmental amenities. *Journal of environmental economics and management, 26*(3), 271-292.

Aizaki, H. (2014). support. BWS: Basic functions for supporting an implementation of best–worst scaling, R package version 0.1-1. In: URL http://CRAN. R-project. org/package= support. BWS.

Akintomide, A. O., & Obasi, U. O. (2019). Intimate patient examinations: The awareness, acceptance and practice preference of transvaginal ultrasound scan among women in a South-southern State of Nigeria. *Journal of Family Medicine and Primary Care, 8*(1), 109.

Akl, E. A., Oxman, A. D., Herrin, J., Vist, G. E., Terrenato, I., Sperati, F., . . . Schünemann, H. (2011). Using alternative statistical formats for presenting risks and risk reductions. *Cochrane Database of Systematic Reviews*(3).

Alemu, M. H., Mørkbak, M. R., Olsen, S. B., & Jensen, C. L. (2013). Attending to the reasons for attribute non-attendance in choice experiments. *Environmental and resource economics, 54*(3), 333-359.

Ali, S., & Ronaldson, S. (2012). Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. *British Medical Bulletin, 103*(1), 21-44. doi:10.1093/bmb/lds020

Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. (2006). Design features of graphs in health risk communication: a systematic review. *Journal of the American Medical Informatics Association, 13*(6), 608-618.

Andermann, A., Blancquaert, I., Beauchamp, S., & Déry, V. (2008). Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bulletin of the World Health Organization, 86*, 317-319.

Arana, J. E., Leon, C. J., & Quevedo, J. L. (2006). The effect of medical experience on the economic evaluation of health policies. A discrete choice experiment. *Social Science and Medicine, 63*(2), 512-524. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1016/j.socscimed.2005.12.008

Armstrong, N., Herbert, G., Aveling, E. L., Dixon-Woods, M., & Martin, G. (2013). Optimizing patient involvement in quality improvement. *Health Expect, 16*(3), e36-47. doi:10.1111/hex.12039

Arora, N., Quaife, M., Hanson, K., Lagarde, M., Woldesenbet, D., Seifu, A., & Crastes dit Sourd, R. (2022). Discrete choice analysis of health worker job preferences in Ethiopia: Separating attribute non-attendance from taste heterogeneity. *Health economics, 31*(5), 806-819.

Atalabi, O. M., Morhason-Bello, I. O., Adekanmi, A. J., Marinho, A. O., Adedokun, B. O., Kalejaiye, A. O., . . . Gbadamosi, S. A. (2012). Transvaginal ultrasonography: a survey of the acceptability and its predictors among a native African women population. *International Journal of Women's Health, 4*, 1.

Bahamonde-Birke, F. J., Navarro, I., & de Dios Ortúzar, J. (2017). If you choose not to decide, you still have made a choice. *Journal of choice modelling, 22*, 13-23.

Baldwin, L.-M., Trivers, K. F., Matthews, B., Andrilla, C. H. A., Miller, J. W., Berry, D. L., . . . Goff, B. A. (2012). Vignette-based study of ovarian cancer screening: do US physicians report adhering to evidence-based recommendations? *Annals of internal medicine, 156*(3), 182-194.

Banks, J., Hollinghurst, S., Bigwood, L., Peters, T. J., Walter, F. M., & Hamilton, W. (2014). Preferences for cancer investigation: a vignette-based study of primary-care attendees. *The Lancet Oncology, 15*(2), 232-240.

Bansal, N., Bhopal, R., Steiner, M., & Brewster, D. H. (2012). Major ethnic group differences in breast cancer screening uptake in Scotland are not extinguished by adjustment for indices of geographical residence, area deprivation, long-term illness and education. *British journal of cancer, 106*(8), 1361-1366.

Bansback, N., Brazier, J., Tsuchiya, A., & Anis, A. (2012). Using a discrete choice experiment to estimate health state utility values. *Journal of health economics, 31*(1), 306-318.

Bansback, N., Harrison, M., Sadatsafavi, M., Stiggelbout, A., & Whitehurst, D. G. (2016). Attitude to health risk in the Canadian population: a cross-sectional survey. *CMAJ open, 4*(2), E284.

Barnett, K., Mercer, S. W., Norbury, M., Watt, G., Wyke, S., & Guthrie, B. (2012). Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *The Lancet, 380*(9836), 37-43. doi:10.1016/S0140-6736(12)60240-2
Bateman, I. J., Burgess, D., Hutchinson, W. G., & Matthews, D. I. (2008). Learning design contingent valuation (LDCV): NOAA guidelines, preference learning and coherent arbitrariness. *Journal of environmental economics and management, 55*(2), 127-141.
Bech, M., Kjaer, T., & Lauridsen, J. (2011). Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. *Health economics, 20*(3), 273-286.
Bechel, D. L., Myers, W. A., & Smith, D. G. (2000). Does patient-centered care pay off? *Jt Comm J Qual Improv, 26*(7), 400-409. doi:10.1016/s1070-3241(00)26033-1
Bennett, K. F., Waller, J., Chorley, A. J., Ferrer, R. A., Haddrell, J. B., & Marlow, L. A. (2018). Barriers to cervical screening and interest in self-sampling among women who actively decline screening. *Journal of Medical Screening, 25*(4), 211-217.
Benning, T. M., Dellaert, B. G. C., Dirksen, C. D., & Severens, J. L. (2014a). Preferences for potential innovations in non-invasive colorectal cancer screening: A labeled discrete choice experiment for a Dutch screening campaign. *Acta Oncologica, 53*(7), 898-908. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.3109/0284186X.2013.877159
Benning, T. M., Dellaert, B. G. C., Severens, J. L., & Dirksen, C. D. (2014b). The effect of presenting information about invasive follow-up testing on individuals' noninvasive colorectal cancer screening participation decision: Results from a discrete choice experiment. *Value in Health, 17*(5), 578-587. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1016/j.jval.2014.04.007
Benzeval, M., Davillas, A., Kumari, M., & Lynn, P. (2014). Understanding society: the UK household longitudinal study biomarker user guide and glossary. *Institute for Social and Economic Research, University of Essex*.
Berchi, C., Dupuis, J.-M., & Launoy, G. (2006). The Reasons of General Practitioners for Promoting Colorectal Cancer Mass Screening in France. *European Journal of Health Economics, 7*(2), 91-98. doi:https://link.springer.com/journal/volumesAndIssues/10198
Bien, D. R., Danner, M., Vennedey, V., Civello, D., Evers, S. M., & Hiligsmann, M. (2017). Patients' Preferences for Outcome, Process and Cost Attributes in Cancer Treatment: A Systematic Review of Discrete Choice Experiments. *Patient, 10*(5), 553-565. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1007/s40271-017-0235-y
Bilger, M., Özdemir, S., & Finkelstein, E. A. (2020). Demand for Cancer Screening Services: Results From Randomized Controlled Discrete Choice Experiments. *Value in Health, 23*(9), 1246-1255.
Birch, S., Melnikow, J., & Kuppermann, M. (2003). Conservative versus aggressive follow up of mildly abnormal Pap smears: testing for process utility. *Health Economics, 12*(10), 879-884.
Bliemer, M. C., Rose, J. M., & Hensher, D. A. (2009). Efficient stated choice experiments for estimating nested logit models. *Transportation Research Part B: Methodological, 43*(1), 19-35.
Bliemer, M. C., Rose, J. M., & Hess, S. (2008). Approximation of Bayesian efficiency in experimental choice designs. *Journal of choice modelling, 1*(1), 98-126.
Board, S. (2009). Preferences and utility. *UCLA, Oct*.
Boland, A., Cherry, G., & Dickson, R. (2017). *Doing a systematic review: A student's guide*: Sage.
Boone, D., Mallett, S., Zhu, S., Yao, G. L., Bell, N., Ghanouni, A., . . . Halligan, S. (2013). Patients' & healthcare professionals' values regarding true- & false-positive diagnosis when colorectal cancer screening by CT colonography: Discrete choice experiment. *PLoS ONE, 8*(12), e80767. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1371/journal.pone.0080767
Boyle, K. J. (2003). Introduction to revealed preference methods. In *A primer on nonmarket valuation* (pp. 259-267): Springer.
Boyle, K. J., Holmes, T. P., Teisl, M. F., & Roe, B. (2001). A comparison of conjoint analysis response formats. *American journal of agricultural economics, 83*(2), 441-454.
Bradley, M., & Daly, A. (1994). Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. *Transportation, 21*(2), 167-184. doi:10.1007/BF01098791
Brazier, J., Akehurst, R., Brennan, A., Dolan, P., Claxton, K., McCabe, C., . . . Tsuchyia, A. (2005). Should patients have a greater role in valuing health states? *Applied health economics and health policy, 4*(4), 201-208.
Brazier, J., Ratcliffe, J., Saloman, J., & Tsuchiya, A. (2017). *Measuring and valuing health benefits for economic evaluation*: OXFORD university press.
Brenner, A., Howard, K., Lewis, C., Sheridan, S., Crutchfield, T., Hawley, S., . . . Pignone, M. (2014). Comparing 3 values clarification methods for colorectal cancer screening decision-making: A

randomized trial in the US and Australia. *Journal of General Internal Medicine, 29*(3), 507-513. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1007/s11606-013-2701-0
Brewer, A. M., & Hensher, D. A. (2000). Distributed work and travel behaviour: the dynamics of interactive agency choices between employers and employees. *Transportation, 27*(1), 117-148.
Bridges, J. (2003). Stated preference methods in health care evaluation: an emerging methodological paradigm in health economics. *Applied health economics and health policy, 2*(4), 213-224.
Bridges, J. F., Hauber, A. B., Marshall, D., Lloyd, A., Prosser, L. A., Regier, D. A., . . . Mauskopf, J. (2011). Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value in health, 14*(4), 403-413.
Brouwer, W. B., Culyer, A. J., van Exel, N. J. A., & Rutten, F. F. (2008). Welfarism vs. extra-welfarism. *Journal of health economics, 27*(2), 325-338.
Brouwer, W. B., van Exel, N. J. A., van den Berg, B., van den Bos, G. A., & Koopmanschap, M. A. (2005). Process utility from providing informal care: the benefit of caring. *Health policy, 74*(1), 85-99.
Buchanan, J., Hill, M., Vass
, C. M., Hammond, J., Riedijk, S., Klapwijk, J. E., . . . Hui, L. (2022). Factor's that impact on women's decision-making around prenatal genomic tests: An international discrete choice survey. *Prenatal Diagnosis*.
Buys, S. S., Partridge, E., Black, A., Johnson, C. C., Lamerato, L., Isaacs, C., . . . Kessel, B. (2011). Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening randomized controlled trial. *Jama, 305*(22), 2295-2303.
Bywall, K. S., Veldwijk, J., Hansson, M. G., Baecklund, E., Raza, K., Falahee, M., & Kihlbom, U. (2021). Does being exposed to an educational tool influence patient preferences? The influence of an educational tool on patient preferences assessed by a discrete choice experiment. *Patient Education and Counseling, 104*(10), 2577-2585. doi:https://doi.org/10.1016/j.pec.2021.03.013
Campbell, D., Boeri, M., Doherty, E., & Hutchinson, W. G. (2015). Learning, fatigue and preference formation in discrete choice experiments. *Journal of Economic Behavior & Organization, 119*, 345-363.
Campbell, D., & Erdem, S. (2019). Including opt-out options in discrete choice experiments: issues to consider. *The Patient-Patient-Centered Outcomes Research, 12*(1), 1-14.
Campbell, D., Hutchinson, W. G., & Scarpa, R. (2008). Incorporating discontinuous preferences into the analysis of discrete choice experiments. *Environmental and resource economics, 41*(3), 401-417.
Cancer Research UK. (2019). Ovarian Cancer Statistics Retrieved from https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer#heading-Two
Cancer Research UK. (2022). Ultrasound scan for ovarian cancer. Retrieved from https://www.cancerresearchuk.org/about-cancer/ovarian-cancer/getting-diagnosed/tests/ultrasound
Cantillo, V., Amaya, J., & Ortúzar, J. d. D. (2010). Thresholds and indifference in stated choice surveys. *Transportation Research Part B: Methodological, 44*(6), 753-763.
Carlsson, F., Kataria, M., & Lampi, E. (2010). Dealing with ignored attributes in choice experiments on valuation of Sweden's environmental quality objectives. *Environmental and resource economics, 47*(1), 65-89.
Carter, J. L., Coletti, R. J., & Harris, R. P. (2015). Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. *Bmj, 350*.
Caussade, S., de Dios Ortúzar, J., Rizzi, L. I., & Hensher, D. A. (2005). Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B: Methodological, 39*(7), 621-640.
Chamot, E., Mulambia, C., Kapambwe, S., Shrestha, S., Parham, G. P., Macwan'gi, M., & Mwanahamuntu, M. H. (2015). Preference for human papillomavirus-based cervical cancer screening: Results of a choice-based conjoint study in Zambia. *Journal of Lower Genital Tract Disease, 19*(2), 119-123. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1097/LGT.0000000000000081
Charness, G., Garcia, T., Offerman, T., & Villeval, M. C. (2020). Do measures of risk attitude in the laboratory predict behavior under risk in and outside of the laboratory? *Journal of Risk and Uncertainty, 60*(2), 99-123.

Charvin, M., Launoy, G., & Berchi, C. (2020). The effect of information on prostate cancer screening decision process: a discrete choice experiment. *BMC Health Services Research, 20*, 1-10.
Cheraghi-Sohi, S., Bower, P., Mead, N., McDonald, R., Whalley, D., & Roland, M. (2007). Making sense of patient priorities: applying discrete choice methods in primary care using 'think aloud' technique. *Fam Pract, 24*(3), 276-282. doi:10.1093/fampra/cmm007
Cheung, K. L., Wijnen, B. F., Hollin, I. L., Janssen, E. M., Bridges, J. F., Evers, S. M., & Hiligsmann, M. (2016). Using best–worst scaling to investigate preferences in health care. *Pharmacoeconomics, 34*(12), 1195-1209.
Coast, J. (2009). Maximisation in extra-welfarism: a critique of the current position in health economics. *Social science & medicine, 69*(5), 786-792.
Coast, J., Al-Janabi, H., Sutton, E. J., Horrocks, S. A., Vosper, A. J., Swancutt, D. R., & Flynn, T. N. (2012). Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. *Health economics, 21*(6), 730-741.
Coast, J., Flynn, T., Sutton, E., Al-Janabi, H., Vosper, J., Lavender, S., . . . Peters, T. (2008). Investigating Choice Experiments for Preferences of Older People (ICEPOP): evaluative spaces in health economics. *Journal of health services research & policy, 13*(3_suppl), 31-37.
Coast, J., & Horrocks, S. (2007). Developing attributes and levels for discrete choice experiments using qualitative methods. *Journal of health services research & policy, 12*(1), 25-30.
Collins, A. (2012). Attribute nonattendance in discrete choice models: measurement of bias, and a model for the inference of both nonattendance and taste heterogeneity.
Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). Mathematical psychology: An elementary introduction.
Corso, P. S., Hammitt, J. K., & Graham, J. D. (2001). Valuing mortality-risk reduction: using visual aids to improve the validity of contingent valuation. *Journal of Risk and Uncertainty, 23*(2), 165-184.
Cowan Bennett, C., & Richards, D. (2000). Patient acceptance of endovaginal ultrasound. *Ultrasound in Obstetrics and Gynecology, 15*(1), 52-55.
Cramer, D. W., Bast, R. C., Berg, C. D., Diamandis, E. P., Godwin, A. K., Hartge, P., . . . Mor, G. (2011). Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer screening trial specimens. *Cancer prevention research, 4*(3), 365-374.
Croissant, Y. (2012). Estimation of multinomial logit models in R: The mlogit Packages. *R package version 0.2-2. URL: http://cran. r-project. org/web/packages/mlogit/vignettes/mlogit. pdf*.
Cronin, R. M., Mayo-Gamble, T. L., Stimpson, S.-J., Badawy, S. M., Crosby, L. E., Byrd, J., . . . Murry, V. M. (2018). Adapting medical guidelines to be patient-centered using a patient-driven process for individuals with sickle cell disease and their caregivers. *BMC hematology, 18*(1), 1-13.
Culyer, A. J. (1989). The normative economics of health care finance and provision. *Oxford review of economic policy, 5*(1), 34-58.
Davey, H. M., Lim, J., Butow, P. N., Barratt, A. L., & Redman, S. (2004). Women's preferences for and views on decision-making for diagnostic tests. *Social science & medicine, 58*(9), 1699-1707. doi:https://doi.org/10.1016/S0277-9536(03)00339-3
Day, B., Bateman, I. J., Carson, R. T., Dupont, D., Louviere, J. J., Morimoto, S., . . . Wang, P. (2012). Ordering effects and choice set awareness in repeat-response stated preference studies. *Journal of environmental economics and management, 63*(1), 73-91. doi:https://doi.org/10.1016/j.jeem.2011.09.001
de Bekker-Grob, E., Rose, J., Donkers, B., Essink-Bot, M.-L., Bangma, C., & Steyerberg, E. (2013). Men's preferences for prostate cancer screening: a discrete choice experiment. *British Journal of Cancer, 108*(3), 533-541.
de Bekker-Grob, E., Rose, J. M., Donkers, B., Essink-Bot, M. L., Bangma, C. H., & Steyerberg, E. W. (2013). Men's preferences for prostate cancer screening: A discrete choice experiment. *British Journal of Cancer, 108*(3), 533-541. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1038/bjc.2013.5
de Bekker-Grob, E. W., Donkers, B., Jonker, M. F., & Stolk, E. A. (2015). Sample size requirements for discrete-choice experiments in healthcare: a practical guide. *The Patient-Patient-Centered Outcomes Research, 8*(5), 373-384.
de Bekker-Grob, E. W., Donkers, B., Veldwijk, J., Jonker, M. F., Buis, S., Huisman, J., & Bindels, P. (2020). What Factors Influence Non-Participation Most in Colorectal Cancer Screening? A Discrete Choice Experiment. *The Patient-Patient-Centered Outcomes Research*, 1-13.
De Bekker-Grob, E. W., Rose, J. M., Donkers, B., Essink-Bot, M. L., Bangma, C. H., & Steyerberg, E. W. (2013). Men's preferences for prostate cancer screening: A discrete choice experiment.

British Journal of Cancer, 108(3), 533-541. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1038/bjc.2013.5
de Bekker-Grob, E. W., Ryan, M., & Gerard, K. (2012). Discrete choice experiments in health economics: a review of the literature. *Health Econ, 21*(2), 145-172. doi:10.1002/hec.1697
de Bekker-Grob, E. W., Ryan, M., & Gerard, K. (2012). Discrete choice experiments in health economics: a review of the literature. *Health Economics, 21*(2), 145-172.
De Brún, A., Flynn, D., Ternent, L., Price, C. I., Rodgers, H., Ford, G. A., . . . Teah, J. (2018). A novel design process for selection of attributes for inclusion in discrete choice experiments: case study exploring variation in clinical decision-making about thrombolysis in the treatment of acute ischaemic stroke. *BMC health services research, 18*(1), 1-14.
Debreu, G. (1959). *Theory of value: An axiomatic analysis of economic equilibrium*: Yale University Press.
Deed, K., Childs, J., & Thoirs, K. (2014). What are the perceptions of women towards transvaginal sonographic examinations? *Sonography, 1*(2), 33-38.
Dekker, T. (2014). Indifference based value of time measures for Random Regret Minimisation models. *Journal of choice modelling, 12*, 10-20.
Department of Health. (2010). Equity and excellence: Liberating the NHS. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213823/dh_117794.pdf
Department of Health. (2015). The NHS Constitution. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/fi le/480482/NHS_Constitution_WEB.pdf
Determann, D., Gyrd-Hansen, D., de Wit, G. A., de Bekker-Grob, E. W., Steyerberg, E. W., Lambooij, M. S., & Bjørnskov Pedersen, L. (2019). Designing unforced choice experiments to inform health care decision making: implications of using opt-out, neither, or status quo alternatives in discrete choice experiments. *Medical Decision Making, 39*(6), 681-692.
Devlin, N. J., Shah, K. K., Feng, Y., Mulhern, B., & van Hout, B. (2018). Valuing health-related quality of life: An EQ-5 D-5 L value set for E ngland. *Health Economics, 27*(1), 7-22.
Dhar, R. (1997). Consumer preference for a no-choice option. *Journal of consumer research, 24*(2), 215-231.
Dhillon, N., Oliffe, J. L., Kelly, M. T., & Krist, J. (2020). Bridging barriers to cervical cancer screening in transgender men: a scoping review. *American journal of men's health, 14*(3), 1557988320925691.
Dochez, V., Caillon, H., Vaucel, E., Dimet, J., Winer, N., & Ducarme, G. (2019). Biomarkers and algorithms for diagnosis of ovarian cancer: CA125, HE4, RMI and ROMA, a review. *Journal of ovarian research, 12*(1), 1-9.
Doherty, E., Hobbins, A., Whitehurst, D. G., & O'Neill, C. (2021). An exploration on attribute non-attendance using discrete choice experiment data from the Irish EQ-5D-5L National Valuation Study. *PharmacoEconomics-Open, 5*(2), 237-244.
Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the european economic association, 9*(3), 522-550.
Dolan, P. (1999). Whose preferences count? *Medical Decision Making, 19*(4), 482-486.
Dolan, P., Gudex, C., Kind, P., & Williams, A. (1996). The time trade-off method: results from a general population study. *Health Economics, 5*(2), 141-154.
Donaldson, C., & Shackley, P. (1997). Does "process utility" exist? A case study of willingness to pay for laparoscopic cholecystectomy. *Social science & medicine, 44*(5), 699-707.
Doubeni, C. A., Doubeni, A. R., & Myers, A. E. (2016). Diagnosis and management of ovarian cancer. *American family physician, 93*(11), 937-944.
Drescher, C., Holt, S. K., Andersen, M. R., Anderson, G., & Urban, N. (2000). Reported ovarian cancer screening among a population-based sample in Washington State. *Obstetrics & Gynecology, 96*(1), 70-74.
Drescher, C. W., Nelson, J., Peacock, S., Andersen, M. R., McIntosh, M. W., & Urban, N. (2004). Compliance of average-and intermediate-risk women to semiannual ovarian cancer screening. *Cancer Epidemiology and Prevention Biomarkers, 13*(4), 600-606.
Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes*: Oxford university press.
Duffy, S. W., Vulkan, D., Cuckle, H., Parmar, D., Sheikh, S., Smith, R. A., . . . Moss, S. M. (2020). Effect of mammographic screening from age 40 years on breast cancer mortality (UK Age

trial): final results of a randomised, controlled trial. *The Lancet Oncology, 21*(9), 1165-1172. doi:10.1016/S1470-2045(20)30398-3
Ellimoottil, C., Marcotte, M., Grace, D., Krasnikov, A., Phillips, J. M., Quek, M. L., . . . Gupta, G. N. (2018). Examining Patient Willingness to Pay for Magnetic Resonance Imaging Guided Prostate Biopsy: Implications in the Era of Health Savings Accounts. *Urology Practice*. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1016/j.urpr.2017.03.003
Ells, C., & Thombs, B. D. (2014). The ethics of how to manage incidental findings. *CMAJ, 186*(9), 655-656.
Elwyn, G., Edwards, A., & Kinnersley, P. (1999). Shared decision-making in primary care: the neglected second half of the consultation. *British Journal of General Practice, 49*(443), 477-482.
Erdem, S., Campbell, D., & Hole, A. R. (2015). Accounting for attribute-level non-attendance in a health choice experiment: does it matter? *Health economics, 24*(7), 773-789.
Fagerlin, A., Zikmund-Fisher, B. J., & Ubel, P. A. (2011). Helping patients decide: ten steps to better risk communication. *Journal of the National Cancer Institute, 103*(19), 1436-1443.
Fallowfield, L., Fleissig, A., Barrett, J., Menon, U., Jacobs, I., Kilkerr, J., & Farewell, V. (2010). Awareness of ovarian cancer risk factors, beliefs and attitudes towards screening: baseline survey of 21 715 women participating in the UK Collaborative Trial of Ovarian Cancer Screening. *British Journal of Cancer, 103*(4), 454-461.
Fenichel, E. P., Lupi, F., Hoehn, J. P., & Kaplowitz, M. D. (2009). Split-sample tests of “no opinion” responses in an attribute-based choice model. *Land Economics, 85*(2), 348-362.
Fiebig, D. G., Haas, M., Hossain, I., Street, D. J., & Viney, R. (2009). Decisions about Pap tests: What influences women and providers? *Social Science and Medicine, 68*(10), 1766-1774. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1016/j.socscimed.2009.03.002
Finn, A., & Louviere, J. J. (1992). Determining the appropriate response to evidence of public concern: the case of food safety. *Journal of Public Policy & Marketing, 11*(2), 12-25.
Fitch, M. I., Deane, K., Howell, D., & Gray, R. E. (2002). Women’s experiences with ovarian cancer: reflections on being diagnosed. *Canadian Oncology Nursing Journal/Revue canadienne de soins infirmiers en oncologie, 12*(3), 152-159.
Florin, D., & Dixon, J. (2004). Public involvement in health care. *Bmj, 328*(7432), 159-161. doi:10.1136/bmj.328.7432.159
Flynn, T. N., Louviere, J. J., Peters, T. J., & Coast, J. (2007). Best–worst scaling: what it can do for health care research and how to do it. *Journal of health economics, 26*(1), 171-189.
Forman, D. (2009). Cancer Incidence and Survival by Major Ethnic Group, England, 2002-2006. *National Cancer Intelligence Network*.
Fremont, A. M., Cleary, P. D., Hargraves, J. L., Rowe, R. M., Jacobson, N. B., & Ayanian, J. Z. (2001). Patient-centered processes of care and long-term outcomes of myocardial infarction. *J Gen Intern Med, 16*(12), 800-808. doi:10.1111/j.1525-1497.2001.10102.x
Fruchter, R. G., & Boyce, J. (1981). Delays in diagnosis and stage of disease in gynecologic cancer. *Cancer Detect Prev, 4*(1-4), 481-486.
Funston, G., Hamilton, W., Abel, G., Crosbie, E. J., Rous, B., & Walter, F. M. (2020a). The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: A population-based cohort study. *PLoS medicine, 17*(10), e1003295.
Funston, G., Hardy, V., Abel, G., Crosbie, E. J., Emery, J., Hamilton, W., & Walter, F. M. (2020b). Identifying ovarian cancer in symptomatic women: a systematic review of clinical tools. *Cancers, 12*(12), 3686.
Funston, G., Mounce, L. T., Price, S., Rous, B., Crosbie, E. J., Hamilton, W., & Walter, F. M. (2021). CA125 test result, test-to-diagnosis interval, and stage in ovarian cancer at diagnosis: a retrospective cohort study using electronic health records. *British Journal of General Practice, 71*(707), e465-e472.
Funston, G., Van Melle, M., Baun, M.-L. L., Jensen, H., Helsper, C., Emery, J., . . . Walter, F. M. (2019). Variation in the initial assessment and investigation for ovarian cancer in symptomatic women: a systematic review of international guidelines. *BMC cancer, 19*(1), 1028.
Garcia-Retamero, R., & Cokely, E. T. (2017). Designing visual aids that promote risk literacy: a systematic review of health research and evidence-based design heuristics. *Human factors, 59*(4), 582-627.
Garcia-Retamero, R., & Galesic, M. (2009). Communicating treatment risk reduction to people with low numeracy skills: a cross-cultural comparison. *American journal of public health, 99*(12), 2196-2202.

Garcia-Retamero, R., Galesic, M., & Gigerenzer, G. (2010). Do icon arrays help reduce denominator neglect? *Medical Decision Making, 30*(6), 672-684.
Gatos, K. C. (2018). A Literature Review of Cervical Cancer Screening in Transgender Men. *Nursing for Women's Health, 22*(1), 52-62. doi:https://doi.org/10.1016/j.nwh.2017.12.008
Gentry-Maharaj, A., Sharma, A., Burnell, M., Ryan, A., Amso, N. N., Seif, M., . . . Rangar, R. (2013). Acceptance of transvaginal sonography by postmenopausal women participating in the United Kingdom Collaborative Trial of Ovarian Cancer Screening. *Ultrasound in obstetrics & gynecology, 41*(1), 73-79.
Georgescu-Roegen, N. (1958). Threshold in Choice and the Theory of Demand. *Econometrica: Journal of the Econometric Society*, 157-168.
Gerard, K., Shanahan, M., & Louviere, J. (2003). Using stated preference discrete choice modelling to inform health care decisionmaking: A pilot study of breast screening participation. *Applied Economics, 35*(9), 1073-1085. doi:10.1080/0003684032000081348
Ghanouni, A., Halligan, S., Taylor, S. A., Boone, D., Plumb, A., Stoffel, S., . . . Von Wagner, C. (2014). Quantifying public preferences for different bowel preparation options prior to screening CT colonography: A discrete choice experiment. *BMJ Open, 4*(4), e004327. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1136/bmjopen-2013-004327
Ghanouni, A., Sanderson, S. C., Pashayan, N., Renzi, C., Von Wagner, C., & Waller, J. (2020a). Attitudes towards risk-stratified breast cancer screening among women in England: A cross-sectional survey. *Journal of Medical Screening, 27*(3), 138-145.
Ghanouni, A., Smith, S. G., Halligan, S., Plumb, A., Boone, D., Yao, G. L., . . . Von Wagner, C. (2013). Public preferences for colorectal cancer screening tests: a review of conjoint analysis studies. *Expert review of medical devices, 10*(4), 489-499.
Ghanouni, A., Waller, J., Stoffel, S. T., Vlaev, I., & von Wagner, C. (2020b). Acceptability of risk-stratified breast screening: Effect of the order of presenting risk and benefit information. *Journal of Medical Screening, 27*(1), 52-56. doi:10.1177/0969141319877669
Gigerenzer, G., Mata, J., & Frank, R. (2009). Public knowledge of benefits of breast and prostate cancer screening in Europe. *Journal of the National Cancer Institute, 101*(17), 1216-1220.
Gilbert, L., Basso, O., Sampalis, J., Karp, I., Martins, C., Feng, J., . . . Takefman, J. (2012). Assessment of symptomatic women for early diagnosis of ovarian cancer: results from the prospective DOvE pilot project. *The Lancet Oncology, 13*(3), 285-291.
Gowdy, J. M., & Mayumi, K. (2001). Reformulating the foundations of consumer choice theory and environmental valuation. *Ecological Economics, 39*(2), 223-237.
Green, P. E., Krieger, A. M., & Agarwal, M. K. (1991). Adaptive conjoint analysis: Some caveats and suggestions. *Journal of Marketing Research, 28*(2), 215-222.
Greene, W. H., & Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological, 37*(8), 681-698.
Groothuis-Oudshoorn, C. G., Fermont, J. M., van Til, J. A., & Ijzerman, M. J. (2014). Public stated preferences and predicted uptake for genome-based colorectal cancer screening. *BMC Medical Informatics and Decision Making, 14*, 18. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1186/1472-6947-14-18
Gupta, K. K., Gupta, V. K., & Naumann, R. W. (2019). Ovarian cancer: screening and future directions. *International Journal of Gynecologic Cancer, 29*(1).
Hall, R., Medina-Lara, A., Hamilton, W., & Spencer, A. E. (2021). Attributes Used for Cancer Screening Discrete Choice Experiments: A Systematic Review. *The Patient-Patient-Centered Outcomes Research*, 1-17.
Hamilton, W. (2010). Cancer diagnosis in primary care. *British Journal of General Practice, 60*(571), 121. doi:10.3399/bjgp10X483175
Hamilton, W., Peters, T. J., Bankhead, C., & Sharp, D. (2009). Risk of ovarian cancer in women with symptoms in primary care: population based case-control study. *Bmj, 339*, b2998. doi:10.1136/bmj.b2998
Hanley, N., Mourato, S., & Wright, R. E. (2001). Choice modelling approaches: a superior alternative for environmental valuatioin? *Journal of economic surveys, 15*(3), 435-462.
Harker, N. (2017). How GPs can offer cancer patients the support they need. Retrieved from https://www.gponline.com/gps-offer-cancer-patients-support-need/cancer/article/1441477
Harrison, G. W., & Rutström, E. E. (2008). Risk aversion in the laboratory. In *Risk aversion in experiments*: Emerald Group Publishing Limited.
Harrison, M., Rigby, D., Vass, C., Flynn, T., Louviere, J., & Payne, K. (2014). Risk as an attribute in discrete choice experiments: a systematic review of the literature. *Patient, 7*(2), 151-170. doi:10.1007/s40271-014-0048-1

Harrison, M., Rigby, D., Vass, C., Flynn, T., Louviere, J., & Payne, K. (2014). Risk as an attribute in discrete choice experiments: a systematic review of the literature. *The Patient-Patient-Centered Outcomes Research, 7*(2), 151-170.
Hauber, A. B., González, J. M., Groothuis-Oudshoorn, C. G., Prior, T., Marshall, D. A., Cunningham, C., . . . Bridges, J. F. (2016). Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value in Health, 19*(4), 300-315.
Hauser, D. J., & Schwarz, N. (2015). It'sa trap! Instructional manipulation checks prompt systematic thinking on "tricky" tasks. *SAGE Open, 5*(2), 2158244015584617.
Hausman, D. M. (2011). *Preference, value, choice, and welfare*: Cambridge University Press.
He, P., Veronesi, M., & Engel, S. (2018). Consistency of risk preference measures: An artefactual field experiment from rural china. *The Journal of Development Studies, 54*(11), 1955-1973.
He, W., Li, M., Zuo, L., Wang, M., Jiang, L., Shan, H., . . . Han, X. (2019). Acupuncture for treatment of insomnia: An overview of systematic reviews. *Complementary therapies in medicine, 42*, 407-416.
Heidenreich, S., Watson, V., Ryan, M., & Phimister, E. (2018). Decision heuristic or preference? Attribute non-attendance in discrete choice problems. *Health Economics, 27*(1), 157-171.
Helter, T. M., & Boehler, C. E. H. (2016). Developing attributes for discrete choice experiments in health: a systematic literature review and case study of alcohol misuse interventions. *Journal of substance use, 21*(6), 662-668.
Henderson, J. T., Webber, E. M., & Sawaya, G. F. (2018). Screening for ovarian cancer: updated evidence report and systematic review for the US preventive services task force. *Jama, 319*(6), 595-606.
Hendrix, N., Hauber, B., Lee, C. I., Bansal, A., & Veenstra, D. L. (2020). Artificial intelligence in breast cancer screening: primary care provider preferences. *Journal of the American Medical Informatics Association*.
Hensher, D., Louviere, J., & Swait, J. (1998). Combining sources of preference data. *Journal of Econometrics, 89*(1-2), 197-221.
Hensher, D. A. (2006). Revealing differences in willingness to pay due to the dimensionality of stated choice designs: an initial assessment. *Environmental and resource economics, 34*(1), 7-44.
Hensher, D. A. (2010). Hypothetical bias, choice experiments and willingness to pay. *transportation research part B: methodological, 44*(6), 735-752.
Hensher, D. A., & Greene, W. H. (2010). Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification. *Empirical economics, 39*(2), 413-426.
Hensher, D. A., Rose, J., & Greene, W. H. (2005a). The implications on willingness to pay of respondents ignoring specific attributes. *Transportation, 32*(3), 203-222.
Hensher, D. A., Rose, J. M., & Black, I. (2008). Interactive agency choice in automobile purchase decisions: the role of negotiation in determining equilibrium choice outcomes. *Journal of Transport Economics and Policy (JTEP), 42*(2), 269-296.
Hensher, D. A., Rose, J. M., & Greene, W. H. (2005b). *Applied choice analysis: a primer*: Cambridge University Press.
Hensher, D. A., Rose, J. M., Rose, J. M., & Greene, W. H. (2005c). *Applied choice analysis: a primer*: Cambridge university press.
Hersch, J., Jansen, J., Barratt, A., Irwig, L., Houssami, N., Howard, K., . . . McCaffery, K. (2013). Women's views on overdiagnosis in breast cancer screening: a qualitative study. *Bmj, 346*.
Hess, S., & Hensher, D. A. (2010). Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. *Transportation Research Part B: Methodological, 44*(6), 781-790.
Hess, S., Hensher, D. A., & Daly, A. (2012). Not bored yet–revisiting respondent fatigue in stated choice experiments. *Transportation Research Part A: Policy and Practice, 46*(3), 626-644.
Hess, S., Stathopoulos, A., Campbell, D., O'Neill, V., & Caussade, S. (2013). It's not that I don't care, I just don't care very much: confounding between attribute non-attendance and taste heterogeneity. *Transportation, 40*(3), 583-607.
Hippisley-Cox, J., & Coupland, C. (2012). Identifying women with suspected ovarian cancer in primary care: derivation and validation of algorithm. *Bmj, 344*.
Hirst, Y., Stoffel, S., Baio, G., McGregor, L., & von Wagner, C. (2018). Uptake of the English Bowel (Colorectal) Cancer Screening Programme: an update 5 years after the full roll-out. *European Journal of Cancer, 103*, 267-273.
Hoffmann, T. C., & Del Mar, C. (2015). Patients' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. *JAMA internal medicine, 175*(2), 274-286.

Hoffrage, U., Gigerenzer, G., Krauss, S., & Martignon, L. (2002). Representation facilitates reasoning: what natural frequencies are and what they are not. *Cognition, 84*(3), 343-352.
Hol, L., De Bekker-Grob, E. W., Van Dam, L., Donkers, B., Kuipers, E. J., Habbema, J. D. F., . . . Essink-Bot, M. L. (2010). Preferences for colorectal cancer screening strategies: A discrete choice experiment. *British Journal of Cancer, 102*(6), 972-980. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1038/sj.bjc.6605566
Hole, A. R. (2006). Small-sample properties of tests for heteroscedasticity in the conditional logit model. *Economics Bulletin, 3*(18), 1-14.
Hole, A. R., Kolstad, J. R., & Gyrd-Hansen, D. (2013). Inferred vs. stated attribute non-attendance in choice experiments: A study of doctors' prescription behaviour. *Journal of Economic Behavior & Organization, 96*, 21-31.
Hole, A. R., Norman, R., & Viney, R. (2016). Response patterns in health state valuation using endogenous attribute attendance and latent class analysis. *Health economics, 25*(2), 212-224.
Hollinghurst, S., Banks, J., Bigwood, L., Walter, F. M., Hamilton, W., & Peters, T. J. (2016). Using willingness to pay to establish patient preferences for cancer testing in primary care. *BMC Medical Informatics and Decision Making, 16*(1), 105. doi:10.1186/s12911-016-0345-9
Holman, L. L., Lu, K. H., Bast Jr, R. C., Hernandez, M. A., Bodurka, D. C., Skates, S., & Sun, C. C. (2014). Risk perception, worry, and test acceptance in average-risk women who undergo ovarian cancer screening. *American journal of obstetrics and gynecology, 210*(3), 257. e251-257. e256.
Holmes, T. P., & Boyle, K. J. (2005). Dynamic learning and context-dependence in sequential, attribute-based, stated-preference valuation questions. *Land Economics, 81*(1), 114-126.
Horsky, D., Nelson, P., & Posavac, S. S. (2004). Stating preference for the ethereal but choosing the concrete: how the tangibility of attributes affects attribute weighting in value elicitation and choice. *Journal of Consumer Psychology, 14*(1-2), 132-140.
Howard, K., & Salkeld, G. (2009). Does attribute framing in discrete choice experiments influence willingness to pay? Results from a discrete choice experiment in screening for colorectal cancer. *Value in Health, 12*(2), 354-363. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1111/j.1524-4733.2008.00417.x
Howard, K., Salkeld, G., McCaffery, K., & Irwig, L. (2008). HPV triage testing or repeat Pap smear for the management of atypical squamous cells (ASCUS) on Pap smear: is there evidence of process utility? *Health Economics, 17*(5), 593-605.
Howard, K., Salkeld, G., Pignone, M., Hewett, P., Cheung, P., Olsen, J., . . . Roberts-Thomson, I. C. (2011). Preferences for CT colonography and colonoscopy as diagnostic tests for colorectal cancer: A discrete choice experiment. *Value in Health, 14*(8), 1146-1152. doi:http://dx.doi.org/10.1016/j.jval.2011.07.012
Howard, K., Salkeld, G. P., Patel, M. I., Mann, G. J., & Pignone, M. P. (2015). Men's preferences and trade-offs for prostate cancer screening: a discrete choice experiment. *Health expectations : an international journal of public participation in health care and health policy, 18*(6), 3123-3135. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1111/hex.12301
Hu, W., Adamowicz, W. L., & Veeman, M. M. (2006). Labeling Context and Reference Point Effects in Models of Food Attribute Demand. *American Journal of Agricultural Economics, 88*(4), 1034-1049. doi:https://doi.org/10.1111/j.1467-8276.2006.00914.x
Huber, J., & Zwerina, K. (1996). The importance of utility balance in efficient choice designs. *Journal of Marketing research, 33*(3), 307-317.
Huber, O. (1980). The influence of some task variables on cognitive operations in an information-processing decision model. *Acta Psychologica, 45*(1-3), 187-196.
Hurley, J. (1998). Welfarism, extra-welfarism and evaluative economic analysis in the health sector. *Health, health care and health economics: Perspectives on distribution*, 373-395.
Hurley, J. (2000). An overview of the normative economics of the health sector. In *Handbook of health economics* (Vol. 1, pp. 55-118): Elsevier.
Hurley, R. (2018). Overdiagnosis and the cancer label. In: British Medical Journal Publishing Group.
Incisive Health. (2014). *Saving lives averting costs: an analysis of the financial implications of achieving earlier diagnosis of colorectal, lung and ovarian cancer. A report prepared for Cancer Research UK.* Retrieved from http://www.incisivehealth.com/uploads/Saving%20lives%20averting%20costs.pdf
Jack, R. H., Møller, H., Robson, T., & Davies, E. A. (2014). Breast cancer screening uptake among women from different ethnic groups in London: a population-based cohort study. *BMJ open, 4*(10), e005586.

Jacobs, I. J., Menon, U., Ryan, A., Gentry-Maharaj, A., Burnell, M., Kalsi, J. K., . . . Cruickshank, D. (2016). Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *The Lancet, 387*(10022), 945-956.

Jelicic, L., Brooker, J., Shand, L., Knight, T., Ricciardelli, L., Denham, G., & Burney, S. (2019). Experiences and health care preferences of women with ovarian cancer during the diagnosis phase. *Psycho-oncology, 28*(2), 379-385.

Jenkins, V., Fallowfield, L., Langridge, C., Barrett, J., Ryan, A., Jacobs, I., . . . Farewell, V. (2015). Psychosocial factors associated with withdrawal from the United Kingdom collaborative trial of ovarian Cancer screening after 1 episode of repeat screening. *International Journal of Gynecologic Cancer, 25*(8).

Jiang, Y., & Punj, G. N. (2010). The effects of attribute concreteness and prominence on selective processing, choice, and search experience. *Journal of the Academy of Marketing Science, 38*(4), 471-489.

Johar, M., Fiebig, D. G., Haas, M., & Viney, R. (2013). Using repeated choice experiments to evaluate the impact of policy changes on cervical screening. *Applied Economics, 45*(14), 1845-1855.

Johnson, F. R., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D. A., . . . Bridges, J. F. (2013). Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value in Health, 16*(1), 3-13.

Johnson, F. R., Yang, J.-C., & Reed, S. D. (2019). The internal validity of discrete choice experiment data: a testing tool for quantitative assessments. *Value in health, 22*(2), 157-160.

Jonker, M. F., Donkers, B., de Bekker-Grob, E. W., & Stolk, E. A. (2018). Effect of level overlap and color coding on attribute non-attendance in discrete choice experiments. *Value in health, 21*(7), 767-771.

Jonker, M. F., Donkers, B., de Bekker-Grob, E., & Stolk, E. A. (2019). Attribute level overlap (and color coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. *Health Economics, 28*(3), 350-363.

Katz, S. J., Belkora, J., & Elwyn, G. (2014). Shared Decision Making for Treatment of Cancer: Challenges and Opportunities. *Journal of Oncology Practice, 10*(3), 206-208. doi:10.1200/jop.2014.001434

Kim, C., Armstrong, M. J., Berta, W. B., & Gagliardi, A. R. (2020). How to identify, incorporate and report patient preferences in clinical guidelines: A scoping review. *Health Expectations, 23*(5), 1028-1036.

Kingsley, D. C., & Brown, T. C. (2010). Preference uncertainty, preference learning, and paired comparison experiments. *Land Economics, 86*(3), 530-544.

Kistler, C. E., Hess, T. M., Howard, K., Pignone, M. P., Crutchfield, T. M., Hawley, S. T., . . . Lewis, C. L. (2015). Older adults' preferences for colorectal cancer-screening test attributes and test choice. *Patient Preference and Adherence, 9*, 1005-1016. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.2147/PPA.S82203

Kitchener, H. C., Gittins, M., Rivero-Arias, O., Tsiachristas, A., Cruickshank, M., Gray, A., . . . Roberts, C. (2016). A cluster randomised trial of strategies to increase cervical screening uptake at first invitation (STRATEGIC). *Health Technology Assessment, 20*(68). doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.3310/hta20680

Kjær, T., Bech, M., Gyrd-Hansen, D., & Hart-Hansen, K. (2006). Ordering effect and price sensitivity in discrete choice experiments: need we worry? *Health Economics, 15*(11), 1217-1228.

Kleij, K.-S., Tangermann, U., Amelung, V. E., & Krauth, C. (2017). Patients' preferences for primary health care–a systematic literature review of discrete choice experiments. *BMC health services research, 17*(1), 1-12.

Kløjgaard, M. E., Bech, M., & Søgaard, R. (2012). Designing a stated choice experiment: the value of a qualitative process. *Journal of choice modelling, 5*(2), 1-18.

Kobayashi, H., Yamada, Y., Sado, T., Sakata, M., Yoshida, S., Kawaguchi, R., . . . Tsuji, Y. (2008). A randomized study of screening for ovarian cancer: a multicenter study in Japan. *International Journal of Gynecologic Cancer, 18*(3).

Kohler, R. E., Gopal, S., Lee, C. N., Weiner, B. J., Reeve, B. B., & Wheeler, S. B. (2017). Breast Cancer Knowledge, Behaviors, and Preferences in Malawi: Implications for Early Detection Interventions From a Discrete Choice Experiment. *Journal of global oncology, 3*(5), 480-489. doi:https://0-dx-doi-org.lib.exeter.ac.uk/10.1200/JGO.2016.005371

Krishnan, K. (1977). Incorporating thresholds of indifference in probabilistic choice models. *Management science, 23*(11), 1224-1233.

Krucien, N., Watson, V., & Ryan, M. (2017). Is best–worst scaling suitable for health state valuation? A comparison with discrete choice experiments. *Health Economics, 26*(12), e1-e16.
Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of political economy, 74*(2), 132-157.
Lancsar, E., & Louviere, J. (2006). Deleting 'irrational'responses from discrete choice experiments: a case of investigating or imposing preferences? *Health economics, 15*(8), 797-811.
Lancsar, E., & Louviere, J. (2008). Conducting discrete choice experiments to inform healthcare decision making. *PharmacoEconomics, 26*(8), 661-677.
Lancsar, E., & Swait, J. (2014). Reconceptualising the external validity of discrete choice experiments. *Pharmacoeconomics, 32*(10), 951-965.
Lee, M. J., & Kang, H. (2018). Designing skin cancer prevention messages: Should we emphasize gains or losses? Message framing, risk type, and prior experience. *American Journal of Health Promotion, 32*(4), 939-948.
Leuraud, K., Jezewski-Serra, D., Viguier, J., & Salines, E. (2013). Colorectal cancer screening by guaiac faecal occult blood test in France: Evaluation of the programme two years after launching. *Cancer epidemiology, 37*(6), 959-967.
Lew, D. K., & Whitehead, J. C. (2020). Attribute non-attendance as an information processing strategy in stated preference choice experiments: Origins, current practices, and future directions. *Marine Resource Economics, 35*(3), 285-317.
Li, S., Liu, S., Ratcliffe, J., Gray, A., & Chen, G. (2019). Preferences for cervical cancer screening service attributes in rural China: a discrete choice experiment. *Patient Preference and Adherence, 13*, 881.
Light, A., Elhage, O., Marconi, L., & Dasgupta, P. (2019). Prostate cancer screening: where are we now? *BJU International, 123*(6), 916-917.
Lim, A., Mesher, D., Gentry-Maharaj, A., Balogun, N., Widschwendter, M., Jacobs, I., . . . Menon, U. (2016). Time to diagnosis of Type I or II invasive epithelial ovarian cancers: a multicentre observational study using patient questionnaire and primary care records. *Bjog, 123*(6), 1012-1020. doi:10.1111/1471-0528.13447
Lim, S. L., Yang, J.-C., Ehrisman, J., Havrilesky, L. J., & Reed, S. D. (2020). Are Videos or Text Better for Describing Attributes in Stated-Preference Surveys? *The Patient - Patient-Centered Outcomes Research, 13*(4), 401-408. doi:10.1007/s40271-020-00416-9
Little, P., Dorward, M., Warner, G., Stephens, K., Senior, J., & Moore, M. (2004). Importance of patient pressure and perceived pressure and perceived medical need for investigations, referral, and prescribing in primary care: nested observational study. *Bmj, 328*(7437), 444.
Lockwood, M. (1996). Non-compensatory preference structures in non-market valuation of natural area policy. *Australian journal of Agricultural economics, 40*(2), 85-101.
Louviere, J., Lings, I., Islam, T., Gudergan, S., & Flynn, T. (2013). An introduction to the application of (case 1) best–worst scaling in marketing research. *International Journal of Research in Marketing, 30*(3), 292-303.
Louviere, J. J. (1988). Conjoint analysis modelling of stated preferences: a review of theory, methods, recent developments and external validity. *Journal of transport economics and policy*, 93-119.
Louviere, J. J., Flynn, T. N., & Carson, R. T. (2010). Discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling, 3*(3), 57-72.
Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). *Best-worst scaling: Theory, methods and applications*: Cambridge University Press.
Lurie, G., Wilkens, L. R., Thompson, P. J., Matsuno, R. K., Carney, M. E., & Goodman, M. T. (2010). Symptom presentation in invasive ovarian carcinoma by tumor histological type and grade in a multiethnic population: a case analysis. *Gynecol Oncol, 119*(2), 278-284. doi:10.1016/j.ygyno.2010.05.028
Luu, J. M., Sergeant, A. K., Anand, S. S., Desai, D., Schulze, K., Knoppers, B. M., . . . Black, S. E. (2021). The impact of reporting magnetic resonance imaging incidental findings in the Canadian alliance for healthy hearts and minds cohort. *BMC medical ethics, 22*(1), 1-15.
Lyratzopoulos, G., Saunders, C. L., Abel, G. A., McPhail, S., Neal, R. D., Wardle, J., & Rubin, G. P. (2015). The relative length of the patient and the primary care interval in patients with 28 common and rarer cancers. *British Journal of Cancer, 112*(1), S35-S40. doi:10.1038/bjc.2015.40
Macdonald, S., Macleod, U., Campbell, N. C., Weller, D., & Mitchell, E. (2006). Systematic review of factors influencing patient and practitioner delay in diagnosis of upper gastrointestinal cancer. *British journal of cancer, 94*(9), 1272-1280.

Maclennan, D., & Williams, N. (1980). Revealed-preference theory and spatial choices: some limitations. *Environment and Planning A, 12*(8), 909-919.
Mandrik, O., Yaumenenka, A., Herrero, R., & Jonker, M. F. (2019). Population preferences for breast cancer screening policies: Discrete choice experiment in Belarus. *PLoS ONE, 14*(11).
Mansfield, C., Ekwueme, D. U., Tangka, F. K. L., Brown, D. S., Smith, J. L., Guy, G. P., . . . Hauber, B. (2018). Colorectal Cancer Screening: Preferences, Past Behavior, and Future Intentions. *Patient*, 1-13. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1007/s40271-018-0308-6
Mansfield, C., Tangka, F. K., Ekwueme, D. U., Smith, J. L., Guy, G. P., Li, C., & Hauber, A. B. (2016). Stated Preference for Cancer Screening: A Systematic Review of the Literature, 1990-2013. *Preventing chronic disease, 13*, E27. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.5888/pcd13.150433
Marley, A., Flynn, T. N., & Australia, V. (2012). Best worst scaling: theory and practice. *Handbook of Choice Modelling. Edward Elgar Publishing, Leeds (UK)*.
Marshall, D., McGregor, E., & Currie, G. (2010). Measuring Preferences for Colorectal Cancer Screening. *The Patient: Patient-Centered Outcomes Research, 3*(2), 79-89. doi:10.2165/11532250-000000000-00000
Marshall, D. A., Johnson, F. R., Phillips, K. A., Marshall, J. K., Thabane, L., & Kulin, N. A. (2007). Measuring patient preferences for colorectal cancer screening using a choice-format survey. *Value in Health, 10*(5), 415-430. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1111/j.1524-4733.2007.00196.x
Marshall, D. A., Johnson, R., Kulin, N. A., Ozdemir, A., Walsh, A., Marshall, J., . . . Phillips, K. (2009). How Do Physician Assessments of Patient Preferences for Colorectal Cancer Screening Tests Differ from Actual Preferences? A Comparison in Canada and the United States Using a Stated-Choice Survey. *Health Economics, 18*(12), 1420-1439. doi:http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291099-1050/issues
Martens, C. E., Crutchfield, T. M., Laping, J. L., Perreras, L., Reuland, D. S., Cubillos, L., . . . Wheeler, S. B. (2016). Why Wait Until Our Community Gets Cancer?: Exploring CRC Screening Barriers and Facilitators in the Spanish-Speaking Community in North Carolina. *Journal of cancer education : the official journal of the American Association for Cancer Education, 31*(4), 652-659. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1007/s13187-015-0890-4
Mbuya Bienge, C., Pashayan, N., Brooks, J. D., Dorval, M., Chiquette, J., Eloy, L., . . . Lévesque, E. (2021). Women's Views on Multifactorial Breast Cancer Risk Assessment and Risk-Stratified Screening: A Population-Based Survey from Four Provinces in Canada. *Journal of personalized medicine, 11*(2), 95.
McFadden, D. (1974). The measurement of urban travel demand. *Journal of public economics, 3*(4), 303-328.
McFadden, D. (1986). The choice theory approach to market research. *Marketing science, 5*(4), 275-297.
Medina-Lara, A., Mujica-Mota, R. E., Kunkwenzu, E. D., & Lalloo, D. G. (2014). Stated preferences for anti-malarial drug characteristics in Zomba, a malaria endemic area of Malawi. *Malaria journal, 13*(1), 1-9.
Meertens, R. M., & Lion, R. (2008). Measuring an individual's tendency to take risks: the risk propensity scale 1. *Journal of Applied Social Psychology, 38*(6), 1506-1520.
Menczer, J. (2000). Diagnosis and treatment delay in gynecological malignancies. Does it affect outcome? *International Journal of Gynecological Cancer, 10*(2), 89-94.
Menon, U., Gentry-Maharaj, A., Burnell, M., Singh, N., Ryan, A., Karpinskyj, C., . . . Raikou, M. (2021). Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *The Lancet, 397*(10290), 2182-2193.
Menon, U., Gentry-Maharaj, A., Hallett, R., Ryan, A., Burnell, M., Sharma, A., . . . Lopes, A. (2009). Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *The lancet oncology, 10*(4), 327-340.
Menon, U., McGuire, A. J., Raikou, M., Ryan, A., Davies, S. K., Burnell, M., . . . Amso, N. N. (2017). The cost-effectiveness of screening for ovarian cancer: results from the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *British Journal of Cancer, 117*(5), 619-627.
Menzel, P., Dolan, P., Richardson, J., & Olsen, J. A. (2002). The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. *Social science & medicine, 55*(12), 2149-2158.

Michalovic, E., Hall, S., Duncan, L. R., Bassett-Gunter, R., & Sweet, S. N. (2018). Understanding the effects of message framing on physical activity action planning: the role of risk perception and elaboration. *International journal of behavioral medicine, 25*(6), 626-636.
Michel-Lepage, A., Ventelou, B., Nebout, A., Verger, P., & Pulcini, C. (2013). Cross-sectional survey: risk-averse French GPs use more rapid-antigen diagnostic tests in tonsillitis in children. *BMJ open, 3*(10), e003540.
Miles, A., Taylor, S. A., Evans, R. E., Halligan, S., Beare, S., Bridgewater, J., . . . Oliver, A. (2019). Patient preferences for whole-body MRI or conventional staging pathways in lung and colorectal cancer: a discrete choice experiment. *European radiology, 29*(7), 3889-3900.
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine, 151*(4), 264-269.
Mooney, G. (1994). *Key issues in health economics* London: Harvester Wheatsheaf.
Mooney, G. (2005). Communitarian claims and community capabilities: furthering priority setting? *Social science & medicine, 60*(2), 247-255.
Mooney, G., & Lange, M. (1993). Ante-natal screening: What constitutes 'benefit'? *Social science & medicine, 37*(7), 873-878.
Moore, P. A., Ziegler, K. M., Lipman, R. D., Aminoshariae, A., Carrasco-Labra, A., & Mariotti, A. (2018). Benefits and harms associated with analgesic medications used in the management of acute dental pain: an overview of systematic reviews. *The Journal of the American Dental Association, 149*(4), 256-265. e253.
Moore, S. F., Price, S. J., Chowienczyk, S., Bostock, J., & Hamilton, W. (2021). The impact of changing risk thresholds on the number of people in England eligible for urgent investigation for possible cancer: an observational cross-sectional study. *Br J Cancer, 125*(11), 1593-1597. doi:10.1038/s41416-021-01541-4
Morris, S., Devlin, N., Parkin, D., & Spencer, A. (2012). *Economic Analysis in Healthcare*: Wiley.
Moser, K., Patnick, J., & Beral, V. (2009). Inequalities in reported use of breast and cervical screening in Great Britain: analysis of cross sectional survey data. *Bmj, 338*, b2025. doi:10.1136/bmj.b2025
Moss, E., Hollingworth, J., & Reynolds, T. (2005). The role of CA125 in clinical practice. *Journal of clinical pathology, 58*(3), 308-312.
Moss, E. L., Moran, A., Reynolds, T. M., & Stokes-Lampard, H. (2013). Views of general practitioners on the role of CA125 in primary care to diagnose ovarian cancer. *BMC women's health, 13*(1), 1-6.
Mühlbacher, A., & Johnson, F. R. (2016). Choice experiments to quantify preferences for health and healthcare: state of the practice. *Applied health economics and health policy, 14*(3), 253-266.
Mühlbacher, A. C., Bridges, J. F., Bethge, S., Dintsios, C., Schwalm, A., Gerber-Grote, A., & Nübling, M. (2017). Preferences for antiviral therapy of chronic hepatitis C: a discrete choice experiment. *The European Journal of Health Economics, 18*(2), 155-165.
Mühlbacher, A. C., Kaczynski, A., Zweifel, P., & Johnson, F. R. (2016). Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. *Health Economics Review, 6*(1), 2. doi:10.1186/s13561-015-0079-x
Mujica-Mota, R. E., Roberts, M., Abel, G., Elliott, M., Lyratzopoulos, G., Roland, M., & Campbell, J. (2015). Common patterns of morbidity and multi-morbidity and their impact on health-related quality of life: evidence from a national survey. *Qual Life Res, 24*(4), 909-918. doi:10.1007/s11136-014-0820-7
Mulley, A. G., Trimble, C., & Elwyn, G. (2012). Stop the silent misdiagnosis: patients' preferences matter. *Bmj, 345*.
Nagle, C. M., Francis, J. E., Nelson, A. E., Zorbas, H., Luxford, K., De Fazio, A., . . . Webb, P. M. (2011). Reducing time to diagnosis does not improve outcomes for women with symptomatic ovarian cancer: a report from the Australian Ovarian Cancer Study Group. *J Clin Oncol, 29*(16), 2253-2258.
Naik, G., Ahmed, H., & Edwards, A. G. (2012). Communicating risk to patients and the public. *British Journal of General Practice, 62*(597), 213-216.
Nash, Z., & Menon, U. (2020). Ovarian cancer screening: Current status and future directions. *Best practice & research Clinical obstetrics & gynaecology, 65*, 32-45.
National Academies of Sciences, E., & Medicine. (2016). *Ovarian cancers: evolving paradigms in research and care*: National Academies Press.
National Screening Committee. (2015). Criteria for appraising the viability, effectiveness and appropriateness of a screening programme. Retrieved from

https://www.gov.uk/government/publications/evidence-review-criteria-national-screening-programmes/criteria-for-appraising-the-viability-effectiveness-and-appropriateness-of-a-screening-programme
Nayaradou, M., Berchi, C., Dejardin, O., & Launoy, G. (2010). Eliciting population preferences for mass colorectal cancer screening organization. *Medical decision making : an international journal of the Society for Medical Decision Making, 30*(2), 224-233.
NCRAS. (2019a). Routes to Diagnosis 2006-2016 workbook (a). Retrieved from http://www.ncin.org.uk/publications/routes_to_diagnosis
NCRAS. (2019b). Stage by Routes to Diagnosis, 2015-2016. Retrieved from http://www.ncin.org.uk/publications/routes_to_diagnosis
Neal, D. E., Feit, E. M., & Etzkorn, J. R. (2018). Patient Preferences for the Treatment of Basal Cell Carcinoma: A Mapping Review of Discrete Choice Experiments. *Dermatol Surg, 44*(8), 1041-1049. doi:10.1097/dss.0000000000001564
Neal, R. D., Allgar, V. L., Ali, N., Leese, B., Heywood, P., Proctor, G., & Evans, J. (2007). Stage, survival and delays in lung, colorectal, prostate and ovarian cancer: comparison between diagnostic routes. *British Journal of General Practice, 57*(536), 212-219.
Neal, R. D., Tharmanathan, P., France, B., Din, N. U., Cotton, S., Fallon-Ferguson, J., . . . Emery, J. (2015). Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *Br J Cancer, 112 Suppl 1*(Suppl 1), S92-107. doi:10.1038/bjc.2015.48
Neuman, T., Neuman, E., & Neuman, S. (2010). Explorations of the effect of experience on preferences for a health-care service. *The Journal of Socio-Economics, 39*(3), 407-419.
Nguyen, T. C., Robinson, J., Whitty, J. A., Kaneko, S., & Nguyen, T. C. (2015). Attribute non-attendance in discrete choice experiments: A case study in a developing country. *Economic Analysis and Policy, 47*, 22-33.
NHS Digital. (2018). *Cervical Screening Programme England 2017-18*. Retrieved from
NHS Digital. (2019). Breast Screening Programme, England 2017-18. Retrieved from https://digital.nhs.uk/data-and-information/publications/statistical/breast-screening-programme/england-2017-18
NHS England. (2016). Achieving World Class Cancer Outcomes: Taking the strategy forward: five year forward. Retrieved from https://www.england.nhs.uk/wp-content/uploads/2016/05/cancer-strategy.pdf
NHS England. (2018). NHS Diagnostic Waiting Times and Activity Data. Retrieved from https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2018/06/DWTA-Report-April-2018.pdf
NHS England. (2022). Cervical screening: quarterly coverage data reports 2021. Retrieved from https://www.gov.uk/government/publications/cervical-screening-quarterly-coverage-data-reports-2021
NHS UK. (2020). Diagnosis: Ovarian Cancer. Retrieved from https://www.nhs.uk/conditions/ovarian-cancer/diagnosis/
NHS UK. (2021a). How to decide if you want breast screening. Retrieved from https://www.nhs.uk/conditions/breast-screening-mammogram/how-to-decide-if-you-want-breast-screening/
NHS UK. (2021b). NHS screening. Retrieved from https://www.nhs.uk/conditions/nhs-screening/
NICE. (2011). Ovarian cancer: recognition and initial management. Retrieved from https://www.nice.org.uk/guidance/CG122
NICE. (2015). Suspected cancer: recognition and referral (NG12). Retrieved from www.nice.org.uk/guidance/ng12
NICE. (2017). Quantitative faecal immunochemical tests to guide referral for colorectal cancer in primary care [DG30]. Retrieved from https://www.nice.org.uk/guidance/DG30
NICE. (2019). Lung cancer: diagnosis and management [DG122]. Retrieved from https://www.nice.org.uk/guidance/ng122
NICE. (2021). Prostate cancer: diagnosis and management [DG131]. Retrieved from https://www.nice.org.uk/guidance/ng131
Nisbett, R. E., & Ross, L. (1980). Human inference: Strategies and shortcomings of social judgment.
Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies* (Vol. 11): sage.
Obadha, M., Barasa, E., Kazungu, J., Abiiro, G. A., & Chuma, J. (2019). Attribute development and level selection for a discrete choice experiment to elicit the preferences of health care providers for capitation payment mechanism in Kenya. *Health economics review, 9*(1), 1-19.

Oberlin, A. M., Pasipamire, T., & Chibwesha, C. J. (2019). Exploring women's preferences for HPV-based cervical cancer screening in South Africa. *International Journal of Gynecology & Obstetrics*.
Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. (2015). Improving risk understanding across ability levels: Encouraging active processing with dynamic icon arrays. *Journal of Experimental Psychology: Applied, 21*(2), 178.
Okeji, M. C., Agwuna, K. K., Ihudiebube-Splendor, C. N., Izge, I. Y., Ekuma, K. K., & Emeter, J. O. (2017). Transvaginal Sonography: perception and attitude of Nigerian women. *BMC women's health, 17*(1), 1-4.
ONS. (2019a). Cancer registration statistics, England: 2017. Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/2017
ONS. (2019b). Cancer survival in England - adults diagnosed. Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancersurvivalratescancersurvivalinenglandadultsdiagnosed
ONS. (2021). *Population estimates by ethnic group, England and Wale*. Retrieved from: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/datasets/populationestimatesbyethnicgroupenglandandwales
Ortiz-Muñoz, B., Aznar-Oroval, E., García, A. G., Peris, A. C., Ballestero, P. P., Yepes, M. S., . . . Garcia, E. G. (2014). HE4, Ca125 and ROMA algorithm for differential diagnosis between benign gynaecological diseases and ovarian cancer. *Tumor Biology, 35*(7), 7249-7258.
Osborne, J. M., Flight, I., Wilson, C. J., Chen, G., Ratcliffe, J., & Young, G. P. (2018). The impact of sample type and procedural attributes on relative acceptability of different colorectal cancer screening regimens. *Patient Preference and Adherence, 12*, 1825-1836. doi:http://dx.doi.org/10.2147/PPA.S172143
Ostermann, J., Brown, D. S., de Bekker-Grob, E. W., Mühlbacher, A. C., & Reed, S. D. (2017). Preferences for health interventions: improving uptake, adherence, and efficiency. *The Patient-Patient-Centered Outcomes Research, 10*(4), 511-514.
Palan, S., & Schitter, C. (2018). Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22-27.
Pan, X., & Zuo, Z. (2020). Exploring the role of utility-difference threshold in choice behavior: An empirical case study of bus service choice. *International Journal of Transportation Science and Technology, 9*(2), 128-136.
Papin-Lefebvre, F., Guillaume, E., Moutel, G., Launoy, G., & Berchi, C. (2017). General practitioners' preferences with regard to colorectal cancer screening organisation Colon cancer screening medico-legal aspects. *Health Policy, 121*(10), 1079-1084. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1016/j.healthpol.2017.08.013
Park, S. H., Lee, B., Lee, S., Choi, E., Choi, E.-B., Yoo, J., . . . Choi, K. S. (2015). A qualitative study of women's views on overdiagnosis and screening for thyroid cancer in Korea. *BMC cancer, 15*(1), 1-8.
Pavlik, E. J., van Nagell Jr, J. R., DePriest, P. D., Wheeler, L., Tatman, J. M., Boone, M., . . . Kryscio, R. K. (1995). Participation in transvaginal ovarian cancer screening: compliance, correlation factors, and costs. *Gynecologic oncology, 57*(3), 395-400.
Pearce, A. M., Mulhern, B. J., Watson, V., & Viney, R. C. (2020). How are debriefing questions used in health discrete choice experiments? An online survey. *Value in health, 23*(3), 289-293.
Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153-163.
Peitzmeier, S. M., Reisner, S. L., Harigopal, P., & Potter, J. (2014). Female-to-male patients have high prevalence of unsatisfactory Paps compared to non-transgender females: implications for cervical cancer screening. *Journal of general internal medicine, 29*(5), 778-784.
Pellat, A., Deyra, J., Coriat, R., & Chaussade, S. (2018). Results of the national organised colorectal cancer screening program with FIT in Paris. *Scientific reports, 8*(1), 1-4.
Peters, Y., & Siersema, P. D. (2020). Public Preferences and Predicted Uptake for Esophageal Cancer Screening Strategies: A Labeled Discrete Choice Experiment. *Clinical and Translational Gastroenterology, 11*(11).
Peterson, E. B., Ostroff, J. S., DuHamel, K. N., D'Agostino, T. A., Hernández, M., Canzona, M. R., & Bylund, C. L. (2016). Impact of provider-patient communication on cancer screening adherence: a systematic review. *Preventive Medicine, 93*, 96-105.
Philips, Z., Ginnelly, L., Sculpher, M., Claxton, K., Golder, S., Riemsma, R., . . . Glanville, J. (2004). Review of guidelines for good practice in decision-analytic modelling in health technology

assessment. In *NIHR Health Technology Assessment programme: Executive Summaries*: NIHR Journals Library.
Phillips, K., Hersch, J., Turner, R., Jansen, J., & McCaffery, K. (2016). The influence of the 'cancer effect'on young women's responses to overdiagnosis in cervical screening. *Patient Education and Counseling, 99*(10), 1568-1575.
Phillips, K., Van, S. B., Marshall, D., Walsh, J., & Thabane, L. (2006). A review of studies examining stated preferences for cancer screening. In: Centers for Disease Control and Prevention.
Phillips, K. A., Johnson, F. R., & Maddala, T. (2002). Measuring what people value: a comparison of "attitude" and "preference" surveys. *Health services research, 37*(6), 1659-1679.
Pignone, M. P., Brenner, A. T., Hawley, S., Sheridan, S. L., Lewis, C. L., Jonas, D. E., & Howard, K. (2012). Conjoint analysis versus rating and ranking for values elicitation and clarification in colorectal cancer screening. *Journal of General Internal Medicine, 27*(1), 45-50. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1007/s11606-011-1837-z
Pignone, M. P., Crutchfield, T. M., Brown, P. M., Hawley, S. T., Laping, J. L., Lewis, C. L., . . . Wheeler, S. B. (2014). Using a discrete choice experiment to inform the design of programs to promote colon cancer screening for vulnerable populations in North Carolina. *BMC Health Services Research, 14*, 611. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1186/s12913-014-0611-4
Pignone, M. P., Howard, K., Brenner, A. T., Crutchfield, T. M., Hawley, S. T., Lewis, C. L., & Sheridan, S. L. (2013). Comparing 3 techniques for eliciting patient values for decision making about prostate-specific antigen screening: A randomized controlled trial. *JAMA Internal Medicine, 173*(5), 362-368. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1001/jamainternmed.2013.2651
Pinsky, P. F., Yu, K., Kramer, B. S., Black, A., Buys, S. S., Partridge, E., . . . Prorok, P. C. (2016). Extended mortality results for ovarian cancer screening in the PLCO trial with median 15 years follow-up. *Gynecologic oncology, 143*(2), 270-275.
Pinto-Prades, J. L., McHugh, N., Donaldson, C., & Manoukian, S. (2019). Sequence effects in time trade-off valuation of hypothetical health states. *Health economics, 28*(11), 1308-1319.
Plott, C. R. (1993). Rational individual behavior in markets and social choice processes.
Plott, C. R., & Zeiler, K. (2005). The willingness to pay-willingness to accept gap, the" endowment effect," subject misconceptions, and experimental procedures for eliciting valuations. *American Economic Review, 95*(3), 530-545.
Plumb, A. A., Boone, D., Fitzke, H., Helbren, E., Mallett, S., Zhu, S., . . . Halligan, S. (2014). Detection of extracolonic pathologic findings with CT colonography: a discrete choice experiment of perceived benefits versus harms. *Radiology, 273*(1), 144-152. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1148/radiol.14131678
Polaris, J. J. Z., & Katz, J. N. (2014). "Appropriate" diagnostic testing: supporting diagnostics with evidence-based medicine and shared decision making. *BMC Research Notes, 7*(1), 922. doi:10.1186/1756-0500-7-922
Porta, M., Gallen, M., Belloc, J., & Malats, N. (1996). Predictors of the interval between onset of symptoms and first medical visit in patients with digestive tract cancer. *Int J Oncol, 8*(5), 941-949. doi:10.3892/ijo.8.5.941
Potoglou, D., Burge, P., Flynn, T., Netten, A., Malley, J., Forder, J., & Brazier, J. E. (2011). Best–worst scaling vs. discrete choice experiments: an empirical comparison using social care data. *Social Science & Medicine, 72*(10), 1717-1727.
Price, S., McManus, J., & Barrett, J. (2019). The transgender population: improving awareness for gynaecologists and their role in the provision of care. *The Obstetrician & Gynaecologist, 21*(1), 11-20.
Raginel, T., Grandazzi, G., Launoy, G., Trocmé, M., Christophe, V., Berchi, C., & Guittet, L. (2020). Social inequalities in cervical cancer screening: a discrete choice experiment among French general practitioners and gynaecologists. *BMC Health Services Research, 20*(1), 1-10.
Ragland, M., Trivers, K. F., Andrilla, C. H. A., Matthews, B., Miller, J., Lishner, D., . . . Baldwin, L.-M. (2018). Physician nonprofessional cancer experience and ovarian cancer screening practices: results from a national survey of primary care physicians. *Journal of Women's Health, 27*(11), 1335-1341.
Ramezani Doroh, V., Delavari, A., Yaseri, M., Sefiddashti, S. E., & Akbarisari, A. (2019). Preferences of Iranian average risk population for colorectal cancer screening tests. *International Journal of Health Care Quality Assurance*.
Ran, T., Cheng, C.-Y., Misselwitz, B., Brenner, H., Ubels, J., & Schlander, M. (2019). Cost-effectiveness of colorectal cancer screening strategies—a systematic review. *Clinical Gastroenterology and Hepatology, 17*(10), 1969-1981. e1915.

Robinson, A., Spencer, A., & Moffatt, P. (2015). A framework for estimating health state utility values within a discrete choice experiment: modeling risky choices. *Medical Decision Making, 35*(3), 341-350.
Robinson, K. M., Christensen, K. B., Ottesen, B., & Krasnik, A. (2012). Diagnostic delay, quality of life and patient satisfaction among women diagnosed with endometrial or ovarian cancer: a nationwide Danish study. *Quality of Life Research, 21*(9), 1519-1525.
Rolfe, J., Windle, J., & Bennett, J. (2015). Benefit transfer: insights from choice experiments. In *Benefit Transfer of Environmental and Resource Values* (pp. 191-208): Springer.
Rose, J., & Hensher, D. A. (2004). Modelling agent interdependency in group decision making. *Transportation Research Part E: Logistics and Transportation Review, 40*(1), 63-79.
Rose, J. M., & Bliemer, M. C. (2009). Constructing efficient stated choice experimental designs. *Transport Reviews, 29*(5), 587-617.
Rose, J. M., & Bliemer, M. C. (2013). Sample size requirements for stated choice experiments. *Transportation, 40*(5), 1021-1041.
Rosenberger, R. S., Peterson, G. L., Clarke, A., & Brown, T. C. (2003). Measuring dispositions for lexicographic preferences of environmental goods: integrating economics, psychology and ethics. *Ecological Economics, 44*(1), 63-76.
Round, T., Gildea, C., Ashworth, M., & Møller, H. (2020). Association between use of urgent suspected cancer referral and mortality and stage at diagnosis: a 5-year national cohort study. *British Journal of General Practice, 70*(695), e389. doi:10.3399/bjgp20X709433
Rozbroj, T., Haas, R., O'Connor, D., Carter, S. M., McCaffery, K., Thomas, R., . . . Buchbinder, R. (2021). How do people understand overtesting and overdiagnosis? Systematic review and meta-synthesis of qualitative research. *Social science & medicine, 285*, 114255.
Rupert, D. J., Poehlman, J. A., Hayes, J. J., Ray, S. E., & Moultrie, R. R. (2017). Virtual versus in-person focus groups: Comparison of costs, recruitment, and participant logistics. *Journal of Medical Internet Research, 19*(3), e6980.
Russell, M. (2005). Does Patient ethnicity or Sonographer gender have any bearing on Patient acceptability of transvaginal ultrasound? *Ultrasound, 13*(3), 170-172.
Ryan, M. (2004). Discrete choice experiments in health care. In: British Medical Journal Publishing Group.
Ryan, M., & Bate, A. (2001). Testing the assumptions of rationality, continuity and symmetry when applying discrete choice experiments in health care. *Applied Economics Letters, 8*(1), 59-63.
Ryan, M., Bate, A., Eastmond, C. J., & Ludbrook, A. (2001). Use of discrete choice experiments to elicit preferences. *BMJ Quality & Safety, 10*(suppl 1), i55-i60.
Ryan, M., Gerard, K., & Amaya-Amaya, M. (2008). Discrete choice experiments in a nutshell. In *Using discrete choice experiments to value health and health care* (pp. 13-46): Springer.
Ryan, M., & Skåtun, D. (2004). Modelling non-demanders in choice experiments. *Health economics, 13*(4), 397-402.
Ryan, M., Watson, V., & Entwistle, V. (2009). Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. *Health economics, 18*(3), 321-336.
Ryan, M., & Wordsworth, S. (2000). Sensitivity of willingness to pay estimates to the level of attributes in discrete choice experiments. *Scottish Journal of Political Economy, 47*(5), 504-524.
Salika, T., Lyratzopoulos, G., Whitaker, K. L., Waller, J., & Renzi, C. (2017). Do comorbidities influence help-seeking for cancer alarm symptoms? A population-based survey in England. *Journal of Public Health, 40*(2), 340-349. doi:10.1093/pubmed/fdx072
Salkeld, G., Ryan, M., & Short, L. (2000). The Veil of Experience: Do Consumers Prefer What They Know Best? *Health economics, 9*(3), 267-270. doi:http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291099-1050/issues
Salkeld, G., Solomon, M., Short, L., Ryan, M., & Ward, J. E. (2003). Evidence-based consumer choice: A case study in colorectal cancer screening. *Australian and New Zealand Journal of Public Health, 27*(4), 449-455.
Salsman, J. M., Pavlik, E., Boerner, L. M., & Andrykowski, M. A. (2004). Clinical, demographic, and psychological characteristics of new, asymptomatic participants in a transvaginal ultrasound screening program for ovarian cancer. *Preventive Medicine, 39*(2), 315-322.
Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica, 5*(17), 61-71.
Sasieni, P., Castanon, A., & Cuzick, J. (2009). Effectiveness of cervical screening with age: population based case-control study of prospectively recorded data. *Bmj, 339*, b2968. doi:10.1136/bmj.b2968

Savage, S. J., & Waldman, D. M. (2008). Learning and fatigue during choice experiments: a comparison of online and mail survey modes. *Journal of Applied Econometrics, 23*(3), 351-371.
Scalia, P., Schubbe, D. C., Lu, E. S., Durand, M.-A., Frascara, J., Noel, G., . . . Elwyn, G. (2021). Comparing the impact of an icon array versus a bar graph on preference and understanding of risk information: Results from an online, randomized study. *PLoS One, 16*(7), e0253644.
Scarpa, R., Zanoli, R., Bruschi, V., & Naspetti, S. (2013). Inferred and stated attribute non-attendance in food choice experiments. *American Journal of Agricultural Economics, 95*(1), 165-180.
Schapira, M. M., Walker, C. M., Cappaert, K. J., Ganschow, P. S., Fletcher, K. E., McGinley, E. L., . . . Jacobs, E. A. (2012). The numeracy understanding in medicine instrument: a measure of health numeracy developed using item response theory. *Medical Decision Making, 32*(6), 851-865.
Schwartz, L. M., Woloshin, S., Sox, H. C., Fischhoff, B., & Welch, H. G. (2000). US women's attitudes to false positive mammography results and detection of ductal carcinoma in situ: cross sectional survey. *Bmj, 320*(7250), 1635-1640.
Schwartz, L. M., Woloshin, S., & Welch, H. G. (2005). Can patients interpret health information? An assessment of the medical data interpretation test. *Medical Decision Making, 25*(3), 290-300.
Schwartz, M., Lerman, C., Daly, M., Audrain, J., Masny, A., & Griffith, K. (1995). Utilization of ovarian cancer screening by women at increased risk. *Cancer Epidemiol Biomarkers Prev, 4*(3), 269-273.
Seay, J., Ranck, A., Weiss, R., Salgado, C., Fein, L., & Kobetz, E. (2017). Understanding transgender men's experiences with and preferences for cervical cancer screening: a rapid assessment survey. *LGBT health, 4*(4), 304-309.
Sen, A. (1993). Capability and well-being73. *The quality of life, 30*.
Shahidi, N., & Cheung, W. Y. (2016). Colorectal cancer screening: Opportunities to improve uptake, outcomes, and disparities. *World journal of gastrointestinal endoscopy, 8*(20), 733.
Sharma, A., Beveridge, H., Fallowfield, L., Jacobs, I., & Menon, U. (2006). Postmenopausal women undergoing transvaginal ultrasound screening prefer not to have chaperones. *BJOG: An International Journal of Obstetrics & Gynaecology, 113*(8), 954-957.
Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., . . . Kristjansson, E. (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *bmj, 358*, j4008.
Shen, J. (2009). Latent class model or mixed logit model? A comparison by transport mode choice data. *Applied Economics, 41*(22), 2915-2924.
Shirley, M. H., Barnes, I., Sayeed, S., Finlayson, A., & Ali, R. (2014). Incidence of breast and gynaecological cancers by ethnic group in England, 2001–2007: a descriptive study. *BMC cancer, 14*(1), 1-10.
Sicsic, J., Krucien, N., & Franc, C. (2016). What are GPs' preferences for financial and non-financial incentives in cancer screening? Evidence for breast, cervical, and colorectal cancers. *Social Science and Medicine, 167*, 116-127. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1016/j.socscimed.2016.08.050
Sicsic, J., Pelletier-Fleury, N., & Moumjid, N. (2018). Women's Benefits and Harms Trade-Offs in Breast Cancer Screening: Results from a Discrete-Choice Experiment. *Value in Health, 21*(1), 78-88. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1016/j.jval.2017.07.003
SIGN. (2018). SIGN 135. Management of epithelial ovarian cancer. A national clinical guideline. Retrieved from https://www.sign.ac.uk/media/1073/sign135_oct2018.pdf
Smith-Bindman, R. (2018). Use of advanced imaging tests and the not-so-incidental harms of incidental findings. *JAMA Internal Medicine, 178*(2), 227-228.
Smith, E. M., & Anderson, B. (1985). The effects of symptoms and delay in seeking diagnosis on stage of disease at diagnosis among women with cancers of the ovary. *Cancer, 56*(11), 2727-2732.
Smith, I. P., DiSantostefano, R. L., de Bekker-Grob, E. W., Levitan, B., Berlin, C., Veldwijk, J., & de Wit, G. A. (2021). Methodological Priorities for Patient Preferences Research: Stakeholder Input to the PREFER Public–Private Project. In (Vol. 14, pp. 449-453): Springer.
Smith, S. M., Campbell, N. C., MacLeod, U., Lee, A. J., Raja, A., Wyke, S., . . . Nicolson, M. C. (2009). Factors contributing to the time taken to consult with symptoms of lung cancer: a cross-sectional study. *Thorax, 64*(6), 523-531. doi:10.1136/thx.2008.096560
Snoswell, C. L., Whitty, J. A., Caffery, L. J., Loescher, L. J., Gillespie, N., & Janda, M. (2018). Direct-to-consumer mobile teledermoscopy for skin cancer screening: Preliminary results

demonstrating willingness to pay in Australia. *Journal of Telemedicine and Telecare, 24*(10), 683-689. doi:http://dx.doi.org/10.1177/1357633X18799582
Soekhai, V., de Bekker-Grob, E. W., Ellis, A. R., & Vass, C. M. (2019). Discrete choice experiments in health economics: past, present and future. *PharmacoEconomics, 37*(2), 201-226.
Spinks, J., Janda, M., Soyer, H. P., & Whitty, J. A. (2016). Consumer preferences for teledermoscopy screening to detect melanoma early. *Journal of Telemedicine and Telecare, 22*(1), 39-46. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1177/1357633X15586701
Stevenson, F. A., Cox, K., Britten, N., & Dundar, Y. (2004). A systematic review of the research on communication between patients and health care professionals about medicines: the consequences for concordance. *Health Expect, 7*(3), 235-245. doi:10.1111/j.1369-7625.2004.00281.x
Stewart, T., Lee, Y. A., & Damiano, E. A. (2020). Do transgender and gender diverse individuals receive adequate gynecologic care? An analysis of a rural academic center. *Transgender health, 5*(1), 50-58.
Stone, D. N., & Schkade, D. A. (1991). Numeric and linguistic information representation in multiattribute choice. *Organizational Behavior and Human Decision Processes, 49*(1), 42-59.
Sud, A., Torr, B., Jones, M. E., Broggio, J., Scott, S., Loveday, C., . . . Jhanji, S. (2020). Effect of delays in the 2-week-wait cancer referral pathway during the COVID-19 pandemic on cancer survival in the UK: a modelling study. *The Lancet Oncology, 21*(8), 1035-1044.
Sundar, S., Rick, C., Dowling, F., Au, P., Snell, K., Rai, N., . . . Davenport, C. (2016). Refining Ovarian Cancer Test accuracy Scores (ROCkeTS): protocol for a prospective longitudinal test accuracy study to validate new risk scores in women with symptoms of suspected ovarian cancer. *BMJ open, 6*(8), e010333.
Szrek, H., Chao, L.-W., Ramlagan, S., & Peltzer, K. (2012). Predicting (un) healthy behavior: A comparison of risk-taking propensity measures. *Judgment and decision making, 7*(6), 716.
Tahmasebi, F., Nath, R., Sokolovsky, N., Scaffidi, J., & Boley, J. (2018). Incidental finding of raised CA125: a cause for concern. *Crit Care Obst Gyne, 5*(01), 1-5.
Tamirisa, N. P., Goodwin, J. S., Kandalam, A., Linder, S. K., Weller, S., Turrubiate, S., . . . Riall, T. S. (2017). Patient and physician views of shared decision making in cancer. *Health Expectations, 20*(6), 1248-1253. doi:https://doi.org/10.1111/hex.12564
Tang, Y.-T., & Chooi, W.-T. (2021). A systematic review of the effects of positive versus negative framing on cancer treatment decision making. *Psychology & health*, 1-26.
Target Ovarian Cancer. (2016). *Pathfinder 2016. Transforming futures for women with ovarian cancer*. Retrieved from https://targetovariancancer.org.uk/sites/default/files/2020-07/Pathfinder%202016.pdf
Target Ovarian Cancer. (2019). Smear tests won't detect ovarian cancer. Retrieved from https://targetovariancancer.org.uk/news/smear-tests-wont-detect-ovarian-cancer
Target Ovarian Cancer. (2022a). Campaigns on awareness. Retrieved from https://targetovariancancer.org.uk/get-involved/campaign/current-campaigns/campaign-awareness
Target Ovarian Cancer. (2022b). Campaigns on diagnosis. Retrieved from https://targetovariancancer.org.uk/get-involved/campaign/current-campaigns/campaign-awareness
Tariq, S., & Woodman, J. (2013). Using mixed methods in health research. *JRSM short reports, 4*(6), 2042533313479197.
Tervonen, T., Schmidt-Ott, T., Marsh, K., Bridges, J. F., Quaife, M., & Janssen, E. (2018). Assessing rationality in discrete choice experiments in health: an investigation into the use of dominance tests. *Value in health, 21*(10), 1192-1197.
Teti, M., Kerr, S., Bauerband, L., Koegler, E., & Graves, R. (2021). A Qualitative Scoping Review of Transgender and Gender Non-conforming People's Physical Healthcare Experiences and Needs. *Frontiers in Public Health, 9*.
Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review, 34*(4), 273.
Timmis, L. (2020). *Stakeholder perspectives on optimal follow-up care: An interdisciplinary economic and psychological investigation into gynaecological cancer.* (Doctor of Philosophy), Bangor University,
Tokuda, Y., Chinen, K., Obara, H., & Joishy, S. K. (2009). Intervals between symptom onset and clinical presentation in cancer patients. *Intern Med, 48*(11), 899-905. doi:10.2169/internalmedicine.48.1720
Torrance, G. W. (1976). Social preferences for health states: an empirical evaluation of three measurement techniques. *Socio-economic planning sciences, 10*(3), 129-136.

Torrance, G. W., Feeny, D., & Furlong, W. (2001). Visual analog scales: do they have a role in the measurement of preferences for health states? In: Sage Publications Sage CA: Thousand Oaks, CA.
Train, K. E. (2009). *Discrete choice methods with simulation*: Cambridge university press.
Trevena, L. J., Bonner, C., Okan, Y., Peters, E., Gaissmaier, W., Han, P. K., . . . Zikmund-Fisher, B. J. (2021). Current challenges when using numbers in patient decision aids: advanced concepts. *Medical Decision Making, 41*(7), 834-847.
Trevena, L. J., BPsych, H. M. D., Barratt, A., Butow, P., & Caldwell, P. (2006). A systematic review on communicating with patients about evidence. *Journal of evaluation in clinical practice, 12*(1), 13-23.
Tversky, A., & Shafir, E. (1992). Choice under conflict: The dynamics of deferred decision. *Psychological science, 3*(6), 358-361.
Ubel, P. A., Loewenstein, G., & Jepson, C. (2003). Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Quality of life Research, 12*(6), 599-607.
UK National Screening Committee. (2015). Criteria for appraising the viability, effectiveness and appropriateness of a screening programme. Retrieved from https://www.gov.uk/government/publications/evidence-review-criteria-national-screening-programmes/criteria-for-appraising-the-viability-effectiveness-and-appropriateness-of-a-screening-programme
van Dam, L., Hol, L., Bekker-Grob, E. W. d., Steyerberg, E. W., Kuipers, E. J., Habbema, J. D. F., . . . Leerdam, M. E. v. (2010). What determines individuals' preferences for colorectal cancer screening programmes? A discrete choice experiment. *European Journal of Cancer, 46*(1), 150-159. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1016/j.ejca.2009.07.014
van Nagell Jr, J. R., DePriest, P. D., Ueland, F. R., DeSimone, C. P., Cooper, A. L., McDonald, J. M., . . . Kryscio, R. J. (2007). Ovarian cancer screening with annual transvaginal sonography: findings of 25,000 women screened. *Cancer, 109*(9), 1887-1896.
van Nagell Jr, J. R., Miller, R. W., DeSimone, C. P., Ueland, F. R., Podzielinski, I., Goodrich, S. T., . . . Pavlik, E. J. (2011). Long-term survival of women with epithelial ovarian cancer detected by ultrasonographic screening. *Obstetrics & Gynecology, 118*(6), 1212-1221.
Van Osch, S., & Stiggelbout, A. (2007). The development of the health-risk attitude scale. *Van Osch, SMC. The Construction of Health State Utilities*, 117-138.
Vass, C., Rigby, D., Tate, K., Stewart, A., & Payne, K. (2018). An exploratory application of eye-tracking methods in a discrete choice experiment. *Medical Decision Making, 38*(6), 658-672.
Vass, C. M., Davison, N. J., Vander Stichele, G., & Payne, K. (2020). A picture is worth a thousand words: the role of survey training materials in stated-preference studies. *The Patient-Patient-Centered Outcomes Research, 13*(2), 163-173.
Vass, C. M., & Payne, K. (2017). Using discrete choice experiments to inform the benefit-risk assessment of medicines: are we ready yet? *PharmacoEconomics, 35*(9), 859-866.
Vass, C. M., Rigby, D., & Payne, K. (2018). Investigating the Heterogeneity in Women's Preferences for Breast Screening: Does the Communication of Risk Matter? *Value in Health, 21*(2), 219-228. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1016/j.jval.2017.07.010
Vass, C. M., Wright, S., Burton, M., & Payne, K. (2018). Scale heterogeneity in healthcare discrete choice experiments: a primer. *The Patient-Patient-Centered Outcomes Research, 11*(2), 167-173.
Veldwijk, J., Lambooij, M. S., de Bekker-Grob, E. W., Smit, H. A., & de Wit, G. A. (2014). The effect of including an opt-out option in discrete choice experiments. *PLoS ONE, 9*(11), e111805.
Viswanathan, M., & Childers, T. L. (1996). Processing of numerical and verbal product information. *Journal of Consumer Psychology, 5*(4), 359-385.
Waddell, A., Lennox, A., Spassova, G., & Bragge, P. (2021). Barriers and facilitators to shared decision-making in hospitals from policy to practice: a systematic review. *Implementation Science, 16*(1), 74. doi:10.1186/s13012-021-01142-y
Waller, J., Douglas, E., Whitaker, K. L., & Wardle, J. (2013). Women's responses to information about overdiagnosis in the UK breast cancer screening programme: a qualitative study. *BMJ open, 3*(4), e002703.
Wang, J., Andrae, B., Sundström, K., Ploner, A., Ström, P., Elfström, K. M., . . . Sparén, P. (2017). Effectiveness of cervical screening after age 60 years according to screening history: Nationwide cohort study in Sweden. *PLoS medicine, 14*(10), e1002414. doi:10.1371/journal.pmed.1002414

Webb, E. J., Meads, D., Lynch, Y., Judge, S., Randall, N., Goldbart, J., . . . Murray, J. (2021). Attribute selection for a discrete choice experiment incorporating a best-worst scaling survey. *Value in health, 24*(4), 575-584.
Weber, E. U. (2010). Risk attitude and preference. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(1), 79-88.
Weber, E. U., Blais, A. R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of behavioral decision making, 15*(4), 263-290.
Weinstein, M. C., Siegel, J. E., Gold, M. R., Kamlet, M. S., & Russell, L. B. (1996). Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *Jama, 276*(15), 1253-1258.
Westwood, M., Ramaekers, B., Lang, S., Grimm, S., Deshpande, S., de Kock, S., . . . Kleijnen, J. (2018). Risk scores to guide referral decisions for people with suspected ovarian cancer in secondary care: a systematic review and cost-effectiveness analysis.
Whitaker, K. L., Ghanouni, A., Zhou, Y., Lyratzopoulos, G., & Morris, S. (2017). Patients' preferences for GP consultation for perceived cancer risk in primary care: a discrete choice experiment. *British Journal of General Practice, 67*(659), e388-e395.
Whitty, J. A., & Oliveira Gonçalves, A. S. (2018). A systematic review comparing the acceptability, validity and concordance of discrete choice experiments and best–worst scaling for eliciting preferences in healthcare. *The Patient-Patient-Centered Outcomes Research, 11*(3), 301-317.
Whitty, J. A., Walker, R., Golenko, X., & Ratcliffe, J. (2014). A think aloud study comparing the validity and acceptability of discrete choice and best worst scaling methods. *PLoS One, 9*(4), e90635. doi:10.1371/journal.pone.0090635
Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*: sage publications.
Wilson, J. M. G., & Jungner, G. (1968). Principles and practice of screening for disease.
Woller-Carter, M. M., Okan, Y., Cokely, E. T., & Garcia-Retamero, R. (2012). *Communicating and distorting risks with graphs: An eye-tracking study.* Paper presented at the Proceedings of the human factors and ergonomics society annual meeting.
Woloshin, S., & Schwartz, L. M. (2011). Communicating data about the benefits and harms of treatment: a randomized trial. *Annals of internal medicine, 155*(2), 87-96.
Woof, V. G., Ruane, H., French, D. P., Ulph, F., Qureshi, N., Khan, N., . . . Donnelly, L. S. (2020). The introduction of risk stratified screening into the NHS breast screening Programme: views from British-Pakistani women. *BMC cancer, 20*(1), 452. doi:10.1186/s12885-020-06959-2
World Ovarian Cancer Coalition. (2018). *The Every Women Study. Summary Report*. Retrieved from https://worldovariancancercoalition.org/wp-content/uploads/2019/11/WOCC-Every-Woman-Study-Summary-Report-Nov-08.pdf
Wortley, S., Wong, G., Kieu, A., & Howard, K. (2014). Assessing stated preferences for colorectal cancer screening: A critical systematic review of discrete choice experiments. *Patient, 7*(3), 271-282. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1007/s40271-014-0054-3
Xiao, J., Wang, F., Wang, M., & Ma, Z. (2022). Attribute nonattendance in COVID-19 vaccine choice: A discrete choice experiment based on Chinese public preference. *Health Expectations*.
Yegoryan, N., Guhl, D., & Klapper, D. (2020). Inferring attribute non-attendance using eye tracking in choice-based conjoint analysis. *Journal of Business Research, 111*, 290-304.
Yepes-Nuñez, J. J., Zhang, Y., Xie, F., Alonso-Coello, P., Selva, A., Schünemann, H., & Guyatt, G. (2017). Forty-two systematic reviews generated 23 items for assessing the risk of bias in values and preferences' studies. *Journal of clinical epidemiology, 85*, 21-31.
Zhang, D. C., Highhouse, S., & Nye, C. D. (2019). Development and validation of the general risk propensity scale (GRiPS). *Journal of behavioral decision making, 32*(2), 152-167.
Zipkin, D. A., Umscheid, C. A., Keating, N. L., Allen, E., Aung, K., Beyth, R., . . . Korenstein, D. (2014). Evidence-based risk communication: a systematic review. *Annals of Internal Medicine, 161*(4), 270-280.
Drescher, C., Holt, S. K., Andersen, M. R., Anderson, G., & Urban, N. (2000). Reported ovarian cancer screening among a population-based sample in Washington State. *Obstetrics & Gynecology, 96*(1), 70-74.
Drescher, C. W., Nelson, J., Peacock, S., Andersen, M. R., McIntosh, M. W., & Urban, N. (2004). Compliance of average-and intermediate-risk women to semiannual ovarian cancer screening. *Cancer Epidemiology and Prevention Biomarkers, 13*(4), 600-606.
Jenkins, V., Fallowfield, L., Langridge, C., Barrett, J., Ryan, A., Jacobs, I., . . . Farewell, V. (2015). Psychosocial factors associated with withdrawal from the United Kingdom collaborative trial of

ovarian Cancer screening after 1 episode of repeat screening. *International Journal of Gynecologic Cancer, 25*(8).
Pavlik, E. J., van Nagell Jr, J. R., DePriest, P. D., Wheeler, L., Tatman, J. M., Boone, M., . . . Kryscio, R. K. (1995). Participation in transvaginal ovarian cancer screening: compliance, correlation factors, and costs. *Gynecologic oncology, 57*(3), 395-400.
Schwartz, M., Lerman, C., Daly, M., Audrain, J., Masny, A., & Griffith, K. (1995). Utilization of ovarian cancer screening by women at increased risk. *Cancer Epidemiol Biomarkers Prev, 4*(3), 269-273.

14 Appendices

Appendix 1.1: Ethical approval certificate for best-worst scaling studies

UEBS
University of Exeter Business School
The University of Exeter
Streatham Court
Rennes Drive
Exeter
EX4 4PU

+44 (0)1392 722523
Web: www.exeter.ac.uk

UEBS Research Ethics Committee

Dear Rebekah Hall

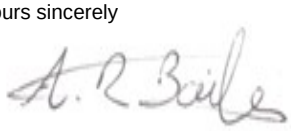
Ethics application - eUEBS003725
An application of Best-Worst Scaling to explore women's preferences towards diagnostic testing

Your project has been reviewed by the UEBS Research Ethics Committee and has received a Favourable opinion.

- Please view your application at https://eethics.exeter.ac.uk/UEBS/ to see comments in full.

If you have received a Favourable with Conditions, Provisional or Unfavourable outcome you are required to re-submit for full review and/or confirm that committee comments have been addressed before you begin your research.
If you have any further queries, please contact your Ethics Officer.

Yours sincerely



Dr. Adrian R. Bailey

Date: 25/08/2020
UEBS Research Ethics Committee

Appendix 1.2: Ethical approval certificate for DCE studies including pilot studies

UNIVERSITY OF EXETER

COLLEGE OF MEDICINE AND HEALTH

College of Medicine and Health
Research Ethics Committee

Certificate of Ethical Approval

Research Institute/Centre: Institute of Health Research

Title of Project: General public preferences for ovarian cancer testing: a discretechoice experiment

Name(s) of Project Research Team member(s): Rebekah Hall, Professor AnneSpencer, Professor Antonieta Medina-Lara and Professor Willie Hamilton

Project Contact Point: Rebekah Hall

This project has been approved for the periodFrom:

12 October 2020

To: 31 May 2022

College of Medicine and Health
Research Ethics Committee approval reference: Oct20/B/261

Signature:

Date: 12 October 2020

Name of Chair Mark Tarrant, PhD

Your attention is drawn of the attached paper "Guidance for Researchers when Ethics Committeeapproval is given", which reminds the researcher of information that needs to be observed when Ethics Committee approval is given.

Application Reference Number: 20/09/261

Appendix 3.1: Relationships between respondent characteristics and acceptance and uptake of ovarian cancer screening

Demographic factors	
Age	No relationship (Drescher *et al.*, 2004) Younger age associated with increased screening (Drescher *et al.*, 2000)
Income	No relationship (Pavlik *et al.*, 1995) Increased income associated with increased screening (Drescher *et al.*, 2000)
Education	No relationship (Drescher *et al.*, 2004; Jenkins *et al.*, 2015) Higher levels of education associated with increased screening (Pavlik *et al.*, 1995)
Employment	Being employed associated with increased screening (Schwartz *et al.*, 1995)
Ethnicity	No relationship (Jenkins *et al.*, 2015)
Health related factors	
Family history of cancer	increased screening behaviour (Drescher *et al.*, 2000; Schwartz *et al.*, 1995) No relationship (Drescher *et al.*, 2004; Jenkins *et al.*, 2015)
Use of HRT	Using HRT associated with reduced screening (Jenkins *et al.*, 2015)
Participation in other screening programmes	No association (Drescher *et al.*, 2004) Previous screening behaviour associated with increased screening (Drescher *et al.*, 2000; Jenkins *et al.*, 2015)
Personal history of cancer	No association (Drescher *et al.*, 2004; Jenkins *et al.*, 2015)
Previous abnormal screening result	No association (Drescher *et al.*, 2004)
Supply-side factors	
Recent contact with a medical professional	Contact with a gynaecologist increased the likelihood of screening (Drescher *et al.*, 2000)
Number of physician per 100,000 patients	No relationship (Pavlik *et al.*, 1995)
Health beliefs and knowledge	
Ovarian cancer worry	Increased worry associated with increased screening (Drescher *et al.*, 2000; Schwartz *et al.*, 1995)
Worried about screening	No relationship (Jenkins *et al.*, 2015)
Perceived risk of ovarian cancer	No relationship (Drescher *et al.*, 2004) Higher perceived risk associated with increased screening (Jenkins *et al.*, 2015; Schwartz *et al.*, 1995)
Ovarian cancer knowledge	No relationship (Drescher *et al.*, 2004)
Increased anxiety	Increased anxiety associated with reduced screening (Jenkins *et al.*, 2015)

Appendix 4.1: Systematic review paper focusing on attributes and attribute selection published in The Patient

The Patient - Patient-Centered Outcomes Research (2022) 15:269–285
https://doi.org/10.1007/s40271-021-00559-3

SYSTEMATIC REVIEW

Attributes Used for Cancer Screening Discrete Choice Experiments: A Systematic Review

Rebekah Hall[1] · Antonieta Medina-Lara[1] · Willie Hamilton[1] · Anne E. Spencer[1]

Accepted: 26 September 2021 / Published online: 21 October 2021

Abstract

Background Evidence from discrete choice experiments can be used to enrich understanding of preferences, inform the (re) design of screening programmes and/or improve communication within public campaigns about the benefits and harms of screening. However, reviews of screening discrete choice experiments highlight significant discrepancies between stated choices and real choices, particularly regarding willingness to undergo cancer screening. The identification and selection of attributes and associated levels is a fundamental component of designing a discrete choice experiment. Misspecification or misinterpretation of attributes may lead to non-compensatory behaviours, attribute non-attendance and responses that lack external validity.

Objectives We aimed to synthesise evidence on attribute development, alongside an in-depth review of included attributes and methodological challenges, to provide a resource for researchers undertaking future studies in cancer screening.

Methods A systematic review was conducted to identify discrete choice experiments estimating preferences towards cancer screening, dated between 1990 and December 2020. Data were synthesised narratively. In-depth analysis of attributes led to classification into four categories: test specific, service delivery, outcomes and monetary. Attribute significance and relative importance were also analysed. The International Society for Pharmacoeconomics and Outcomes Research conjoint analysis checklist was used to assess the quality of reporting.

Results Forty-nine studies were included at full text. They covered a range of cancer sites: over half (26/49) examined colorectal screening. Most studies elicited general public preferences (34/49). In total, 280 attributes were included, 90% (252/280) of which were significant. Overall, test sensitivity and mortality reduction were most frequently found to be the most important to respondents.

Conclusions Improvements in reporting the identification, selection and construction of attributes used within cancer screening discrete choice experiments are needed. This review also highlights the importance of considering the complexity of choice tasks when considering risk information or compound attributes. Patient and public involvement and stakeholder engagement are recommended to optimise understanding of unavoidably complex choice tasks throughout the design process. To ensure quality and maximise comparability across studies, further research is needed to develop a risk-of-bias measure for discrete choice experiments.

Appendix 4.2: Quality assessment for systematic reviews included in the overview using the adapted AMSTAR2

Questions 1 and 8 were excluded as they were not relevant to systematic reviews of discrete experiments. The maximum score was 8. The AMSTAR checklist was supplemented by an additional question which assessed whether the systematic reviews performed a risk of bias assessment for the studies which it included.

	Phillips et al, 2006	Marshall et al, 2010	Ghanouni et al, 2013	Wortley et al, 2014	Mansfield et al,, 2016
1. Was the research question clearly defined? (i.e. Were a description and motivation for the research question, target population, key outcomes provided)	✓	✓	✓	✓	✓
2. Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?	✗	✗	✗	✓	✗
3. Did the review authors explain their selection of the study designs for inclusion in the review?	✓	✓	✓	✓	✓
4. Did the review authors use a comprehensive literature search strategy?	✗	✓	✓	✓	✓
5. Did the review authors perform study selection in duplicate?	✓	**?**	**?**	✓	**?**
6. Did the review authors perform data extraction in duplicate?	✓	**?**	**?**	✓	**?**
7. Did the review authors provide a list of excluded studies and justify the exclusions?	✗	✗	✗	✓	✗
8. Did the review authors describe the included studies in adequate detail? (populations, outcomes only)	✓	✓	✓	✓	✓
9. **Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?**	✗	✗	✓	✓	✗
10. Did the review authors report on the sources of funding for the studies included in the review?	✓	✓	✓	✓	✓
11. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	✓	✓	✓	✓	✓
Total AMSTAR2 score (out of 11)	7	6	7	11	6

Q1. Ammended- wording unchanged but answers amended -original question uses PICO format which is not relevant/necessary to preference-based studies– instead we focused on the motivation of the study and description of the target population, type of preference study and key outcomes
Q8. Ammended- wording unchanged but answers amended-we focused on descriptions of the population and key outcomes only
Q9. Ammended- wording unchanged but answers amended- we focused on whether the quality of studies had been checked using a checklist/tool
Q11-15- Excluded- no quantitative/meta-analysis was not performed in any of the reviews

Appendix 4.3: Search terms used to identify discrete choice studies

Database: MEDLINE; Strategy:
1. "discrete choice".tw.
2. (conjoint adj1 (analys?s or measurement* or stud* or "choice experiment*")).tw.
3. "part-worth utilit*".tw.
4. "functional measurement*".tw.
5. "paired comparison*".tw.
6. "pairwise choice*".tw.
7. (stated adj1 (preference* or choice*)).tw.
8. or/1-7
9. exp animals/ not humans.sh.
10. 8 not 9
11. limit 16 to yr="2000 -Current"

Database: MEDLINE(R) In-Process & Other Non-Indexed Citation; Strategy:
1. "discrete choice".tw.
2. (conjoint adj1 (analys?s or measurement* or stud* or "choice experiment*")).tw.
3. "part-worth utilit*".tw.
4. "functional measurement*".tw.
5. "paired comparison*".tw.
6. "pairwise choice*".tw.
7. (stated adj1 (preference* or choice*)).tw.
8. or/1-7
9. limit 17 to yr =”2000 –Current”

Database: EMBASE; Strategy:
1. "discrete choice".tw.
2. (conjoint adj1 (analys?s or measurement* or stud* or "choice experiment*")).tw.
3. "part-worth utilit*".tw.
4. "functional measurement*".tw.
5. "paired comparison*".tw.
6. "pairwise choice*".tw.
7. (stated adj1 (preference* or choice*)).tw.
8. or/1-7
9. exp animal/ not human/
10. 8 not 9
11. limit 17 to yr="2000 -Current"

Database: PsycINFO; Strategy:
1. "discrete choice".tw.
2. (conjoint adj1 (analys?s or measurement* or stud* or "choice experiment*")).tw.
3. "part-worth utilit*".tw.
4. "functional measurement*".tw.
5. "paired comparison*".tw.
6. "pairwise choice*".tw.
7. (stated adj1 (preference* or choice*)).tw.
8. or/1-7
9. limit 17 to yr="2000 -Current"

Database: HMIC (Health Management Information Consortium); Strategy:
1. "discrete choice".tw.
2. (conjoint adj1 (analys?s or measurement* or stud* or "choice experiment*")).tw.
3. "part-worth utilit*".tw.
4. "functional measurement*".tw.
5. "paired comparison*".tw.
6. "pairwise choice*".tw.
7. (stated adj1 (preference* or choice*)).tw.
8. or/1-7
9. limit 17 to yr="2000 -Current"

Database: Web of Science; Strategy:
1. TITLE: ("discrete choice")
2. TITLE: (conjoint near/0 (analys?s or measurement* or stud* or "choice experiment*"))
3. TITLE: (part-worth utilit*)
4. TITLE: (functional measurement*)
5. TITLE: (paired comparison*)
6. TITLE: (pairwise choice*)
7. TITLE: (stated near/0 (preference* or choice*))
8. #7 OR #6 OR #5 OR #4 OR #3 OR #2 OR #1

Database: EconLit; Strategy:
1. TI "discrete choice" OR AB "discrete choice"
2. TI (conjoint N0 (analys?s or measurement* or stud* or "choice experiment*") OR AB (conjoint N0 (analys?s or measurement* or stud* or "choice experiment*")
3. TI "part-worth utilit*" OR AB "part-worth utilit*"
4. TI "functional measurement*" OR AB "functional measurement*"
5. TI "paired comparison*" OR AB "paired comparison*"
6. TI "pairwise choice*" OR AB "pairwise choice*"
7. TI (stated N0 (preference* or choice*)) OR AB (stated N0 (preference* or choice*))
8. S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7

Database: NHS EED; Strategy:
1. "discrete choice":ti or "discrete choice":ab
2. conjoint near/1 (analys?s or measurement* or stud* or "choice experiment*"):ti or conjoint near/1 (analys?s or measurement* or stud* or "choice experiment*"):ab
3. "part-worth utilit*":ti or "part-worth utilit*":ab
4. "functional measurement*":ti or "functional measurement*":ab
5. "paired comparison*":ti or "paired comparison*":ab
6. "pairwise choice*":ti or "pairwise choice*":ab
7. stated near/1 (preference* or choice*):ti or stated near/1 (preference* or choice*):ab
8. #1 or #2 or #3 or #4 or #5 or #6 or #7 in Economic Evaluations

Appendix 4.4: ISPOR Good Practice Guidance Checklist (Bridges et al, 2011)

1. Was a well-defined research question stated and is conjoint analysis an appropriate method for answering it?
1.1 Were a well-defined research question and a testable hypothesis articulated? 1.2 Was the study perspective described, and was the study placed in a particular decision-making or policy context? 1.3 What is the rationale for using conjoint analysis to answer the research question?
2. Was the choice of attributes and levels supported by evidence?
2.1 Was attribute identification supported by evidence (literature reviews, focus groups, or other scientific methods)? 2.2 Was attribute selection justified and consistent with theory? 2.3 Was level selection for each attribute justified by the evidence and consistent with the study perspective and hypothesis?
3. Was the construction of tasks appropriate?
3.1 Was the number of attributes in each conjoint task justified (that is, full or partial profile)? 3.2 Was the number of profiles in each conjoint task justified? 3.3 Was (should) an opt-out or a status-quo alternative (be) included?
4. Was the choice of experimental design justified and evaluated?
4.1 Was the choice of experimental design justified? Were alternative experimental designs considered? 4.2 Were the properties of the experimental design evaluated? 4.3 Was the number of conjoint tasks included in the data-collection instrument appropriate?
5. Were preferences elicited appropriately, given the research question?
5.1 Was there sufficient motivation and explanation of conjoint tasks? 5.2 Was an appropriate elicitation format (that is, rating, ranking, or choice) used? Did (should) the elicitation format allow for indifference? 5.3 In addition to preference elicitation, did the conjoint tasks include other qualifying questions (for example, strength of preference, confidence in response, and other methods)?
6. Was the data collection instrument designed appropriately?
6.1 Was appropriate respondent information collected (such as sociodemographic, attitudinal, health history or status, and treatment experience)? 6.2 Were the attributes and levels defined, and was any contextual information provided? 6.3 Was the level of burden of the data-collection instrument appropriate? Were respondents encouraged and motivated?
7. Was the data-collection plan appropriate?
7.1 Was the sampling strategy justified (for example, sample size, stratification, and recruitment)? 7.2 Was the mode of administration justified and appropriate (for example, face-to-face, pen-and-paper, web-based)? 7.3 Were ethical considerations addressed (for example, recruitment, information and/or consent, compensation)?
8. Were statistical analyses and model estimations appropriate?
8.1 Were respondent characteristics examined and tested? 8.2 Was the quality of the responses examined (for example, rationality, validity, reliability)? 8.3 Was model estimation conducted appropriately? Were issues of clustering and subgroups handled appropriately?
9. Were the results and conclusions valid?
9.1 Did study results reflect testable hypotheses and account for statistical uncertainty? 9.2 Were study conclusions supported by the evidence and compared with existing findings in the literature? 9.3 Were study limitations and generalizability adequately discussed?
10. Was the study presentation clear, concise, and complete?
10.1 Was study importance and research context adequately motivated? 10.2 Were the study data-collection instrument and methods described? 10.3 Were the study implications clearly stated and understandable to a wide audience?
Highlighted questions were particularly relevant to the research question

Appendix 4.5: Results of quality assessment using the ISPOR Good Practice Guidance Checklist (Bridges et al, 2011)

	Q1.	Q2.	Q3.	Q4.	Q5.	Q6.	Q7.	Q8.	Q9.	Q10.	Total score
Ryan and Wordsworth (2000)	3	3	3	2	0	2	2	3	3	2	23
Salkeld, et al. (2000)	3	2	1	1	0	0	0	1	2	2	12
Gerard, et al. (2003)	3	3	3	3	0	1	2	3	3	2	23
Salkeld, et al. (2003)	2	3	1	3	2	2	3	2	1	3	22
Arana, et al. (2006)	3	3	1	1	2	2	1	1	1	3	18
Berchi, et al. (2006)	2	3	2	2	0	2	1	2	3	2	19
Marshall, et al. (2007)	3	3	3	2	1	3	1	3	3	3	25
Fiebig, et al. (2009)	3	3	3	3	0	2	2	2	2	2	22
Howard and Salkeld (2009)	3	3	3	3	0	2	2	3	3	2	24
Marshall, et al. (2009)	3	3	3	3	1	3	3	3	3	3	28
Hol, et al. (2010)	3	3	3	2	1	3	2	3	3	3	26
Nayaradou, et al. (2010)	2	3	3	2	0	2	0	2	3	2	19
van Dam, et al. (2010)	3	3	3	2	2	3	2	3	3	3	27
Howard, et al. (2011)	3	3	2	3	1	3	3	3	3	3	27
Pignone, et al. (2012)	3	3	1	2	3	3	1	2	2	3	23
Boone, et al. (2013)	3	2	3	2	2	3	3	2	3	3	26
de Bekker-Grob, et al. (2013)	3	2	3	2	2	3	3	3	3	3	27
Johar, et al. (2013)	3	3	3	3	0	2	2	2	2	2	22
Pignone, et al. (2013)	3	3	2	1	2	2	3	1	3	3	23
Benning, et al. (2014a)	3	3	2	2	1	3	0	3	3	3	23
Benning, et al. (2014b)	3	3	3	2	1	3	0	2	3	3	23
Brenner, et al. (2014)	3	3	2	1	2	2	2	1	2	3	21
Ghanouni, et al. (2014)	3	2	3	3	2	3	3	1	3	3	26
Groothuis-Oudshoorn, et al. (2014)	3	3	3	2	1	3	3	2	3	3	26
Pignone, et al. (2014)	2	3	3	2	2	3	3	2	3	3	26
Plumb, et al. (2014)	2	2	3	1	1	3	3	2	3	3	23
Chamot, et al. (2015)	2	3	2	1	1	3	2	1	3	3	21
Howard, et al. (2015)	3	3	3	2	0	3	2	2	3	2	23
Kistler, et al. (2015)	2	3	3	1	2	3	3	2	3	3	25
Kitchener, et al. (2016)	3	3	3	3	2	3	3	1	3	3	27
Martens, et al. (2016)	3	3	3	2	0	2	1	1	3	2	20
Sicsic, et al. (2016)	3	3	3	3	1	2	3	1	3	2	24
Spinks, et al. (2016)	2	3	1	3	0	1	1	0	3	2	16
Kohler, et al. (2017)	3	3	2	2	2	3	2	2	3	3	25
Papin-Lefebvre, et al. (2017)	2	3	2	2	0	2	3	1	3	2	20
Ellimoottil, et al. (2018)	2	3	1	1	0	2	1	1	3	2	16
Mansfield, et al. (2018)	3	3	2	3	1	3	2	1	3	3	24
Osborne, et al. (2018)	2	1	0	2	2	1	1	1	3	2	15
Sicsic, et al. (2018)	3	3	3	1	1	3	3	1	3	3	24
Snoswell, et al. (2018)	2	3	3	3	0	2	3	0	3	2	21
Vass, et al. (2018)	3	3	3	3	2	3	2	3	3	3	28
Li, et al. (2019)	2	3	3	1	1	2	3	2	3	2	22
Mandrik, et al. (2019)	2	3	3	3	1	2	3	2	3	2	24
Miles, et al. (2019)	3	2	2	0	2	3	3	3	3	3	21
Oberlin, et al. (2019)	2	3	1	1	1	1	2	2	2	2	17
Ramezani Doroh, et al. (2019)	3	3	0	1	1	1	1	1	3	2	16
Bilger, et al. (2020)	3	3	2	2	2	3	2	3	3	3	26
Charvin, et al. (2020)	3	3	3	2	1	3	2	3	3	3	26
de Bekker-Grob, et al. (2020)	3	3	3	3	1	3	3	3	3	3	28
Hendrix, et al. (2020)	3	3	3	3	0	2	3	2	3	2	24
Peters et al, (2020)	3	3	3	2	2	3	3	3	3	3	28
Raginel et al. (2020)	3	3	2	2	2	3	2	1	3	3	24
Average score	2.7	2.8	2.4	2.1	2.0	2.4	2.1	1.9	2.8	2.6	23.8

Appendix *4.*6: Summary data for DCE studies included in the systematic review

Author (date)	Country	Cancer site	Intervention(s)	Population(s) (response rate)	Method of administration	Attribute Selection process	Choice tasks
Ryan and Wordsworth (2000)	UK	Cervical	Multiple- Pap smear, Liquid-based cytology	641 general public- women mixed screening experience, 18-65yrs (32%)	Self-completed postal questionnaire	Policy questions	Per respondent: 6 or 7 Experimental design: 25 Blocks: 4
Salkeld, et al. (2000)	Australia	Colorectal	Single-FOBT	336 patients- previously screened, >50yrs (56%)	Self-completed postal questionnaire	Focus groups with previously screened individuals	Per respondent: 16 Experimental design: 16 Blocks: 1
Gerard, et al. (2003)	Australia	Breast	Single- Standard breast screening, rescreening rates	87 patients (48%)	Self-completed postal questionnaire	Literature review	Per respondent: 16 Experimental design: 32 Blocks: 2
Salkeld, et al. (2003)	Australia	Colorectal	Single-FOBT	301 general public- mixed experience of screening, 50-70yrs (73%)	Face-to-face interviews	Focus groups and ratings survey with general public	Per respondent: 18 Experimental design: 16 Blocks: 2
Arana, et al. (2006)	Spain	Cervical	Single- standard cervical screening	480 general public- Female students, mixed screening experience (NS), 467 HCPs (NS)	Face-to-face interviews with self-completed questionnaire	Trial data and focus groups	Per respondent: 8 Experimental design: 16 Blocks: 2
Berchi, et al. (2006)	France	Colorectal	Single-FOBT	294 HCPs (42.6%)	Self-completed postal questionnaire	Literature review, interviews with experts	Per respondent: 3 Experimental design: 9 Blocks: 3
Marshall, et al. (2007)	Canada	Colorectal	Multiple- Colonoscopy, FOBT, CT colonoscopy, Sigmoidoscopy, barium enema	547 general public- mixed experience of screening, 40-60yrs (52%)	Self-completed postal questionnaire	Literature review, expert opinion, focus groups	Per respondent: 12 Experimental design: 40 Blocks: 4

Author (date)	Country	Cancer site	Intervention(s)	Population(s) (response rate)	Method of administration	Attribute Selection process	Choice tasks
Fiebig, et al. (2009)	Australia	Cervical	Multiple-Pap smear, Liquid-based cytology, HPV testing, Automated screening	167 general public- women, mixed screening experience, 18-69yrs (NS), 215 HCPs (NS)	Self-completed postal questionnaire	Literature review, current policy	Per respondent: 32 Experimental design: 512 Blocks: 16
Howard and Salkeld (2009)	Australia	Colorectal	Single-FOBT	1157 patients-previously screened (60%)	Self-completed postal questionnaire	Literature review	Per respondent: 18 Experimental design: 256 Blocks: 16
Marshall, et al. (2009)	USA/Canada	Colorectal	Multiple-CT, Barium enema, Colonography, FOBT	1599 general public- mixed screening experience, 47-70yrs (NS), 100 HCPs (NS)	Self-completed online questionnaire	Literature review, focus groups with patients, previous DCE, expert opinion	Per respondent: 12 Experimental design: 33 Blocks: 3
Hol, et al. (2010)	Netherlands	Colorectal	Multiple- FOBT, Sigmoidoscopy, Colonoscopy	489 screening-naïve individuals (33%), 549 previously screening individuals, 50-75yrs (32%)	Self-completed postal questionnaire	Literature review, expert opinion, interviews with patients due to be screened	Per respondent: 12 Experimental design: 84 Blocks: 7
Nayaradou, et al. (2010)	France	Colorectal	Multiple-FOBT, Blood sample	656 general public- mixed screening experience, 50-74yrs (33%)	Self-completed postal questionnaire	Literature review, interviews with experts	Per respondent: 3-4 Experimental design: 14 Blocks: 4
van Dam, et al. (2010)	Netherlands	Colorectal	Multiple- FOBT, FIT, Sigmoidoscopy	156 Screening naïve individuals (31%) 124 Screening participants (59%), 50-75yrs	Self-completed postal questionnaire	Literature review, expert opinion, interviews with target populations	Per respondent: 16 Experimental design: 16 Blocks: 1
Howard, et al. (2011)	Australia	Colorectal	Multiple- CT colonoscopy, Colonoscopy	130 patients with experience of both tests (84.4%)	Self-completed postal questionnaire or telephone interview with nurse	Literature review, interviews with patients, ranking exercise with patients and doctors	Per respondent: 33 Experimental design: 32 Blocks: 1
Pignone, et al. (2012)	USA	Colorectal	Multiple- FOBT, Sigmoidoscopy, Colonoscopy, CT colonography	50 general public- mixed screening experience, 48-75yrs (47%)	Self-completed postal questionnaire	Literature review and previous work by authors	Per respondent: 16 Experimental design: 15 Blocks:1
Boone, et al. (2013)	UK	Colorectal	Single- CT colonography	75 general public-mixed screening experience, >55yrs (67%) 50 HCPs (81%)	Face-to-face interviews (Patients), self-completed online questionnaire (HCPs)	None	Per respondent: 10 Experimental design: 10 Blocks: 1

Author (date)	Country	Cancer site	Intervention(s)	Population(s) (response rate)	Method of administration	Attribute Selection process	Choice tasks
de Bekker-Grob, et al. (2013)	Netherlands	Prostate	Single- PSA screening	459 general public- males, mixed screening experience, 55-75yrs (46%)	Self-completed postal questionnaire	Literature review, interviews with experts and eligible screening candidates	Per respondent: 16 Experimental design: 32 Blocks: 2
Johar, et al. (2013)	Australia	Cervical	Multiple- standard test, liquid based test	295 previously screened women, 18-69yrs (NS)	Self-completed postal questionnaire	Previous DCE	Per respondent: 32 Experimental design: 512 Blocks: 16
Pignone, et al. (2013)	USA and Australia	Prostate	Single- PSA screening	302 general public- males, mixed screening experience, 50-70yrs (85%)	Self-completed online questionnaire	Literature review, previous work by group	Per respondent: 16 Experimental design: 16 Blocks: 1
Benning, et al. (2014a)	Netherlands	Colorectal	Multiple- Blood test, FOBT	1571 general public- mixed screening experience, 55-75yrs (60.8%)	Self-completed online questionnaire	Literature review	Per respondent: 13 Experimental design: NS Blocks: 6
Benning, et al. (2014b)	Netherlands	Colorectal	Multiple- Blood test, FOBT	631general public-mixed screening experience, 55-77yrs (56.5%)	Self-completed online questionnaire	Literature review	Per respondent: 13 Experimental design: NS Blocks: 3
Brenner, et al. (2014)	USA and Australia	Colorectal	Multiple- FOBT, Sigmoidoscopy, Colonoscopy, CT colonography	306 general public- mixed screening experience, 50-75yrs (86.3%)	Self-completed online questionnaire	Literature review, previous work by authors	Per respondent: 16 Experimental design: 16 Blocks: 1
Ghanouni, et al. (2014)	UK	Colorectal	Single- CT colonography	607 general public- mixed screening experience, 45-54yrs (77%)	Self-completed online questionnaire	None- chosen based on research question	Per respondent: 3-4 Experimental design: 18 Blocks: 6
Groothuis-Oudshoorn, et al. (2014)	UK and Netherlands	Colorectal	Multiple- Nanopill, FOBT, sigmoidoscopy	1356 general public- mixed screening experience, 50-74yrs (61%)	Self-completed online questionnaire	Literature review	Per respondent: 14 Experimental design: 13,986 Blocks:999

Author (date)	Country	Cancer site	Intervention(s)	Population(s) (response rate)	Method of administration	Attribute Selection process	Choice tasks
Pignone, et al. (2014)	USA	Colorectal	Multiple- FOBT, Sigmoidoscopy, Colonoscopy, CT colonography	150 general public- mixed screening experience, 50-75yrs (85.7%)	Paper questionnaire completed in-person during focus-group session	Literature review, expert opinion	Per respondent: 16 Experimental design: 16 Blocks: 1
Plumb, et al. (2014)	UK	Colorectal	Single- CT colonography	52 general public- mixed screening experience, 55-69yrs (24.8%), 50 HCPs (52.1%)	Face-to-face interviews for patients, interviews or online self-completed questionnaire- HCPs	None- chosen based on research question	Per respondent: 23 Experimental design: 23 Blocks: 1
Chamot, et al. (2015)	Zambia	Cervical	Multiple- Visual inspection, urine sample, swab by woman, swab by professional	208 patients- women, recently screened (87.4%)	Face-to-face computer-aided interviews	Literature review, focus groups with women and screening personnel, expert opinion	Per respondent: 10 Experimental design: 512 Blocks: Random draws
Howard, et al. (2015)	Australia	Prostate	Single- PSA screening	602 general public- males, mixed screening experience, 40-69yrs (83.4%)	Self-completed online questionnaire	Literature review, expert opinion, interviews with men	Per respondent: 15 Experimental design: NS Blocks: NS
Kistler, et al. (2015)	USA	Colorectal	Multiple- FOBT, Sigmoidoscopy, Colonoscopy, CT colonography	116 general public- mixed experience 70-90yrs (55.2%)	Face-to-face computer-aided interviews	Literature review, interviews with target population	Per respondent: 10 Experimental design: NS Blocks: 15
Kitchener, et al. (2016)	UK	Cervical	Multiple- Self-sampling kits, Nurse-led screening, timed appointment	222 general public- screening naïve women (5.5%)	Self-completed questionnaire- online or paper	Interviews with women from general public, literature review, assumptions of authors	Per respondent: 12 Experimental design: 12 Blocks: 1
Martens, et al. (2016)	USA	Colorectal	Multiple- FOBT, Sigmoidoscopy, Colonoscopy, CT colonography	38 general public- mixed screening experience, Spanish speaking, 50-75yrs (NS)	Paper questionnaire completed in-person during focus-group session	Literature review, expert opinion	Per respondent: 16 Experimental design: 16 Blocks:1

Author (date)	Country	Cancer site	Intervention(s)	Population(s) (response rate)	Method of administration	Attribute Selection process	Choice tasks
Sicsic, et al. (2016)	France	Colorectal, Breast, Cervical	Single- FOBT	HCPs- Total- 333 GPs Colorectal- 114 (85%) Breast- 108 (80%) Cervical- 111 (83%)	Self-completed online questionnaire	Literature review, interviews with target populations	Per respondent: 12 Experimental design: 12 Blocks: 1
Spinks, et al. (2016)	Australia	Melanoma	Multiple- Teledermoscopy, self-examination, Examination by professional	35 patients- previously screened, high risk, 50-64yrs (70%)	Self-completed questionnaire- online or paper	Literature review, previous surveys with target population	Per respondent: 12 Experimental design: NS Blocks: NS
Kohler, et al. (2017)	Malawi	Breast	Multiple- Self-examination, clinical breast exam	213 general public- women, mixed knowledge of screening and breast cancer (NS)	Face-to-face tablet-based interviews	Literature review, interviews with local women	Per respondent: 9 Experimental design: NS Blocks: NS
Papin-Lefebvre, et al. (2017)	France	Colorectal	Single- FOBT	326 GPs (17%)	Self-completed postal questionnaire	Existing literature	Per respondent: 5-6 Experimental design: 12 Blocks: 2
Ellimoottil, et al. (2018)	USA	Prostate	Multiple- MRI transrectal ultrasound guided prostate biopsy, Transrectal ultrasound-guided prostate biopsy	146 patients- males from an outpatient urology clinic, 55-70yrs (NS)	Self-completed questionnaire completed on tablet	Literature review, interviews and ranking exercise with patients	Per respondent: NS Experimental design: NS Blocks: NS
Mansfield, et al. (2018)	USA	Colorectal	Multiple- FOBT, Sigmoidoscopy, Colonoscopy, FIT	2067 General public- mixed experience, 50-75yrs (64%)	Self-completed online questionnaire	Literature review, ranking exercise with experts	Per respondent: 5 Experimental design: NS Blocks: 16
Osborne, et al. (2018)	France	Breast	Single- Mammogram	812 general public- women, mixed screening experience, 40-74yrs (74%)	Self-completed online questionnaire	Literature review, focus groups and interviews with women	Per respondent: 8 Experimental design: 16 Blocks: 2
Sicsic, et al. (2018)	UK	Breast	Single- Mammogram	1018 general public- women, mixed screening experience, 18-70yrs (9%)	Self-completed online questionnaire	Literature review, expert opinion, interviews with general public and patient representative	Per respondent: 11 Experimental design: 44 Blocks: 4

Author (date)	Country	Cancer site	Intervention(s)	Population(s) (response rate)	Method of administration	Attribute Selection process	Choice tasks
Snoswell, et al. (2018)	Australia	Skin	Multiple- Teledermoscopy, self-examination, examination by professional	113 patients with experience of tests (trial participants) (~50%)	Self-completed online questionnaire	Previous DCE, trial data	Per respondent: 12 Experimental design: 24 Blocks: 2
Vass, et al. (2018)	Australia	Colorectal	Multiple- FOBT, FIT, Blood, Saliva	1282 general public- mixed experience 50-74yrs (42.7%)	Self-completed online/ postal questionnaire	Unclear- ranking exercise	Per respondent: 9 Experimental design: 27 Blocks: 3
Li, et al. (2019)	South Africa	Cervical	Single- HPV-based screening	298 patients- women attending clinic, 18yrs+ (NS)	In person-computer assisted	Qualitative work with women visiting clinic	Per respondent: 10 Experimental design: NS Blocks: Random draws
Mandrik, et al. (2019)	China	Cervical	Single- existing test	405 general public- women, mixed screening experience, 30-65yrs (NS)	Face-to-face interviews	Literature review, interviews with women, expert opinion	Per respondent: 10 Experimental design: 27 Blocks: 3
Miles, et al. (2019)	UK	Lung, colorectal	Multiple- Whole-body MRI, conventional staging (e.g. combined CT, ultrasound, PET)	132 Patients- lung/colorectal cancer as part of trial (experience of WB-MRI and conventional staging methods) (52%)	Self-completed postal questionnaire	Clinical trial	Per respondent: 9 Experimental design: 18 Blocks: 2
Oberlin, et al. (2019)	Belarus	Breast	Multiple- mammogram, physical examination only, mammogram plus physical examination	428 general public- women, mixed screening experience, 50-69yrs (89%)	Face-to-face interviews	Literature review, interviews with women and HCPs	Per respondent: 18 Experimental design: 72 Blocks: 4
Ramezani Doroh, et al. (2019)	Iran	Colorectal	Multiple- FOBT, FIT, Fecal DNA test, Sigmoidoscopy, Colonoscopy, Barium Enema	500 patients referred to teaching hospital (reason for referral unspecified), 50-75yrs (NS)	Not specified	Literature review, expert opinion	Per respondent: 7 or 8 Experimental design: 36 Blocks: 5
Bilger, et al. (2020)	Singapore	Breast, Cervical	Single-mammogram/pap smear	Breast- 400 general public- women, mixed screening experience, 40-65yrs (NS) Cervical- 401 general public women, mixed screening experience, 25-65yrs (NS)	Face-to-face interview using iPad	Literature review, focus groups and interviews with women	Per respondent-10 Experimental design- 72 Blocks-9

Author (date)	Country	Cancer site	Intervention(s)	Population(s) (response rate)	Method of administration	Attribute Selection process	Choice tasks
Charvin, et al. (2020)	France	Prostate	Single- PSA with digital rectal exam	1023 general public men, 50-75yrs (NS)	Self-completed online questionnaire	Literature review, expert opinion	Per respondent-7 Experimental design- 7 Blocks: 2 (experimental arms only)
de Bekker-Grob, et al. (2020)	Netherlands	Colorectal	Single-FIT	406 general public, mixed screening experience. 50-75yrs (NS)	Self-completed online questionnaire	Literature review, expert opinion, focus groups/group interviews with men	Per respondent-16 Experimental design- NS Blocks-10
Hendrix, et al. (2020)	USA	Breast	Single- AI-enhanced mammogram	91 HCPs-primary care providers (6%)	Self-completed online questionnaire	Literature review, interviews with HCPs	Per respondent-15 Experimental design-15 Blocks-1
Peters et al, (2020)	Netherlands	Oesophagus	Multiple- upper endoscopy, transnasal endoscopy, pill on a string, breath test, blood test	554 general public, 50-74yrs (36.9%)	Self-completed postal questionnaire	Literature review, expert opinion, focus groups with public	Per respondent:15 Experimental design: 130 Blocks:10
Raginel et al. (2020)	France	Cervical	Multiple- pap smear, self-administered swab	123 HCPs- GPs and Gynaecologists (15.6% postal, 2.8% email)	Self-completed online or postal questionnaire	Literature review, stakeholder interviews	Per respondent: 11 Experimental design: 11 Blocks: 1

Appendix 5.1: Attribute analysis of diagnostic DCEs identified during attribute identification stage (n=50)

	Total number of attributes*	Aim/purpose of test	Test procedure	Accuracy	Sensitivity	Specificity	Risk of inconclusive result	Timeliness/accuracy	Side effects	Access to treatment	Appointment/test duration	Appointment availability	Decision-making responsibility	Follow-up testing	Location	Notification of negative result	Supportive services	Privacy/anonymity	Results delivery method	Timing of test/results (gestation)	Type of HCP	HCP characteristics	Waiting time-appointment	Waiting time- results	Cost	Potential benefit of diagnosis	Occurrence rate of condition	Severity of condition	Other
Antenatal testing																													
Barrett et al, 2017	4	●		●					●											●									
Beulen et al, 2015	7	●			●	●			●											●				●	●				
Bishop et al, 2004	3				●				●			●																	
Carroll et al, 2013	4				●															●				●	●				
Chan et al, 2009	3	●																						●	●				
Hill et al, 2012	4	●		●					●											●									
Hill et al, 2014	3			●					●											●									
Hill et al, 2016	4	●		●					●											●									
Hill et al, 2017	3			●					●											⊙									
Lewis et al, 2006a	3				●				●											●									
Lewis et al, 2006b	3				●				●											●									
Lund et al, 2018	4	●		●					●											●									
Miller et al, 2015	5	●			●															●						●			●
Ryan et al, 2005	3	●																						●	●				
Wright et al, 2017	4												●				⊙			⊙					●				
Wright et al, 2018	4												●				⊙			⊙					●				
Genetic testing																													
Blumenschein et al 2016	5	●			●																				●	●		●	
Buchanan et al, 2016	6			●	●												⊙				●			●	●				
Davidson et al, 2019	5		●		○		○																	○	●				
Hall et al, 2006	1 2	○			●										●									○	● ●		● ●	⊙	○ ○ ●
Kilambi et al, 2015	4				●													●							●		●		
Knight et al, 2014	4				●													●							●		●		
Marshall et al, 2019	6		⊙		●																			●	●	● ●			
Najafzadeh et al, 2013	7		⊙		⊙	⊙			● ⊙															⊙	●				
Payne et al, 2011	5		○	●													●				●			●					
Peyron et al, 2018	5					⊙						●	⊙				⊙								⊙				
Plothner et al, 2018	5	⊙		⊙													⊙								⊙		⊙		
Regier et al, 2009	3				●																			●	●				
Severin et al, 2013	6	○								○															●	●	●	●	
Severin et al, 2015	5	⊙																							⊙	⊙	⊙	⊙	
Veldwijk et al, 2016	4												⊙													⊙	⊙	●	
Weymann et al, 2018	4		○		●																			●	●				
Sexual health testing																													
D'Elbée et al, 2018	4													⊙		⊙							●		●				
Eaton et al, 2019	6		⊙		●					⊙								⊙		●				●					
Indravuldh et al, 2017	10		○								○			⊙		⊙ ○	○			⊙	○ ○				●				
Johnson et al, 2010	6		⊙					⊙						⊙		⊙	⊙								⊙				
Llewellyn et al, 2013	6	●													●			●		●			○	●					
Miners et al, 2012	6	●													●			●		●			●	●					
Ostermann et al, 2015	7		⊙							⊙	⊙			⊙			⊙			⊙	⊙ ⊙								
Pan et al, 2019	7		●								●			●			● ●			●					●				
Phillips et al, 2002	6		●					● c						⊙		○	● c								●				
Strauss et al, 2018a	7		⊙								⊙	○		○		⊙				⊙					⊙				
Strauss et al, 2018b	6		○									⊙		⊙		●				○					⊙				
Wijnen et al, 2019	5		⊙							●	●			⊙			⊙												
Zanolini et al, 2018	4		⊙											○		⊙									●				
Other																													
Bryan et al, 1998	4												●												○	●			●
Muhlbacher et al,2016	3		●			●																			●				
Radley et al, 2015	5													● ●								●		●	●				
Whitty et al, 2015	5		●						● c				●											●	●				
Wijnen et al, 2014	6		● ●									●	● ●	●															

Significance of attributes: ●= significant at 95% confidence level, ○= not significant at 95% confidence level, ⊙= mixed evidence or unclear
Type of attributes: ● contains risk information, ● contains time information, ●contains risk and time information, ●contains no risk or time information
* Some attributes spanned multiple categories

Appendix 5.2: Best-worst scaling questionnaire

What features of tests for ovarian cancer are most important to women?

You are being invited to take part in a research study that aims to understand attitudes around testing for possible ovarian cancer. Before you decide whether to take part in this study it is important for you to understand why the research is being done and what taking part will involve.

Please read the following information carefully and take time to consider whether you wish to take part.

What is the aim of the research?
We want to understand what characteristics of diagnostic testing for possible ovarian cancer are most important to women.

Who can participate?
We are interested in exploring the views of females over the age of 40. You must be able to complete the survey in English to take part.

What do I have to do?
Complete a survey that will last about 35 minutes. There are no other commitments associated with participating.

What type of information will I be asked?
The survey includes questions about your background, health history and your views on ovarian cancer testing. There are no right or wrong answers- the purpose of the survey is to better understand the opinion of women.

Is the survey confidential?
All your answers to the survey are completely confidential. You will not be asked your name or any other identifying information. Only those directly involved in the research will have access to your data. Your information will be stored and managed according to a law called the Data Protection Act (2018).

What happens to the results of the survey?
This survey forms part of a larger study which aims to understand preferences around testing for cancer. It is hoped the results will help improve the delivery of cancer diagnostic services. The results of this study may be published but your personal information or responses will not be identifiable.

Who is funding the research?
The project is part of a PhD project at the University of Exeter. The research is funded by Cancer Research UK.

If you have any other questions or require more information about this study, please contact Prof Anne Spencer at A.E.Spencer@exeter.ac.uk or Rebekah Hall at rh591@exeter.ac.uk

Consent Form

*I have read the information provided on the previous page and wish to take part in the study.

I grant permission for the information collected during this survey to be used in the researcher's publications on this topic.

- ○ I agree
- ○ I disagree

*Please enter your Prolific ID:

Section 1: Background questions

In this section, we are interested in getting to know you a little better by asking a few background questions.

***What is your age?**

Which of the following best describes your current relationship status?

- ○ Married
- ○ Widowed
- ○ Divorced
- ○ Separated
- ○ In a domestic partnership or civil union
- ○ Single, but cohabiting with a significant other
- ○ Single, never married
- ○ I prefer not to say

How many children do you have?

What is your ethnicity?

- ○ White/Caucasian
- ○ Mixed-White and Black African
- ○ Mixed- White and Black Caribbean
- ○ Mixed- White and Asian
- ○ Asian- Indian
- ○ Asian- Pakistani
- ○ Asian- Bangladeshi
- ○ Asian-Chinese
- ○ Black-African
- ○ Black- Caribbean
- ○ Arab
- ○ I prefer not to say
- ○ Other:

Which of the following categories best describes your employment status?

- ○ Employed, working full-time
- ○ Employed, working part-time
- ○ Self-employed or freelance
- ○ Not employed, looking for work
- ○ Not employed, NOT looking for work
- ○ Not employed, NOT looking for work
- ○ Retired
- ○ Long-term sick or disabled, not able to work
- ○ Student
- ○ Volunteering
- ○ Looking after home or relative
- ○ Prefer not to say

What is your highest level of education?

- ○ 1-4 O levels/CSEs/GCSEs (any grades), Entry level, Foundation Diploma
- ○ NVQ Level 1, Foundation GNVQ, Basic Skills
- ○ 5+ O levels (passes)/ CSEs (grade 1)/ GCSEs (grades A*-C), School Certificate, 1 A level/ 2-3 AS levels/VCEs, Higher Diploma
- ○ NVQ Level 2, Intermediate GNVQ, City and Guilds Craft, BTEC First/General Diploma, RSA Diploma
- ○ Apprenticeship
- ○ 2+ A levels/VCEs, 4+ AS levels, Higher School Certificate, Progression/Advanced Diploma
- ○ NVQ Level 3, Advanced GNVQ, City and Guilds Advanced Craft, ONC, OND, BTEC National, RSA Advanced Diploma
- ○ Undergraduate degree (BA, BSc)
- ○ Postgraduate degree (e.g. MSc, MA, PhD)
- ○ NVQ Level 4-5, HNC, HND, RSA Higher Diploma, BTEC Higher Level
- ○ Professional qualifications (e.g. teaching, nursing, accounting)
- ○ Other vocational/ work-related qualifications
- ○ Foreign qualifications
- ○ No qualifications
- ○ I prefer not to say

To what extent do you agree with the following statements:

	Totally agree	2	3	Neither agree or disagree	5	6	Totally disagree
I think I take good care of my body	○	○	○	○	○	○	○
I don't want to have to consider the consequences for my health in everything that I do	○	○	○	○	○	○	○
It is important to me that I organise my life so that I will later enjoy good health	○	○	○	○	○	○	○
If it concerns my health, then I see myself as someone who avoids risks	○	○	○	○	○	○	○
Uncertainty about the consequences of a medical intervention is, in general, part of the game	○	○	○	○	○	○	○
My health means everything to me	○	○	○	○	○	○	○
When I look back at my past, I think that, in general, I did take risks with my health	○	○	○	○	○	○	○
If my doctor cannot offer me certainty about the possible consequences of a medical intervention, then I would rather not undergo it	○	○	○	○	○	○	○
We want to make sure you are paying attention. Please select 'Totally disagree' for this row.	○	○	○	○	○	○	○
Safety first, where my health is concerned	○	○	○	○	○	○	○
To enjoy good health now and in the future, I am prepared to forgo a lot	○	○	○	○	○	○	○
People say that I take risks with my health because of my habits	○	○	○	○	○	○	○
I am not very fussy about my health	○	○	○	○	○	○	○
In general I would estimate that I would not have much of a problem with undergoing a high risk operation	○	○	○	○	○	○	○

Section 2- Your views on ovarian cancer testing

In the next set of questions, we want to learn more about what women like you think about testing for ovarian cancer.

In each question you will be shown a list of features based on the tests that might be offered to a woman who is experiencing symptoms that may be caused by ovarian cancer.

For each question, we would like you to select which feature of a medical test you think is most important and which feature is least important if you were given a choice about what test to take. The features will change between questions so please only consider the features listed in each case. Please only choose one feature as most important and one feature as the least important.

During this section we are trying to measure your preferences for 25 different features of diagnostic testing. To do this we need to ask quite a lot of very similar questions. The task may start to feel repetitive but each question is slightly different and asks about different combinations of the features we are interested in hearing your opinions on. Please take your time and try and pay close attention.

There are no right or wrong answers. We are simply interested in your opinion.

To help you get used to this style of question, please complete the example question about choosing a restaurant to eat in below:

Considering only the features below, which is the most important and which is the least important when choosing a restaurant to eat at?

Most important		Least important
○	Distance from home	○
○	Taste of the food	○
○	Price of the food	○
○	Customer service	○
○	Variety of food on the menu	○

For each question, we would like you to select which feature of a medical test you think is most important and which feature is least important if you were given a choice about what test to take. The features will change between questions so please only consider the features listed in each case. Please only choose one feature as most important and one feature as the least important.

You can hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Chance of dying from ovarian cancer	○
○	How test results are returned	○
○	Notification of negative test results	○
○	Test procedure	○
○	Pre-test support	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Staff attitude	○
○	Test location	○
○	How test results are returned	○
○	Gender of health care provider	○
○	Travel time	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Chance of an inconclusive result	○
○	Out of pocket costs	○
○	Chance of dying from ovarian cancer	○
○	Time away from usual activities	○
○	Time to notification of results	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Pain and discomfort	○
○	Time to notification of results	○
○	Information included with the invitation	○
○	Openness of health care providers	○
○	How test results are returned	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Waiting time for the test	○
○	Pre-test support	○
○	Who explains the results	○
○	Test location	○
○	Out of pocket costs	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Number of follow up tests	○
○	Gender of health care provider	○
○	Pre-test support	○
○	Information included with the invitation	○
○	Time away from usual activities	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Out of pocket costs	○
○	Post-test support	○
○	Chance that the test will miss cancer in a patient who actually does have the disease	○
○	How test results are returned	○
○	Chance of diagnosing another condition	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Who explains the results	○
○	Openness of health care providers	○
○	Gender of health care provider	○
○	Chance that the test will miss cancer in a patient who actually does have the disease	○
○	Chance of dying from ovarian cancer	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Gender of health care provider	○
○	Chance of an inconclusive result	○
○	Post-test support	○
○	Choice of appointment time	○
○	Test procedure	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Choice of appointment time	○
○	Chance that the test will miss cancer in a patient who actually does have the disease	○
○	Time to notification of results	○
○	Number of follow up tests	○
○	Test location	○

Hover over each feature for a more detailed description

Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Test duration	○
○	Information included with the invitation	○
○	Test location	○
○	Chance of dying from ovarian cancer	○
○	Post-test support	○

Hover over each feature for a more detailed description

Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Chance of diagnosing another condition	○
○	Information included with the invitation	○
○	Chance of an inconclusive result	○
○	Staff attitude	○
○	Who explains the results	○

Hover over each feature for a more detailed description

Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Pre-test support	○
○	Chance of unnecessary further invasive testing	○
○	Post-test support	○
○	Time to notification of results	○
○	Staff attitude	○

Hover over each feature for a more detailed description

Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Pain and discomfort	○
○	Chance of diagnosing another condition	○
○	Time away from usual activities	○
○	Test procedure	○
○	Test location	○

Hover over each feature for a more detailed description

Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Chance that the test will miss cancer in a patient who actually does have the disease	○
○	Pre-test support	○
○	Travel time	○
○	Chance of an inconclusive result	○
○	Pain and discomfort	○

Hover over each feature for a more detailed description

Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Test duration	○
○	Staff attitude	○
○	Chance that the test will miss cancer in a patient who actually does have the disease	○
○	Notification of negative test results	○
○	Time away from usual activities	○

Hover over each feature for a more detailed description

Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Notification of negative test results	○
○	Out of pocket costs	○
○	Choice of appointment time	○
○	Travel time	○
○	Information included with the invitation	○

Hover over each feature for a more detailed description

Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Who explains the results	○
○	Time away from usual activities	○
○	Chance of unnecessary further invasive testing	○
○	How test results are returned	○
○	Choice of appointment time	○

Hover over each feature for a more detailed description

Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Gender of health care provider	○
○	Notification of negative test results	○
○	Chance of diagnosing another condition	○
○	Time to notification of results	○
○	Waiting time for the test	○

Hover over each feature for a more detailed description

Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Test procedure	○
○	Staff attitude	○
○	Openness of health care providers	○
○	Out of pocket costs	○
○	Number of follow up tests	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Time to notification of results	○
○	Test procedure	○
○	Travel time	○
○	Who explains the results	○
○	Test duration	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Openness of health care providers	○
○	Choice of appointment time	○
○	Test duration	○
○	Pre-test support	○
○	Chance of diagnosing another condition	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Chance of dying from ovarian cancer	○
○	Pain and discomfort	○
○	Staff attitude	○
○	Waiting time for the test	○
○	Choice of appointment time	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Chance of unnecessary further invasive testing	○
○	Test duration	○
○	Out of pocket costs	○
○	Pain and discomfort	○
○	Gender of health care provider	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Travel time	○
○	Chance of dying from ovarian cancer	○
○	Number of follow up tests	○
○	Chance of diagnosing another condition	○
○	Chance of unnecessary further invasive testing	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	How test results are returned	○
○	Number of follow up tests	○
○	Waiting time for the test	○
○	Test duration	○
○	Chance of an inconclusive result	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Test location	○
○	Notification of negative test results	○
○	Chance of an inconclusive result	○
○	Chance of unnecessary further invasive testing	○
○	Openness of health care providers	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Time away from usual activities	○
○	Travel time	○
○	Openness of health care providers	○
○	Post-test support	○
○	Waiting time for the test	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Information included with the invitation	○
○	Waiting time for the test	○
○	Test procedure	○
○	Chance of unnecessary further invasive testing	○
○	Chance that the test will miss cancer in a patient who actually does have the disease	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Pain and discomfort	○
○	Chance of diagnosing another condition	○
○	Time away from usual activities	○
○	Test procedure	○
○	Test location	○

Hover over each feature for a more detailed description

Considering only the features below, which is the most important and which is the least important when choosing a test for possible ovarian cancer?

Most important		Least important
○	Post-test support	○
○	Who explains the results	○
○	Pain and discomfort	○
○	Number of follow up tests	○
○	Notification of negative test results	○

Hover over each feature for a more detailed description

How easy or difficult did you find making choices between the most and least important characteristics in the last section?

- ○ Very easy
- ○ Easy
- ○ Neither easy nor difficult
- ○ Difficult
- ○ Very difficult

During the last section you were shown different combinations of 25 features related to testing for possible ovarian cancer. You have already shown that some features are more important than others to you but were there any features that you would consider completely unimportant?

Please select any features that would never be an important issue to you if you were making a decision about whether to undergo a medical test for possible ovarian cancer.

- [ ] Chance that the test will miss cancer in a patient who actually does have the disease
- [ ] Chance of dying from ovarian cancer
- [ ] Choice of appointment time
- [ ] Who explains the results
- [ ] Pain and discomfort
- [ ] Notification of negative test results
- [ ] Chance of diagnosing another condition
- [ ] Pre-test support
- [ ] Test procedure
- [ ] Staff attitude
- [ ] Post-test support
- [ ] Time away from usual activities
- [ ] Chance of unnecessary further invasive testing
- [ ] Travel time
- [ ] Time to notification of results
- [ ] Openness of health care providers
- [ ] Number of follow up tests
- [ ] Chance of an inconclusive result
- [ ] Out of pocket costs
- [ ] Gender of health care provider
- [ ] How test results are returned
- [ ] Test location
- [ ] Test duration
- [ ] Information included with the invitation
- [ ] Waiting time for the test

During the last section you were shown different combinations of 25 features related to testing for possible ovarian cancer. You have already shown that some features are more important than others to you but were there any features that you would consider completely unimportant?

Please select any features that would always be an important issue to you if you were making a decision about whether to undergo a medical test for possible ovarian cancer.

- [ ] Chance that the test will miss cancer in a patient who actually does have the disease
- [ ] Chance of dying from ovarian cancer
- [ ] Choice of appointment time
- [ ] Who explains the results
- [ ] Pain and discomfort
- [ ] Notification of negative test results
- [ ] Chance of diagnosing another condition
- [ ] Pre-test support
- [ ] Test procedure
- [ ] Staff attitude
- [ ] Post-test support
- [ ] Time away from usual activities
- [ ] Chance of unnecessary further invasive testing
- [ ] Travel time
- [ ] Time to notification of results
- [ ] Openness of health care providers
- [ ] Number of follow up tests
- [ ] Chance of an inconclusive result
- [ ] Out of pocket costs
- [ ] Gender of health care provider
- [ ] How test results are returned
- [ ] Test location
- [ ] Test duration
- [ ] Information included with the invitation
- [ ] Waiting time for the test

These are all the features you were asked about in the last section:

- Waiting time for the test
- Who explains the results
- Chance of diagnosing another condition
- Staff attitude
- Chance of unnecessary further invasive testing
- Openness of health care providers
- Out of pocket costs
- Test location
- Chance that the test will miss cancer in a patient who actually does have the disease
- Chance of dying from ovarian cancer
- Pain and discomfort
- Pre-test support
- Post-test support
- Travel time
- Number of follow up tests
- Gender of health care provider
- Test duration
- Choice of appointment time
- Notification of negative test results
- Test procedure
- Time away from usual activities
- Time to notification of results
- Chance of an inconclusive result
- How test results are returned
- Information included with the invitation

Imagine you were experiencing symptoms of possible ovarian cancer and there are two different tests available. Is there any other information about the different tests available you would like to know about before you decided what test to have?

Please list any thoughts below:

Section 4: A little more about you

In this section we will ask a little more about you and your views around health care.

In general, how would you rate your overall health?

- ○ Very poor
- ○ Poor
- ○ Fair
- ○ Good
- ○ Very good

Compared to the average woman of your age, would you say that you are more or less likely to get ovarian cancer?

- ○ Very high risk
- ○ High risk
- ○ Average risk
- ○ Low risk
- ○ Very low risk

Was this person your:

- ○ Blood relative
- ○ Non-blood relative (e.g. step-relatives, in-law, adoptive relatives)Non-blood relative
- ○ Friend
- ○ Acquaintance/ work colleague
- ○ Other:

Have you every undergone testing for possible ovarian cancer?

- ○ Yes
- ○ No
- ○ I don't know
- ○ I prefer not to say

What test(s) did you have?

ⓘ Check all that apply

- ☐ Blood test
- ☐ Transvaginal ultrasound/internal ultrasound
- ☐ CT scan
- ☐ Laparoscopy

(Also known as keyhole surgery this is a small surgical procedure that allows access inside the abdomen (tummy) and pelvis without having to make large incisions in the skin.)

- ☐ I don't know
- ☐ Other:

Have you ever had a transvaginal ultrasound?

- ○ Yes
- ○ No
- ○ I don't know
- ○ I prefer not to say

ⓘ *A transvaginal ultrasound is a type of pelvic ultrasound used to examine female reproductive organs. This is an internal examination.*

Based on the information above, what is your favourite chocolate bar?
Please select 'KitKat' below to show you have read this question.

- ○ Twix
- ○ Dairy Milk
- ○ Snickers
- ○ KitKat
- ○ Bounty

Have you ever taken contraceptive (birth control) pills?

- ○ Yes
- ○ No
- ○ I don't know
- ○ I prefer not to say

We are interested in whether you have ever taken contraceptive pills at any point in your life. You do not currently have to be taking them to answer 'yes'

How many years in total have you taken contraceptive pills?

- ○ Less than 1 year
- ○ Between 1 and 5 years
- ○ Between 5 and 10 years
- ○ More than 10 years
- ○ I don't remember

Total amount of time taking contraceptive pills across your lifetime

Have you ever taken hormone replacement therapy (HRT) medication?

- ○ Yes
- ○ No
- ○ I don't know
- ○ I prefer not to say

Hormone replacement therapy (HRT) are medications used to ease the symptoms of menopause. They are usually taken in the form of tablets but could also be implants, gels/creams or skin patches.

How many years in total have you taken hormone replacement therapy medication?

- ○ Less than 1 year
- ○ Between 1 and 5 years
- ○ Between 5 and 10 years
- ○ More than 10 years
- ○ I don't remember

Total amount of time taking HRT across your lifetime

Which of the following applies to you?

- ○ I smoke cigarettes every day
- ○ I smoke cigarettes, but not every day
- ○ I do not smoke cigarettes at all, but do smoke tobacco of some kind (e.g. e-cigarettes/vaping, pipe cigars)
- ○ I have stopped smoking completely in the last year
- ○ I stopped smoking completely more than a year ago
- ○ I have never been a smoker (i.e. smoked regularly for a year or more)
- ○ I don't know
- ○ I prefer not to say

Have you ever had a hysterectomy?

- ○ Yes
- ○ No
- ○ I don't know
- ○ I prefer not to say

A hysterectomy is a surgical procedure to remove the womb (uterus)

Have you or blood relative ever been diagnosed with a BRCA1 or BRCA2 mutation?

- ○ Yes-I have been tested and diagnosed with a BRCA1 or BRCA2 mutation
- ○ Yes- An immediate family member has been tested and diagnosed with a BRCA 1 or BRCA2 mutation (child, parent, sibling)
- ○ Yes- An extended family member has been tested and diagnosed with a BRCA 1 or BRCA2 mutation (aunt, uncle, grandparent, cousin)
- ○ No- to my knowledge no one in my family has been tested and diagnosed with a BRCA 1 or BRCA2 mutation
- ○ I don't know
- ○ I prefer not to say

BRCA1 and BRCA2 are genes that help prevent tumours developing. Damage or mutations to these genes increase a persons chance of developing cancer. Mutations are typically inherited from a parent.

Have you ever been diagnosed with endometriosis?

- ○ Yes
- ○ No
- ○ I don't know
- ○ I prefer not to say

Endometriosis is a condition where tissue similar to the lining of the womb starts to grow in other places, such as the ovaries and fallopian tubes.

Section 4: A little more about you

In this section we will ask a little more about you and your views around health care

We would like to see how people's personality traits might be linked to their views on testing for cancer.
Please select an answer for all of the following statements.

I see myself as someone who...

	Disagree strongly	Disagree a little	Neither agree or disagree	Agree a little	Agree strongly
...is reserved	○	○	○	○	○
...is generally trusting	○	○	○	○	○
...tends to be lazy	○	○	○	○	○
...is relaxed, handles stress well	○	○	○	○	○
...has few artistic interests	○	○	○	○	○
...is outgoing, sociable	○	○	○	○	○
...tends to find fault with others	○	○	○	○	○
...does a thorough job	○	○	○	○	○
...gets nervous easily	○	○	○	○	○
...has an active imagination	○	○	○	○	○

How much confidence and trust do you have in general practitioners (GPs)?

- ○ A great deal
- ○ A lot
- ○ A moderate amount
- ○ A little
- ○ None at all

When seeking help for medical issues, how much do you feel able to be involved in decisions about the treatment process?

- ○ A great deal
- ○ A lot
- ○ A moderate amount
- ○ A little
- ○ Not at all

When seeking help for medical issues, how much do you wish to be involved in decisions about the treatment process?

- ○ A great deal
- ○ A lot
- ○ A moderate amount
- ○ A little
- ○ Not at all

In general, how willing are you to take risks?

- ○ 1- unwilling to take risks
- ○ 2
- ○ 3
- ○ 4
- ○ 5
- ○ 6
- ○ 7
- ○ 8
- ○ 9
- ○ 10- fully prepared to take risks

Submit

Thank you for completing this survey.

Important: Please click here to return to prolific and confirm your submission

---

If you have any questions or concerns about the survey please contact Rebekah Hall at rh591@exeter.ac.uk or Prof Anne Spencer at A.E.Spencer@exeter.ac.uk

Postal address: University of Exeter Medical School, Room 1.15, South Cloisters, St Luke's Campus, Magdalen Road, City, Exeter, EX1 2LU

More information about ovarian cancer and the tests available can be found at :

Cancer Research UK: https://www.cancerresearchuk.org/about-cancer/ovarian-cancer

Target ovarian cancer: https://www.targetovariancancer.org.uk

The Eve Appeal: https://eveappeal.org.uk/gynaecological-cancers/ovarian-cancer/

Please speak to your GP if you are concerned about your risk of ovarian cancer.

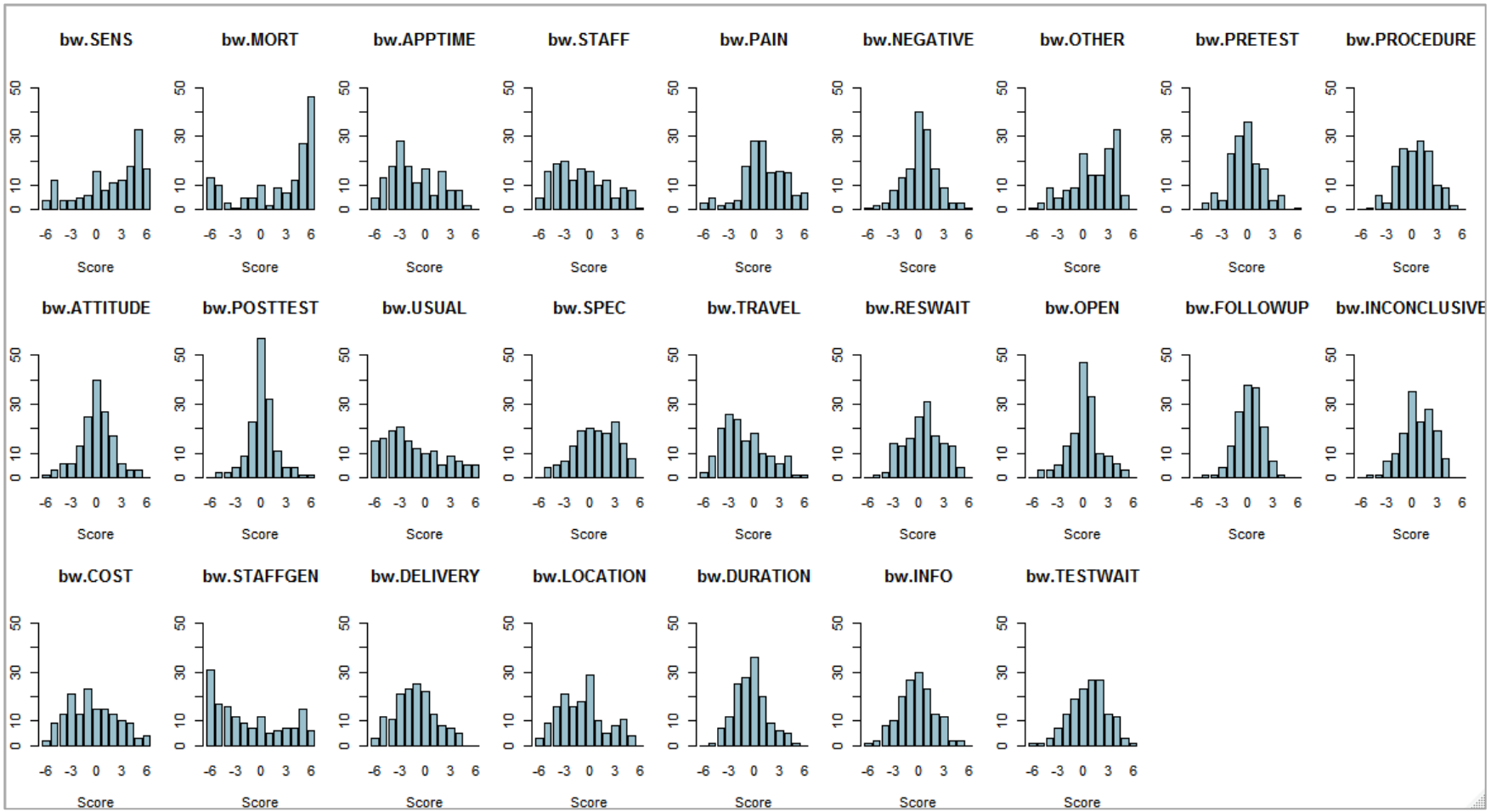
Appendix 5.3: Sociodemographic and health-related characteristics of respondents from the BWS study

Characteristic	n (%)
Age, mean (SD)	51.4 (9.1)
Ethnicity	
White	120 (80.0)
Asian	8 (5.3)
Black	3 (2.0)
Mixed	3 (2.0)
Other	9 (6.0)
Not reported	7 (4.7)
Children, mean (SD)	1.3 (1.3)
Relationship status	
Married	75 (50.0)
In a relationship	22 (14.6)
Single	19 (12.7)
Divorced/separated	26 (17.3)
Widowed	6 (4.0)
Not reported	2 (1.3)
Education	
No qualifications	1 (0.7)
GCSE	37 (24.7)
A-Level/ College	25 (16.7)
Undergraduate	41 (27.3)
Post-graduate	35 (23.3)
Other	9 (6.0)
Not reported	2 (1.3)
Employment	
Employed, full-time	47 (31.1)
Part-time	32 (21.2)
Self-employed	23 (15.2)
Not employed	11 (7.2)
Retired	14 (9.3)
Other	18 (12.0)
Not reported	5 (3.3)

Descriptive characteristics- health-related

Characteristic	n (%)
Self-reported overall health	
Very good	18 (12.0)
Good	61 (40.7)
Fair	47 (31.3)
Poor	14 (9.3)
Very poor	4 (2.7)
Not reported	6 (4.0)
Perceived risk of ovarian cancer	
Very low	12 (8.0)
Low	33 (22.0)
Average	83 (55.3)
High	12 (8.0)
Very high	4 (2.7)
Not reported	6 (4.0)
Ovarian cancer-related worry	
A great deal	4 (2.7)
A lot	25 (16.7)
A moderate amount	43 (28.7)
A little	47 (31.3)
Not at all	26 (17.3)
Not reported	5 (3.3)
Personal history of cancer	17 (13.3)
Knew someone who was diagnosed with ovarian cancer	25 (16.7)
Previously tested for ovarian cancer	40 (26.7)
Previously undergone a TVUS (any reason)	50 (33.3)
Cervical screening	
Always attends/ attended	75 (50.0)
Irregularly attends/attended	37 (24.7)
Never attended/stopped attending	37 (24.7)
Unknown	1 (0.7)
How much confidence and trust in GPs	
A great deal/a lot	59 (39.4)
A moderate amount	53 (35.3)
A little	25 (16.7)
None at all	2 (1.3)
Unknown	11 (7.3)
How much do you feel able to be involved in medical decisions	
A great deal	12 (8.0)
A lot	22 (14.7)
A moderate amount	58 (38.7)
A little	34 (22.7)
Not at all	17 (11.3)
Unknown	7 (4.7)
How much do you wish to be involved in medical decisions	
A great deal	60 (40.0)
A lot	67 (44.7)
A moderate amount	19 (12.7)
A little	3 (2.0)
Not at all	1 (0.7)
Task difficulty	
Very east/easy	63 (42.0)
Neither easy or difficult	30 (20.0)
Very difficult/difficult	54 (38.0)

Appendix 5.4: Frequency of individual attribute scores for each attribute during BWS

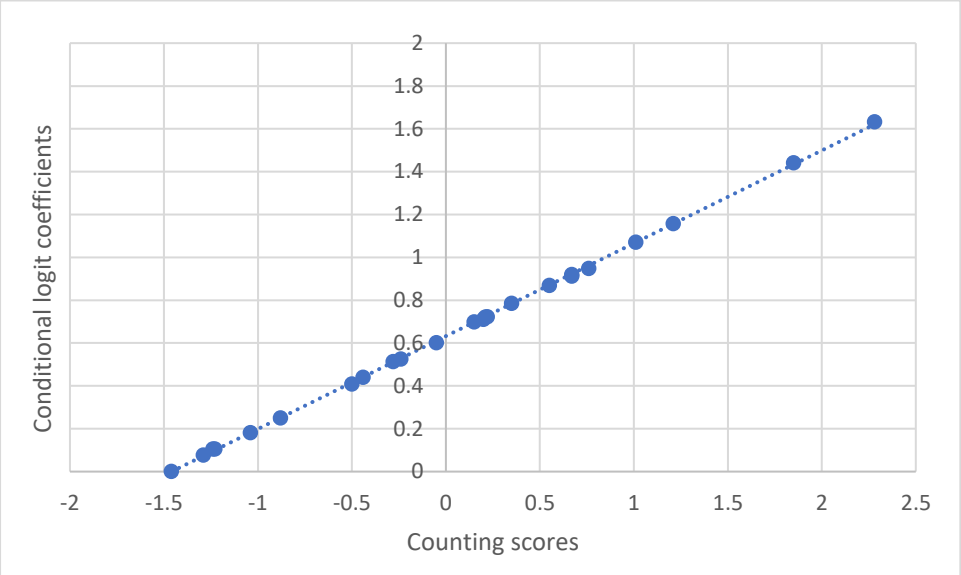


Appendix 5.5: Results of logit model

The importance of each attribute was estimated relative to “time away from usual activities”, which ranked lowest during the counting analysis. All attributes had positive coefficient estimates and most were statistically significant at the 95% level, confirming the relative importance of all attributes compared to “time away from usual activities”. Non-significant attributes were those with the lowest importance, confirming the clustering effect at the bottom of the latent importance scale observed during counting analysis.

Attributes	Conditional logit		Ranking
	Coefficient	SE	
Sensitivity	1.44***	0.096	2
Chance of dying from ovarian cancer	1.63***	0.103	1
Choice of appointment time	0.08	0.085	23
Who explains the results	0.18*	0.087	20
Pain and discomfort	1.07***	0.084	4
Notification of negative test results	0.72***	0.078	11
Chance of diagnosing another condition	1.16***	0.088	3
Pre-test support	0.53***	0.078	15
Test-procedure	0.79***	0.081	9
Staff attitude	0.60***	0.077	14
Post-test support	0.70***	0.076	13
Time away from usual activities	Ref	Ref	25
Specificity	0.95***	0.085	5
Travel time	0.11	0.081	21
Time to notification of test results	0.87***	0.082	8
Openness of healthcare providers	0.72***	0.78	10
Number of follow up tests	0.71***	0.078	12
Chance of an inconclusive result	0.92***	0.077	6
Out-of-pocket costs	0.41***	0.083	18
Gender of healthcare provider	0.001	0.091	24
How test results are returned	0.11	0.082	22
Test location	0.25***	0.083	19
Test duration	0.44***	0.077	17
Information included with the invitation	0.51***	0.081	16
Waiting time for the test	0.91***	0.084	6

Appendix 5.6: Comparison of the conditional logit coefficients and counting scores during BWS

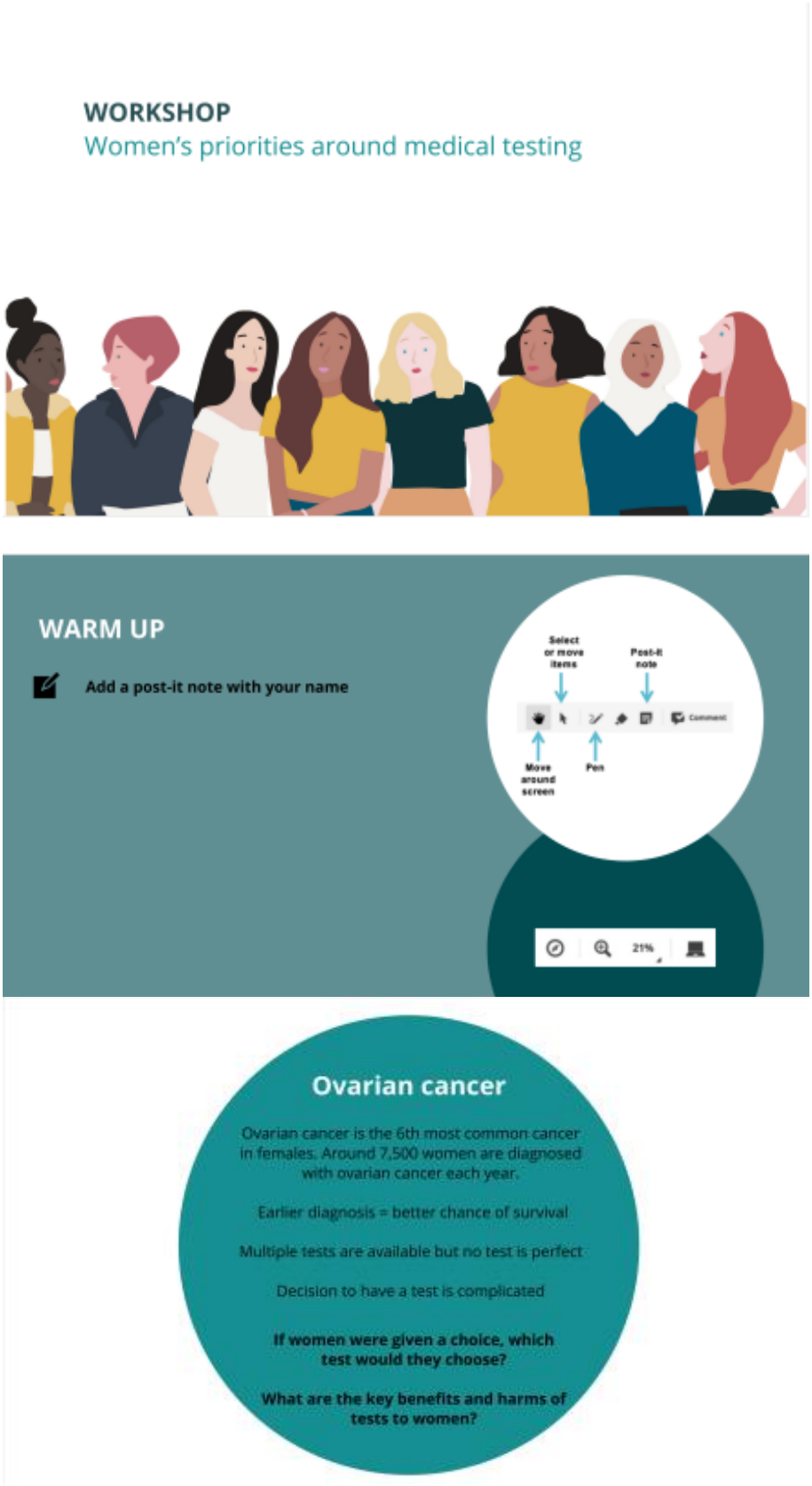


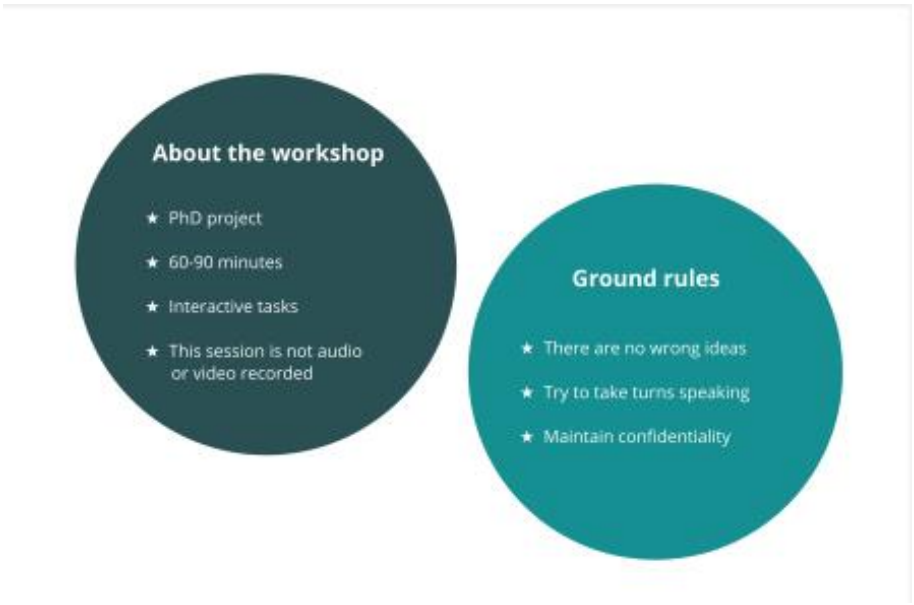
Appendix 5. 7: Results from the Health Risk Attitude Scale during the BWS study

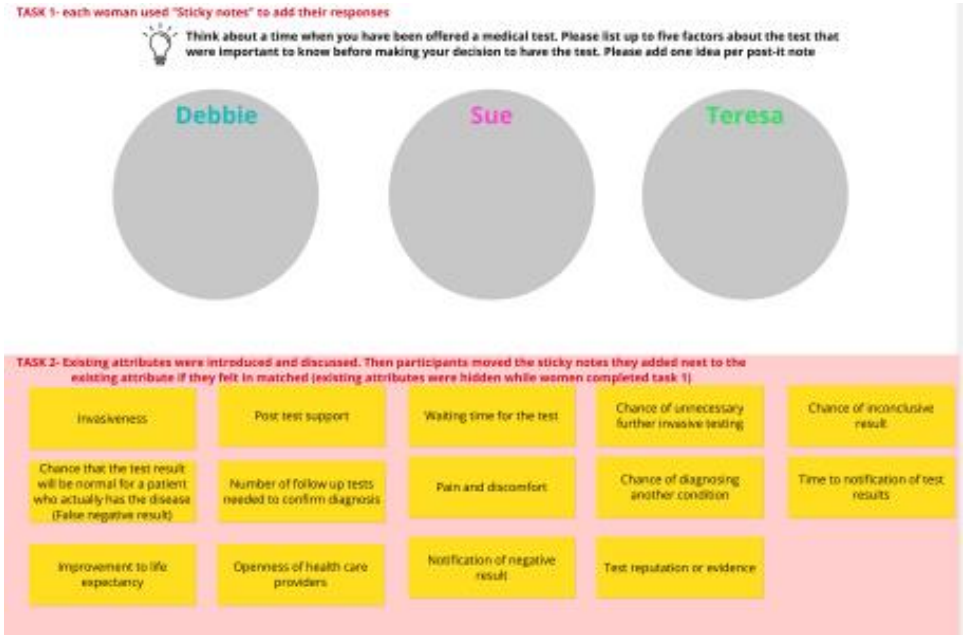
Responses were analysed based on a previous study (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4933649/pdf/cmajo.20150071.pdf).
Each response item was based on a Likert scale from 1-10. Higher scores indicated higher risk aversion. Scores from each question were summed to give a final risk attribute classification. A total score of <52 indicated risk seeking behaviour and score >52 indicated reek-aversion.
*= reverse scored items

Question		Mean score (SD)	Risk averse	Neutral	Risk seeking
1.	I think I take good care of my body.*	4.85 (1.4)	101 (68.7%)	19 (12.9%)	27 (18.4%)
2.	I don't want to have to consider the consequences for my health in everything that I do	3.5 (1.6)	37 (25.2%)	25 (17.0%)	85 (57.8%)
3.	It is important to me that I organize my life so that I will later enjoy good health.*	5.32 (1.3)	111 (75.5%)	24 (16.5%)	12 (8.2%)
4.	If it concerns my health, then I see myself as someone who avoids risks.*	4.84 (1.4)	89 (60.5%)	33 (22.4%)	25 (17.0%)
5.	Uncertainty about the consequences of a medical intervention is, in general, part of the game.	3.52 (1.3)	30 (20.4%)	41 (27.9%)	76 (51.7%)
6.	My health means everything to me.*	5.54 (1.4)	111 (75.5%)	21 (14.3%)	15 (10.2%)
7.	When I look back at my past, I think that, in general, I did take risks with my health.	3.22 (1.8)	40 (27.2%)	14 (9.5%)	93 (63.3%)
8.	If the doctor cannot offer me certainty about the possible consequences of a medical intervention, then I would rather not undergo it.*	4.34 (1.6)	67 (45.6%)	33 (22.4%)	47 (32%)
9.	Safety first, where my health is concerned.*	5.54 (1.2)	116 (78.9%)	24 (16.3%)	7 (4.8%)
10.	To enjoy good health now and in the future, I am prepared to forego a lot.*	4.8 (1.4)	89 (60.5%)	30 (20.4%)	28 (19.0%)
11.	People say that I take risks with my health because of my habits.	4.82 (1.9)	87 (59.2%)	18 (12.2%)	42 (28.6%)
12.	I'm not very fussy about my health.	4.85 (1.5)	86 (58.5%)	35 (23.8%)	26 (17.7%)
13.	In general I would estimate that I would not have much of a problem with undergoing a high risk operation.	4.13 (1.8)	54 (36.7%)	36 (24.5%)	57 (38.8%)
Total		**59.4 (9.4)**	**115 (78.2%)**	**4 (2.7%)**	**28 19.0%)**

Appendix 5.8: Conceptboard template used in workshops




About the workshop
★ PhD project
★ 60-90 minutes
★ Interactive tasks
★ This session is not audio or video recorded
Ground rules
★ There are no wrong ideas
★ Try to take turns speaking
★ Maintain confidentiality


TASK 1- each woman used "Sticky notes" to add their responses
Think about a time when you have been offered a medical test. Please list up to five factors about the test that were important to know before making your decision to have the test. Please add one idea per post-it note
Debbie
Sue
Teresa
TASK 2- Existing attributes were introduced and discussed. Then participants moved the sticky notes they added next to the existing attribute if they felt in matched (existing attributes were hidden while women completed task 1)
Invasiveness
Post test support
Waiting time for the test
Chance of unnecessary further invasive testing
Chance of inconclusive result
Chance that the test result will be normal for a patient who actually has the disease (False negative result)
Number of follow up tests needed to confirm diagnosis
Pain and discomfort
Chance of diagnosing another condition
Time to notification of test results
Improvement to life expectancy
Openness of health care providers
Notification of negative result
Test reputation or evidence

THANK YOU FOR JOINING OUR WORKSHOP

Appendix 9: Questionnaire distributed to workshop participants

Importance of features of ovarian cancer tests- survey

You have been invited to take part in this survey because you participated in a workshop on medical testing and indicated you would be willing to consider taking part in future related research.
The workshops focused on medical testing in general but we would not like to focus more on ovarian cancer.
The questionnaire should take no more than 30 minutes to complete including some background reading about ovarian cancer.
If you are happy to take part in this questionnaire, please return your response to Rebekah Hall (rh591@exeter.ac.uk) by DATE.

Instructions

To begin with we would like to provide you with some more information about ovarian cancer testing to help you complete the questionnaire.

Before beginning the questionnaire, please read the information below.

PREZI EMBEDDED HERE

Once you have read the information, please continue to the questionnaire on the next page.

Task 1:

The table below shows a list of characteristics relating to medical testing that were identified as important to women during a series of workshops, including the one you participated in.

We would like to know how important each of the characteristics is to you in the context of ovarian cancer. The characteristics vary between the different tests and these differences are shown in the second column of the table.

Please indicate how important you consider each characteristic after considering the possible differences between the tests. A score 1 indicates the characteristic is not very important and 5 indicates the characteristic is very important to you.

If you do not understand the wording or definition of a characteristic, please indicate this by checking the box under that heading.

Please note: A number of additional important issues were raised during the workshops. Factors relating to follow up care and treatment were particularly common, however, we have not included them here since they are unlikely to differ between the different tests.

Characteristic of tests	**Differences between the tests**	**I do not understand this characteristic**	**Not important at all** 1	2	3	4	**Very important** 5
If 100 people ***with ovarian cancer*** had this test, the test would*:*	A: correctly identify 75 people with cancer but miss 25 cases B: correctly identify 85 people with cancer but miss 15 cases C: correctly identify 95 people with cancer but miss 5 cases	☐	☐	☐	☐	☐	☐
If 100 ***without ovarian cancer*** had this test, the test would**:**	A: correctly identify 75 people without cancer but 25 people would have unnecessary further tests B: correctly identify 85 people without cancer but 15 people would have unnecessary further tests **C:** correctly identify 95 people without cancer but 5 people would have unnecessary further tests	☐	☐	☐	☐	☐	☐
How established is the test? *Definition: How long has the test been used for diagnosing ovarian cancer*	A: Less than 5 years B: Less than 10 years C: More than 10 years	☐	☐	☐	☐	☐	☐
Notification of negative results *Definition: Are you contacted and told your results even if the test suggests you do not have cancer?*	A: Yes B: No	☐	☐	☐	☐	☐	☐
Communication skills of the health care providers *Definition: Ability of staff to listen and explain things clearly throughout the testing process*	A: Good B: Fair C: Poor	☐	☐	☐	☐	☐	☐

If the test is positive, are any follow up tests before seeing a specialist? *Definition: Do you need any more tests from your GP before seeing a specialist for a final diagnosis?*	A: Yes B: No	☐	☐	☐	☐	☐	☐
Pain and discomfort during the test	A: None B: Mild C: Moderate	☐	☐	☐	☐	☐	☐
Waiting time for the test *Definition: The time between your GP recommending the test and having the test*	A: 1 week B: 4 weeks C: 6 weeks	☐	☐	☐	☐	☐	☐
What conditions can the test identify? *Definition: Most women will not have ovarian cancer but the test may be able to identify other conditions. Some conditions may be the cause of the symptoms but other conditions might be completely unrelated (not caused by the symptoms that you went to the GP about)*	A: Ovarian cancer only B: Ovarian cancer plus alternative conditions that caused similar symptoms (e.g. ovarian cysts) C: Ovarian cancer plus other unrelated conditions	☐	☐	☐	☐	☐	☐
Is the test invasive? *Definition: Does the test require cutting the skin or inserting instruments into the body*	A: Yes B: No	☐	☐	☐	☐	☐	☐
What does the test entail? *Definition: The process of having the test*	A: Blood test B: Internal ultrasound C: Blood test plus internal ultrasound	☐	☐	☐	☐	☐	☐
Post-test support *Definition: the type of support available after having the test*	A: None B: Leaflets and written information C: Follow up with nurse	☐	☐	☐	☐	☐	☐
Chance of an inconclusive result *Definition: The probability the result of the test is unclear or borderline and you would need to repeat the test*	A: 5 people out of 100 B: 15 people out of 100 C: 25 people out of 100	☐	☐	☐	☐	☐	☐
Waiting time for results *Definition: The time between having the test and receiving the results*	A: Immediately B: 1 week C: 2 weeks	☐	☐	☐	☐	☐	☐
Improvement to life expectancy *Definition: Would having the test improve the chance of survival due to better treatment options*	A: Yes B: No	☐	☐	☐	☐	☐	☐

Task 2:

In this part of the survey, we would like to understand how acceptable each of the characteristics would be to you if they were a feature of test you were recommended by your GP based on symptoms you were experiencing.

Characteristic of tests	I do not understand this characteristic	Highly unacceptable 1	2	3	4	Highly acceptable 5
If 100 people ***with ovarian cancer*** had this test, the test would*:* correctly identify 75 people with cancer but miss 25 cases	☐	☐	☐	☐	☐	☐
If 100 people ***with ovarian cancer*** had this test, the test would*:* correctly identify 85 people with cancer but miss 15 cases	☐	☐	☐	☐	☐	☐
If 100 people ***with ovarian cancer*** had this test, the test would*:* correctly identify 95 people with cancer but miss 5 cases	☐	☐	☐	☐	☐	☐
If 100 ***without ovarian cancer*** had this test, the test would**:** correctly identify 75 people without cancer but 25 people would have unnecessary further tests	☐	☐	☐	☐	☐	☐
If 100 ***without ovarian cancer*** had this test, the test would**:** correctly identify 85 people without cancer but 15 people would have unnecessary further tests	☐	☐	☐	☐	☐	☐
If 100 ***without ovarian cancer*** had this test, the test would**:** correctly identify 95 people without cancer but 5 people would have unnecessary further tests	☐	☐	☐	☐	☐	☐
The test been used to diagnose ovarian cancer for **less than 5 years**	☐	☐	☐	☐	☐	☐
The test been used to diagnose ovarian cancer **for less than 10 years**	☐	☐	☐	☐	☐	☐
The test been used to diagnose ovarian cancer for **more than 10 years**	☐	☐	☐	☐	☐	☐
You **are** notified if your test results are negative	☐	☐	☐	☐	☐	☐
You **are not** notified if your test results are negative	☐	☐	☐	☐	☐	☐
Health care providers have **good** communication skills	☐	☐	☐	☐	☐	☐
Health care providers have **fair** communication skills	☐	☐	☐	☐	☐	☐
Health care providers have **poor** communication skills	☐	☐	☐	☐	☐	☐
If the test is positive, you **will need** follow up tests before seeing a specialist	☐	☐	☐	☐	☐	☐
If the test is positive, you **will not need** follow up tests before seeing a specialist	☐	☐	☐	☐	☐	☐
The test causes **no** pain or discomfort	☐	☐	☐	☐	☐	☐
The test causes **mild** pain or discomfort	☐	☐	☐	☐	☐	☐
The test causes **moderate** pain and discomfort	☐	☐	☐	☐	☐	☐
The waiting time for the test is **1 week**	☐	☐	☐	☐	☐	☐
The waiting time for the test is **4 weeks**	☐	☐	☐	☐	☐	☐

The waiting time for the test is **6 weeks**	☐	☐	☐	☐	☐	☐
The test can identify **ovarian cancer only**	☐	☐	☐	☐	☐	☐
The test can identify **ovarian cancer plus alternative conditions that caused similar symptoms (e.g. ovarian cysts)**	☐	☐	☐	☐	☐	☐
The test can identify **ovarian cancer plus other unrelated conditions**	☐	☐	☐	☐	☐	☐
The test is **invasive**	☐	☐	☐	☐	☐	☐
The test is **non-invasive**	☐	☐	☐	☐	☐	☐
The test involves a **blood test**	☐	☐	☐	☐	☐	☐
The test involves an **internal ultrasound**	☐	☐	☐	☐	☐	☐
The test involves a **blood test plus an internal ultrasound**	☐	☐	☐	☐	☐	☐
Post-test support **is not provided**	☐	☐	☐	☐	☐	☐
Post-test support involves **leaflets and written information**	☐	☐	☐	☐	☐	☐
Post-test support involves **a follow up with a nurse**	☐	☐	☐	☐	☐	☐
5 people out of 100 receive an inconclusive result	☐	☐	☐	☐	☐	☐
15 people out of 100 receive an inconclusive result	☐	☐	☐	☐	☐	☐
25 people out of 100 receive an inconclusive result	☐	☐	☐	☐	☐	☐
You receive the results **immediately**	☐	☐	☐	☐	☐	☐
You receive the results **in 1 week**	☐	☐	☐	☐	☐	☐
You receive the results **in 2 weeks**	☐	☐	☐	☐	☐	☐
Having the test **would** improve the chance of survival	☐	☐	☐	☐	☐	☐
Having the test **would not** improve the chance of survival	☐	☐	☐	☐	☐	☐

Additional comments about the survey:

Thank you!

Please return the completed survey to Rebekah Hall (rh591@exeter.ac.uk) by Date. If you have any additional questions, please feel free to get in touch.

Please contact your GP if you have any concerns about your risk of ovarian cancer.

For more information about ovarian cancer, we recommend the following charities:

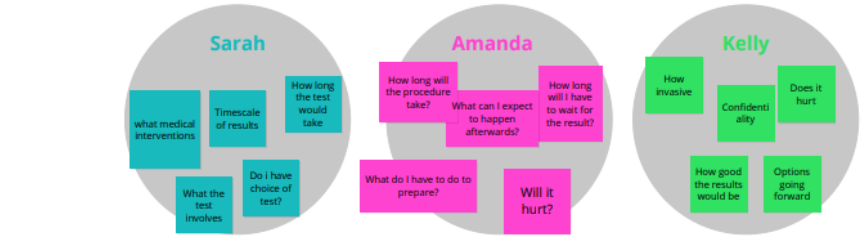
Target ovarian cancer: https://www.targetovariancancer.org.uk/

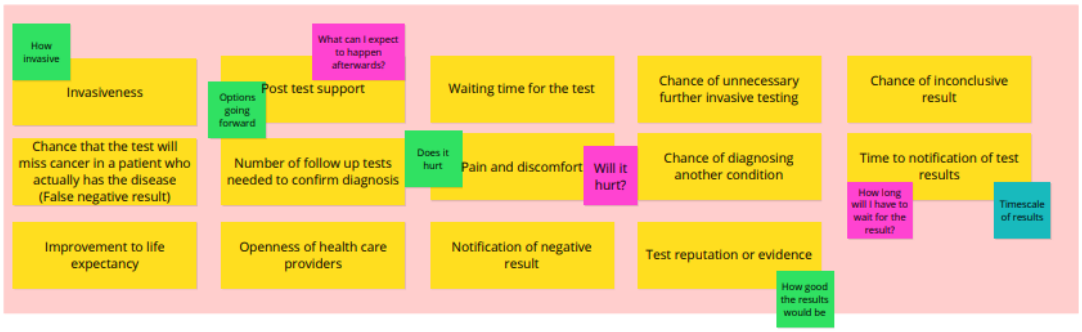
Ovacome: https://www.ovacome.org.uk/

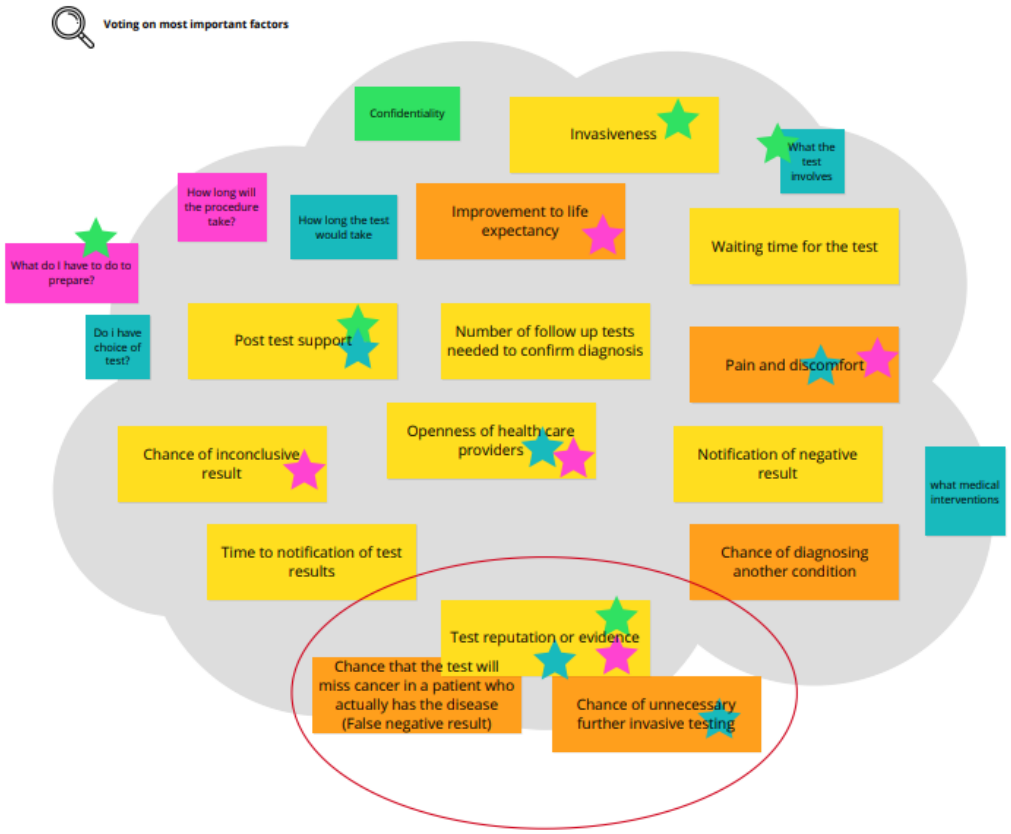
Ovarian cancer action: https://ovarian.org.uk/

Appendix 5.10: Pseudo-anonymised results from workshop tasks

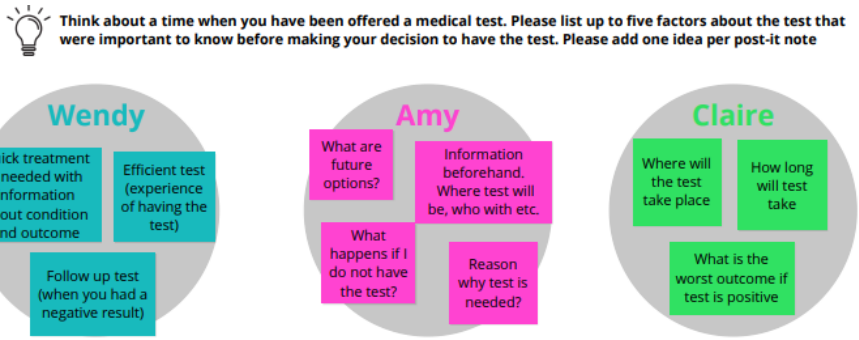
a. Workshop 1

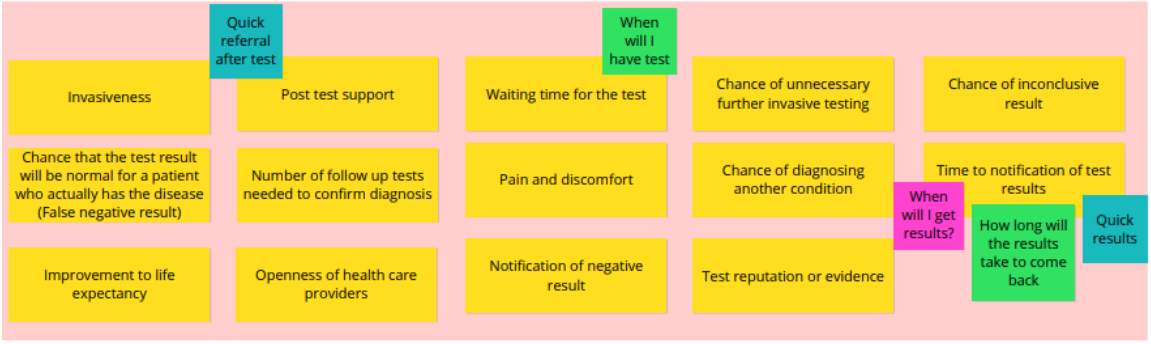


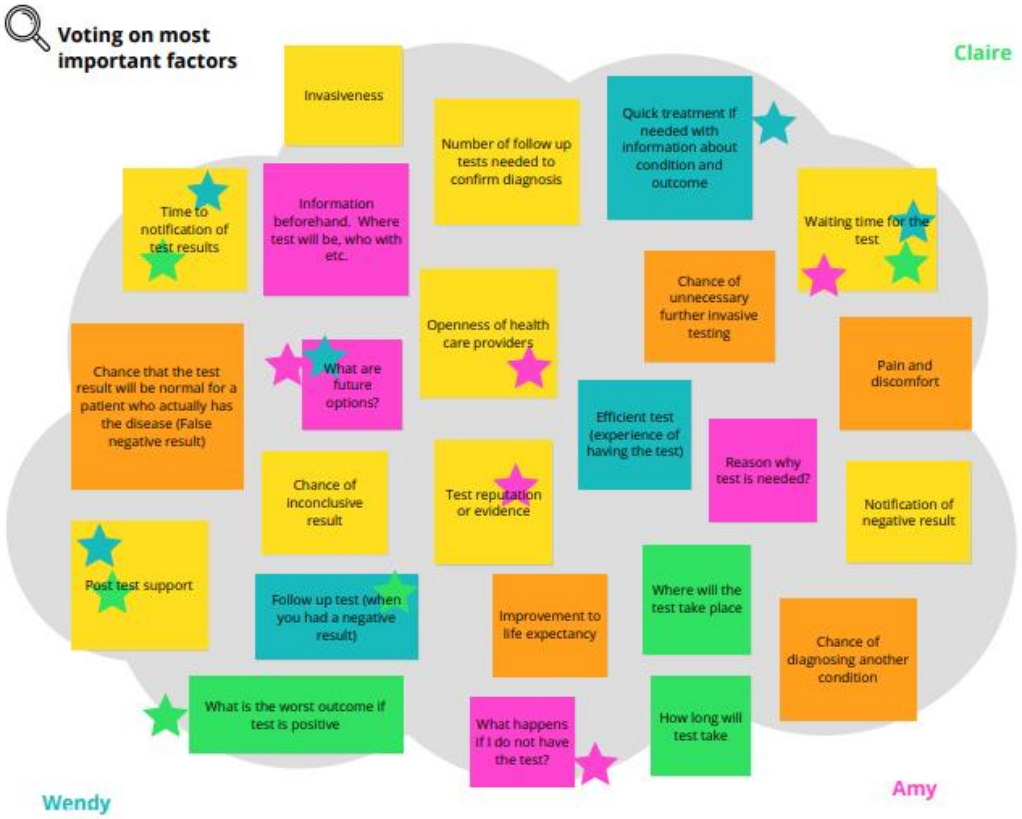




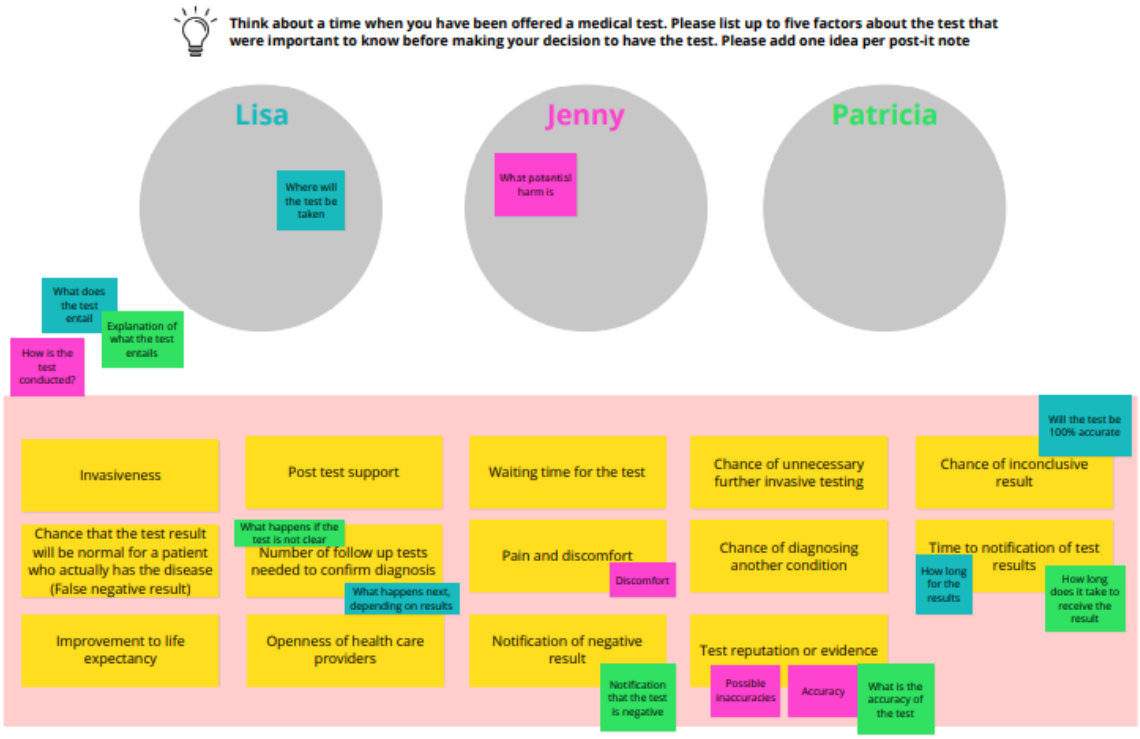
b. Workshop 2

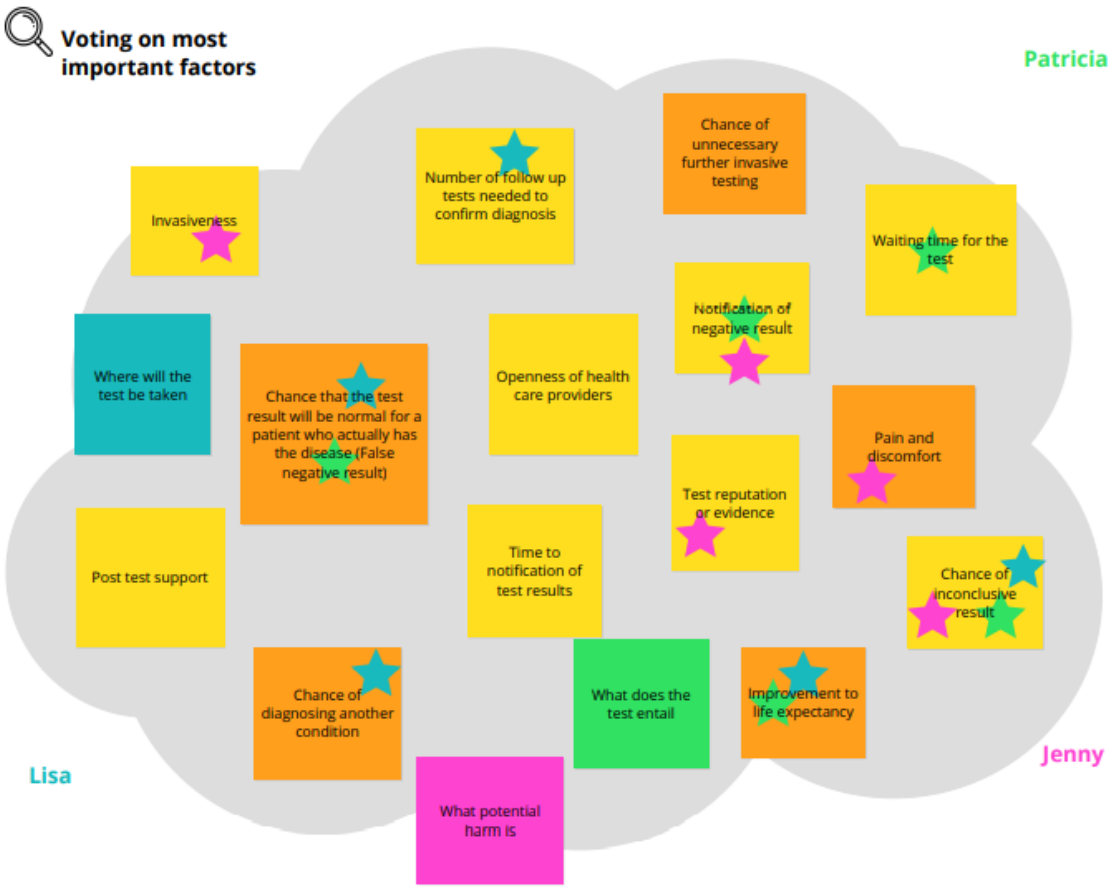

Think about a time when you have been offered a medical test. Please list up to five factors about the test that were important to know before making your decision to have the test. Please add one idea per post-it note
Wendy
Quick treatment if needed with information about condition and outcome
Efficient test (experience of having the test)
Follow up test (when you had a negative result)
Amy
What are future options?
Information beforehand. Where test will be, who with etc.
What happens if I do not have the test?
Reason why test is needed?
Claire
Where will the test take place
How long will test take
What is the worst outcome if test is positive


Quick referral after test
When will I have test
Invasiveness
Post test support
Waiting time for the test
Chance of unnecessary further invasive testing
Chance of inconclusive result
Chance that the test result will be normal for a patient who actually has the disease (False negative result)
Number of follow up tests needed to confirm diagnosis
Pain and discomfort
Chance of diagnosing another condition
Time to notification of test results
When will I get results?
How long will the results take to come back
Quick results
Improvement to life expectancy
Openness of health care providers
Notification of negative result
Test reputation or evidence


Voting on most important factors
Claire
Invasiveness
Number of follow up tests needed to confirm diagnosis
Quick treatment if needed with information about condition and outcome
Time to notification of test results
Information beforehand. Where test will be, who with etc.
Waiting time for the test
Chance of unnecessary further invasive testing
Openness of health care providers
Chance that the test result will be normal for a patient who actually has the disease (False negative result)
What are future options?
Pain and discomfort
Efficient test (experience of having the test)
Reason why test is needed?
Chance of inconclusive result
Test reputation or evidence
Notification of negative result
Post test support
Follow up test (when you had a negative result)
Improvement to life expectancy
Where will the test take place
Chance of diagnosing another condition
What is the worst outcome if test is positive
What happens if I do not have the test?
How long will test take
Wendy
Amy

c. Workshop 3


Think about a time when you have been offered a medical test. Please list up to five factors about the test that were important to know before making your decision to have the test. Please add one idea per post-it note
Lisa
Jenny
Patricia
Where will the test be taken
What potential harm is
What does the test entail
Explanation of what the test entails
How is the test conducted?
Invasiveness
Post test support
Waiting time for the test
Chance of unnecessary further invasive testing
Chance of inconclusive result
Will the test be 100% accurate
Chance that the test result will be normal for a patient who actually has the disease (False negative result)
What happens if the test is not clear
Number of follow up tests needed to confirm diagnosis
What happens next, depending on results
Pain and discomfort
Discomfort
Chance of diagnosing another condition
Time to notification of test results
How long for the results
How long does it take to receive the result
Improvement to life expectancy
Openness of health care providers
Notification of negative result
Notification that the test is negative
Test reputation or evidence
Possible inaccuracies
Accuracy
What is the accuracy of the test


Voting on most important factors
Patricia
Invasiveness
Number of follow up tests needed to confirm diagnosis
Chance of unnecessary further invasive testing
Waiting time for the test
Notification of negative result
Where will the test be taken
Chance that the test result will be normal for a patient who actually has the disease (False negative result)
Openness of health care providers
Pain and discomfort
Test reputation or evidence
Post test support
Time to notification of test results
Chance of inconclusive result
Chance of diagnosing another condition
What does the test entail
Improvement to life expectancy
Lisa
What potential harm is
Jenny

Appendix 5.11: Summary of attributes excluded during the finalisation stage of attribute development

Attribute	Levels	Rankings		Reason for exclusion
		BWS	PPI	
Improvement to life expectancy	Yes, No	1	4	Unlikely to differ between tests for symptomatic population. Should be considered for screening
Pain and discomfort	None, mild, moderate	4	15	Low importance once levels were introduced.
Specificity	75,85,95%	5	9	Specificity is likely to be very high for all tests in symptomatic population (>90%). More applicable to screening.
Chance of an inconclusive result	75,85,95%	7	11	Overlaps with sensitivity and specificity.
What does the test entail?	Blood test, internal ultrasound, blood test + internal ultrasound	9	14	Low importance. Inclusion would mean DCE was limited to labelled design.
Is the test invasive	Yes, no	-	13	Low importance
Notification of negative result	Yes, no	11	3	Unlikely to systematically differ between tests. Overlap with communication skills. More applicable to screening
Number of follow up tests (BWS)/ Are follow up tests needed? (PPI)	Yes, no	12	12	Low importance
Post-test support	None, leaflets, HCP	13	9	Low importance. Tests are intermediate so follow up care/support is more relevant after final diagnosis
Acceptability to GPs	Widely accepted, criticised by some	-	4	Driven by accuracy of the test. Previous criticism of CA125 unlikely to still be present particularly given recently published high sens/spec
Length of use	<5 years, 5-10 years, >10 years	-	16	Low importance

Appendix 6.1: Overview of rationality checks used within DCE studies

Monotonicity (or non-satiation)

During this test, respondents are presented with two test profiles where one of the alternatives is irrefutably better than the other (test A in the example below). This is to test the underlying assumption of monotonicity (i.e. more/better is always preferable). This test is often presented as a warm-up question or pre-requisite for inclusion where failure to choose the more desirable option may be interpreted as a failure to understand the task as a whole.

Q3	Test A	Test B
Test sensitivity	95%	75%
Identifiable conditions	Ovarian cancer plus related conditions	Ovarian cancer only
Time to completion of testing	2 weeks	8 weeks
Communication skills of HCP	Good	Poor
Which test would you choose?	☐	☒

Test-retest stability

Rational decision theory suggests choices should be consistent over time . To test the stability of choices, an early choice task was repeated later in the questionnaire to check whether the respondent will choose the same test.

Example:

Respondents complete the same choice sets in question 1 and question 10

Q1	Test A	Test B
Test sensitivity	65%	85%
Identifiable conditions	Ovarian cancer plus unrelated conditions	Ovarian cancer only
Time to completion of testing	5 weeks	8 weeks
Communication skills of HCP	Good	Fair
Which test would you choose?	☐	☒

Q10	Test A	Test B
Test sensitivity	65%	85%
Identifiable conditions	Ovarian cancer plus unrelated conditions	Ovarian cancer only
Time to completion of testing	5 weeks	8 weeks
Communication skills of HCP	Good	Fair
Which test would you choose?	☐	☒

Transitivity

Transitivity is one of the axioms of utility theory and is requirement for rational preferences. Transitivity means that if Test A is preferred to Test B and Test B is preferred to Test C then Test A **must** be preferred to test C (example below demonstrates a failure of transitivity). Transitivity requires two additional choice tasks are completed by respondents. Perhaps for this reason, transitivity tends to be less tested than other validity tests within the health care literature.

Q2	Test A	Test B
Test sensitivity	85%	75%
Identifiable conditions	Ovarian cancer plus unrelated conditions	Ovarian cancer plus related conditions
Time to completion of testing	8 weeks	5 weeks
Communication skills of HCP	Poor	Good
Which test would you choose?	☒	☐

Q5	Test C	Test B
Test sensitivity	65%	75%
Identifiable conditions	Ovarian cancer only	Ovarian cancer plus related conditions
Time to completion of testing	2 weeks	5 weeks
Communication skills of HCP	Fair	Good
Which test would you choose?	☐	☒

Q7	Test A	Test C
Test sensitivity	85%	65%

Identifiable conditions	Ovarian cancer plus unrelated conditions	Ovarian cancer only
Time to completion of testing	8 weeks	2 weeks
Communication skills of HCP	Poor	Fair
Which test would you choose?	☐	☒

Non-trading – flat-lining

This is a straightforward inspection of responses to check whether any participant has repeated chosen a choice task in the same position throughout all question (e.g. selected "Test A" for all tasks). The probability that the more preferred alternative will always appear in the same position is small enough to suggest that respondents are not fully evaluating the profiles being presented. This check reduces the likelihood of low-quality or unconsidered responses and is particularly important to ensure the quality of data given the use of an online panel for recruitment and administration (Tervonen *et al.*, 2018). No additional choice tasks are required to check for flat-lining behaviours.

Caveats of rationality tests

Decisions of how many and which choice tasks to include is important to consider. The inclusion of validity checks often requires the introduction of one or more additional choice sets outside of the experimental design. As a result, rationality checks reduce the overall efficiency of a DCE and may increase respondent burden and fatigue (Johnson *et al.*, 2019) . Rationality questions may also distract or alter respondent behaviour if respondents notice the inclusion of such checks and feel they are being "tested" in some way.

Failure to act rationally according to behavioural theory does not necessarily mean that an individual is irrational with their choices. Instead, motivations for responses may not fully captured within the confines of data collection. For example, Ryan *et al.* (2009) found that several respondents who appeared to fail the monotonicity had actually inferred additional information about the alternatives (e.g. quality of a test based on the cost) and on this basis, choices subsequently appeared “rational”.

Just as the interpretation of “irrational” responses is debated, so is how such responses should be managed within analysis. Some researchers remove irrational respondents prior to analysis. However, excluding people who fail rationality tests even where failures are genuine, may introduce bias if there is a systematic reason for failures (e.g.. lower health literacy and numeracy skills) (Tervonen *et al.*, 2018) . In such instances, removing “irrational” responders may increase the internal validity of the experiment but lower the external validity. Instead, many researchers opt to keep the full sample but may choose to perform sensitivity checks to understand how model estimates differ between those who “pass” and “fail” rationality check questions (Pearce *et al.*, 2020; Ryan & Bate, 2001).

Appendix 6.2: Recruitment poster for think-aloud interview pilot study

CANCER RESEARCH UK

UNIVERSITY OF EXETER

Volunteers needed!

Women and people with ovaries

We want to hear from you!

PREFERENCES FOR OVARIAN CANCER TESTING: an interview study

We are looking for volunteers over 40 years old to take part in an online interview. We want to understand more about women's preferences towards diagnostic tests and the thought processes behind how people make decisions when given choices.

Interviews will take place online via Microsoft Teams.

The interview will last around 45 minutes. You will receive a £10 amazon gift voucher for your time.

To take part or for more information. Please contact Rebekah Hall at rh591@exeter.ac.uk

Appendix 6.3: Information sheet for think-aloud interview pilot study

UNIVERSITY OF EXETER

COLLEGE OF MEDICINE AND HEALTH

CANCER RESEARCH UK

INFORMATION SHEET FOR INTERVIEW PARTICIPANTS
What features of tests for ovarian cancer are most important?
VERSION NUMBER [3]: DATE [09/10/20]

You are being invited to take part in an interview study that aims to understand how women and people with ovaries make decisions about medical testing for ovarian cancer. This study forms part of a PhD project funded by Cancer Research UK and conducted at the University of Exeter. Before you decide whether to take part in the study it is important that you understand why the research is being done and what taking part will involve. Please take time to read the following information carefully. Please ask if there is anything that is not clear or if you would like more information.

What is the aim of the project?
Traditionally decisions about when and how to test for cancer have been made by doctors. We would like to understand more about the decisions women and people with ovaries might make if they were given more of a choice. In particular, we want to learn what characteristics of diagnostic testing are most important to women and people with ovaries and how they make choices about medical tests.

Why have I been invited to take part?
We are approaching you because we are seeking responses from people with ovaries over the age of 40. You do not need to have any prior knowledge of ovarian cancer or have been previously tested for ovarian cancer to take part in the study. You must speak fluent English to take part. **Please do not take part in this study if you have ever undergone a procedure to remove both of your ovaries.**

What will I be asked to do?
Should you agree to take part, you will be invited to complete an online interview. The interview should last around 45-60 minutes up to a maximum of 90 minutes. During the interview you will complete a survey where you will be asked to imagine you are experiencing symptoms of ovarian cancer. You will be shown descriptions of two different imaginary medical tests and asked which test you would prefer to have.

What are the possible disadvantages and risks of taking part?
Participating in the research is not anticipated to cause you any disadvantages or discomfort. Some questions may be considered sensitive. You do not have to provide responses to any questions you feel uncomfortable answering. During the interview you will be asked to imagine you might have cancer, for some this may cause anxiety. Resources for help and information relating to ovarian cancer will be provided at the end of the interview.

Will I be paid?

You will also receive a £10 Amazon gift voucher for participating in this research.

Can I change my mind and withdraw from the study?
You may withdraw from the study at any time, without any disadvantage to yourself of any kind.
What data or information will be collected and what use will be made of it?
All your responses during the interview are completely confidential. Your interview will be

recorded digitally and transcribed. Transcripts will be made available for you to review after the interview.

Any data will be managed according to a law called the Data Protection Act (2018). Anonymised interview data will be stored securely on an encrypted password protected computer for a period of five years. Only member of the research team will have access to your data.
This project involves an open-questioning technique where the precise nature of the questions asked have not been determined in advance, but will depend on the way in which the interview develops. Consequently, although the Research Ethics Committee is aware of the general areas to be explored in the interview, the Committee has not been able to review the precise questions to be used. In the event that the line of questioning does develop in such a way that you feel hesitant or uncomfortable, you are reminded of your right to decline to answer any particular question(s) and also that you may withdraw from the project at any stage without any disadvantage to yourself of any kind.

The results of the study may be published in academic journals or conferences but any included data will be anonymised and you will not be individually identifiable.

How will my information be kept confidential?
The University of Exeter processes personal data for the purposes of carrying out research in the public interest. The University will endeavour to be transparent about its processing of your personal data and this information sheet should provide a clear explanation of this. If you do have any queries about the University's processing of your personal data that cannot be resolved by the research team, further information may be obtained from the University's Data Protection Officer by emailing informationgovernance@exeter.ac.uk or at www.exeter.ac.uk/ig/

What if I have any questions?
If you have any questions about our project, either now or in the future, please feel free to contact either Rebekah Hall (email: rh591@exeter.ac.uk) or Anne Spencer (email: a.e.spencer@exeter.ac.uk)

Complaints
If you have any complaints about the way in which this study has been carried out please contact the Chair of the College of Medicine and Health Research Ethics Committee:-
Mark Tarrant, PhD
Chair of the CMH Research Ethics Committee
Email: cmhethics@exeter.ac.uk

This project has been reviewed and approved by the
University of Exeter College of Medicine and Health Research Ethics Committee
(REF NUMBER: 20/09/261)

Appendix 6.4: Consent form for think-aloud pilot interviews

UNIVERSITY OF EXETER
COLLEGE OF MEDICINE AND HEALTH

CANCER RESEARCH UK

CONSENT FORM FOR INTERVIEW PARTICIPANTS

What features of tests for ovarian cancer are most important?

VERSION NUMBER [3]: DATE [09/10/20]

Participant Identification Number:

		Please initial box
1.	I confirm that I have read the information sheet dated [3] version no. [09/10/20] for the above project. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily	
2.	I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without my legal rights being affected.	
3.	I understand that a digital recording and transcript of my interview will be retained and stored securely for a period of five years.	
4.	I understand that relevant sections of the data collected during the study may be looked at by members of the research team, individuals from the University of Exeter, Cancer Research UK or regulatory authorities for audit purposes	
5.	I understand that this study will involve an open-questioning technique, and I can decline to answer any particular question(s) without any disadvantage	
6.	I understand that the results of the study may be published in an academic publication but included data will be anonymised and my anonymity will be preserved	
7.	I understand that my travel expenses will be reimbursed in line with standard university rates and I will receive £10 gift voucher for my participation	

Please confirm that you have read and understood the following information:

I agree to take part in the study:
Yes □
No □

.....	……..	.....
(Printed name of participant)	(Signature of participant)	(Date)
.....	……..	.....
(Printed name of researcher)	(Signature of researcher)	(Date)

This project has been reviewed and approved by the University of Exeter College of Medicine and Health Research Ethics Committee (REF NUMBER: 20/09/261)

Appendix 6.5: Interview schedule

Introduction to purpose of the study and verbal consent (8 mins)

Introduction of interviewers

Verbal consent
(Written consent received prior to interview)
Reminder of the information sheet. Recap of consent form (free to withdraw, interview is recorded for transcription at which point it will be deleted etc.) Ask if they are happy to go ahead and start interview.

Background to study
We are interested in testing the usability of the questionnaire we have developed, it not a test of the user themselves, questions are based on your own opinion and there are no right or wrong answers

(Recording starts here)

Introduction to thinking-aloud (5 mins)

Explain that we're interested in testing the system that we're about to present and that we're not testing the user specifically

Ask the participant to tell us everything they're thinking about from the moment they read the task and when they complete it.

Tell them that they don't need to plan/think out what they want to say. Just act as if they're by themselves Emphasize that the important thing is to keep talking

Explain that if you notice long periods of silence you may interrupt and remind them to keep talking.

You will not be able to respond to any questions during the observation but will happily answer any questions at the end of the task

Warm up exercise (5 mins)

To help the user get a feel for actually performing a Think-Aloud, perform an example of think-aloud e.g. "Please think aloud as you name how many windows are in your house?"

Think-aloud tasks (20 mins)

Introduction to the task: discrete choice experiments and attributes and levels included

Remind participant: most important thing is to keep talking, do not worry about making sense or explaining what you are doing to me. Act as if you are alone and are thinking out loud as you complete tasks.

Does the user have any questions before tasks begin?

During tasks: interviewers will remain silent unless participant stops talking then probes will be used to encourage talking. If the user becomes frustrated, take a quick break

Probes:

- What are you thinking now?
- Why did you choose that option?
- What did you consider when making that choice?
-

Follow up questions (5 mins)

What did you think of the questions?

Did you have difficulty understanding any of the tasks?
Did you have difficulty understanding any of the wording?
Were you satisfied with the choices you made?
Would you have answered differently if you were alone?

Wrap up discussion (5 mins)

Any further feedback or thoughts?
Any questions?
Thank you for completing the interview

Appendix 6.6: Key participant demographics from quantitative pilot study

	Version 1	Version 2	Version 3	Version 4
Age				
Mean (SD)	49.9 (8.7)	52.5 (11.1)	52.4 (7.7)	51.4 (9.3)
Range	41-77	40-74	40-65	40-80
Ethnicity, n (%)				
White	24 (96%)	23 (92%)	24 (96%)	23 (92%)
Other	1 (4%)	1 (4%)	1 (4%)	2 (8%)
Prefer not to say	-	1 (4%)	-	-
Previously tested for ovarian cancer, n (%)				
Yes	3 (12%)	4 (16%)	3 (12%)	2 (8%)
No	21 (84%)	20 (80%)	21 (84%)	23 (92%)
Don't know	1 (4%)	1 (4%)	-	-
Self-reported health, n (%)				
Very good	7 (28%)	5 (20%)	4 (16%)	1 (4%)
Good	10 (40%)	14 (56%)	12 (48%)	19 (76%)
Fair	7 (28%)	6 (24%)	6 (24%)	5 (20%)
Poor	-	-	3 (12%)	-
Very poor	-	-	-	-
Prefer not to say	1 (4%)	-	-	-

Appendix 6.7: Ngene syntax for the final experimental design

```
design
;alts(m1) = alt1*, alt2*, alt3
;alts(m2) = alt1*, alt2*, alt3
;alts(m3) = alt1*, alt2*, alt3
;rows = 16
;block= 2
;eff = m1(mnl, d,median)+ m2(mnl, d,median)+ m3(mnl, d,median)

;bdraws= sobol(500)
;model(m1) :
U(alt1)=  bsens[(n,0.0623,0.01088)] * sens[65,75,85,95] + btime[(n,-
0.2516,0.083)] * time[1,2,3,4] + bconditions.dummy[(n,0.6704,0.1565)] *
conditions[1,0] + bcomm.dummy[(n,0.9280,0.257)|(n,1.3418,0.2253)] *
comm[1,2,0] /
U(alt2)=  bsens        * sens              + btime         * time
+ bconditions.dummy        * conditions      + bcomm.dummy               *
comm        /
U(alt3)= b3[-1.990]

;model(m2) :
U(alt1)=  bsens[(n,0.0604,0.0103)] * sens[65,75,85,95] + btime[(n,-
0.3124,0.0871)] * time[1,2,3,4] + bconditions.dummy[(n,0.4131,0.2021)] *
conditions[1,0] + bcomm.dummy[(n,0.5670,0.2053)|(n,6654,0.2236)] *
comm[1,2,0] /
U(alt2)=  bsens        * sens              + btime         * time
+ bconditions.dummy        * conditions      + bcomm.dummy               *
comm        /
U(alt3)= b3[-1.048]


;model (m3) :
U(alt1)= bsens[(n,0.0876,0.01266)] * sens[65,75,85,95] + btime[(n,-
0.5232,0.1279)] * time[1,2,3,4] + bconditions.dummy[(n,0.5536,0.2240)] *
conditions[1,0] + bcomm.dummy[(n,0.8383,0.2639)|(n,1.052,0.2842)] *
comm[1,2,0] /
U(alt2)= bsens                    * sens              + btime
* time          + bconditions.dummy         * conditions      + bcomm.dummy
* comm   /
U(alt3)= b3[-1.08]

$
```

Appendix 6.8: Experimental design for final data collection

The final design was split in to two blocks to address how the number of choice tasks impacts preference behaviour. Respondents completed all questions but the order of completion was randomised across respondents to either; Block A then Block B or Block B then Block A. There were two transitivity questions per respondent (T1 and T2 or T3 and T4 depending on order of block completion). Further details are provided in the Chapter 8.

	Choice task	Alternative 1				Alternative 2			
		Accuracy	Timing	Conditions	Communication	Accuracy	Timing	Conditions	Communication
Block 1	1	85	3	Related conditions	Fair	75	1	Cancer only	Poor
	2	85	3	Related conditions	Fair	85	2	Cancer only	Good
	3	95	2	Cancer only	Poor	65	3	Related conditions	Good
	4	75	2	Cancer only	Fair	95	3	Related conditions	Good
	5	65	1	Related conditions	Fair	95	4	Cancer only	Poor
	6	75	4	Cancer only	Good	65	1	Related conditions	Poor
	7	95	2	Cancer only	Poor	65	3	Related conditions	Fair
	8	65	3	Related conditions	Good	95	2	Cancer only	Fair
Block 2	9	75	3	Related conditions	Poor	85	2	Cancer only	Fair
	10	85	4	Cancer only	Fair	85	1	Related conditions	Poor
	11	95	1	Cancer only	Good	75	4	Related conditions	Fair
	12	65	1	Cancer only	Good	85	4	Related conditions	Poor
	13	65	1	Related conditions	Poor	95	4	Cancer only	Fair
	14	95	2	Related conditions	Fair	75	3	Cancer only	Good
	15	85	4	Related conditions	Good	75	1	Cancer only	Fair
	16	75	4	Cancer only	Poor	65	2	Related conditions	Good
Rationality checks questions	T1	85	3	Related conditions	Fair	75	1	Cancer only	Poor
	T2	75	1	Cancer only	Poor	85	2	Cancer only	Good
	T3	85	4	Cancer only	Fair	75	2	Related conditions	Good
	T4	75	2	Related conditions	Good	85	1	Related conditions	Poor
	S1	85	3	Related conditions	Fair	75	1	Cancer only	Poor
	S9	75	3	Related conditions	Poor	85	2	Cancer only	Fair
	M	95	2	Related conditions	Good	75	3	Cancer only	Fair

Appendix 6.9: Final version of the survey

What features of tests for ovarian cancer are most important?

INFORMATION SHEET FOR PARTICIPANTS

VERSION NUMBER [3]: DATE [09/10/20]

Thank you for showing an interest in this survey. Please take time to read the following information carefully before deciding whether or not to take part.

What is the aim of the project?
Traditionally decisions about when and how to test for cancer have been made by doctors. In this study we would like to understand more about the decisions women and people with ovaries might make if they were given more of a choice. We aim to understand the attitudes and preferences of women and people with ovaries relating to testing for possible ovarian cancer. In particular, we want to learn what characteristics of diagnostic testing are most important.

Why have I been invited to take part?
We are approaching you because we are seeking responses from women and people with ovaries over the age of 40. You do not need to have any prior knowledge of ovarian cancer and you do not have to have been previously tested for ovarian cancer to take part in the study. You must be able to complete the survey in English to take part. **Please do not take part in this survey if you have ever undergone a procedure to remove both of your ovaries.**

What will I be asked to do?
Should you agree to take part, you will be asked to complete a survey lasting about 15 minutes. During the survey you will be asked to imagine you are experiencing symptoms of ovarian cancer. You will be shown descriptions of two different medical tests and asked which test you would prefer to have. In total, we will ask you about 16 pairs of tests. At the end of the survey we will ask you some additional questions about yourself. This will help us to better understand how attitudes might vary from person to person.

What are the possible disadvantages and risks of taking part?
Participating in the research is not anticipated to cause you any disadvantages or discomfort. Some questions may be considered sensitive, however, you do not have to provide responses to any questions you feel uncomfortable answering. During the survey you will be asked to imagine you might have cancer, for some this may cause anxiety.

Will I be paid?
You will receive a payment of £1.80 for completing the survey. Payments will be paid directly into your Prolific account within 10 working days.

Please note: In order to receive the payment you must complete the full survey and click the completion link at the end of the survey. To ensure responses are high quality, an "attention check" question has been included within the survey. You must correctly complete this question correctly to receive payment.

Can I change my mind and withdraw from the project?
If you decide you no longer wish to take part during the survey, simply exit the webpage to withdraw. Your incomplete responses will be permanently deleted. If you decide to withdraw after submitting your responses, please contact us via your Prolific account or directly by email. You can withdraw from the study for up to 14 days after completion and do no not have to give a reason. After 14 days it may no longer be possible to withdraw your submission because anonymization will mean we cannot link responses to individual participants.

Is the survey confidential?
All your answers to the survey are completely confidential and anonymous. You will not be asked your name or any other identifying information. Your responses will be securely stored on an encrypted password protected computer and managed according to a law called the Data Protection Act (2018). Your anonymised data will be stored for a period of five years.

In line with the Cancer Research UK data sharing guidelines, your data may be shared with other researchers in the future at our discretion. Any shared data will be fully anonymised. For more information: https://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/submission-of-a-data-sharing-and-preservation-strategy/data-sharing-guidelines

The results of the study may be published in academic journals or conferences but any included data will not be individually identifiable.

The University of Exeter processes personal data for the purposes of carrying out research in the public interest. The University will endeavour to be transparent about its processing of your personal data and this information sheet should provide a clear explanation of this. If you do have any queries about the University's processing of your personal data that cannot be resolved by the research team, further information may be obtained from the University's Data Protection Officer by emailing dataprotection@exeter.ac.uk or at www.exeter.ac.uk/dataprotection

What if I have any questions?
If you have any questions about our project, either now or in the future, please feel free to contact Rebekah Hall by emailing rh591@exeter.ac.uk

Complaints
If you have any complaints about the way in which this study has been carried out please contact the Chair of the College of Medicine and Health Research Ethics Committee:-

Mark Tarrant, PhD
Chair of the CMH Research Ethics Committee
Email: cmhethics@exeter.ac.uk

This project has been reviewed and approved by the
University of Exeter College of Medicine and Health Research Ethics Committee (REF NUMBER: 20/09/261)

Consent Form

- I understand that my participation is voluntary and that I am free to withdraw for up to 14 days without giving any reason and without my legal rights being affected.
- I understand that my data from the study will be fully anonymised and will be looked at by members of the research team and may potentially be shared with other researchers in future if appropriate.
- I understand that relevant sections of the data collected during the study may be looked at by individuals from the University of Exeter, Cancer Research UK or regulatory authorities for audit purposes
- I understand that the results of the study may be published in academic journals but my anonymity will be preserved
- I understand that my anonymised data will be securely stored on an encrypted password protected computed for a period of five years.
- I understand that in order to receive payment for this survey I must complete the full survey and click the link at the end of the survey. I must also correctly complete three attention check questions.

I confirm that I have read the information above and agree to take part in the study:

- ○ Yes
- ○ No

Please enter your ProlificID

Have you ever had a medical procedure that involved the removal of both of your ovaries?

- ○ Yes
- ○ No

Thank you for agreeing to take this survey

The survey will present some information about ovarian cancer and describe some tests. We will then ask you some questions about your health and about cancer testing. Later we will ask you to consider different tests for ovarian cancer.

Let's start with some information about ovarian cancer…

Next

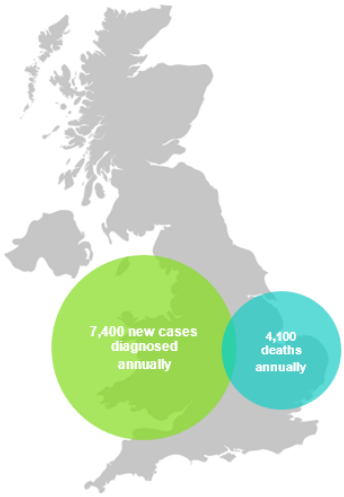
Ovarian cancer occurs when the cells in and around the ovaries and fallopian tubes become abnormal, grow out of control and form a lump called a "tumour".

Ovarian cancer is the 6th most common cancer for women in the UK. Over 7,000 women are diagnosed annually. Most of these cases occur in women over the age of 40.

Diagnosing ovarian cancer early means more treatment options are available and inceases the chance of being cured or living longer.

There are a few tests that can be used to diagnose ovarian cancer. These tests differ in many ways.

We would like to find out what kind of tests women like you would prefer to have.



Part 1: Ovarian cancer knowledge

We would like to learn more about your knowledge and experience of ovarian cancer.

What is your age?

Have you ever undergone testing for possible ovarian cancer?

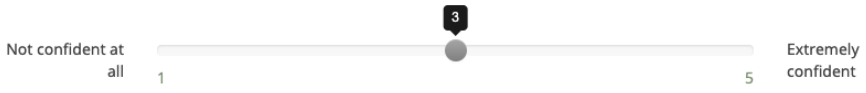
- Yes
- No
- I don't know
- Prefer not to say

What test(s) did you have?

Check all that apply

- [ ] Blood test
- [ ] Transvaginal/internal ultrasound
- [ ] CT scan
- [ ] Laparoscopy
- [ ] I don't know
- [ ] Other:

How confident are you that you would notice a symptom of ovarian cancer?



Which of the following do you recognise as a symptom of ovarian cancer?

Check all that apply

- [ ] Feeling constantly bloated
- [ ] A swollen tummy
- [ ] Discomfort in your tummy
- [ ] Persistent indigestion or feeling sick
- [ ] Discomfort in your pelvic area
- [ ] A change in bowel habits
- [ ] Back pain
- [ ] Pain during sex
- [ ] Feeling full quickly or loss of appetite
- [ ] Feeling tired all the time
- [ ] Unintentional weight loss
- [ ] Needing to pee more often or more urgently than usual
- [ ] None

If you had a symptom that you thought might be a sign of ovarian cancer, how long would it take you to go to the doctors from the time from first noticed the symptom?

- ( ) Immediately, no wait
- ( ) Up to 1 week
- ( ) 1-2 weeks
- ( ) 2-4 weeks
- ( ) More than a month

Next

In the next few pages, we will describe four key characteristics of ovarian cancer tests:

1. **Accuracy**
2. **Identifiable conditions**
3. **Time to diagnosis**
4. **Communication throughout the testing process**

Later we will ask you to choose between tests described in terms of these four different characteristics.

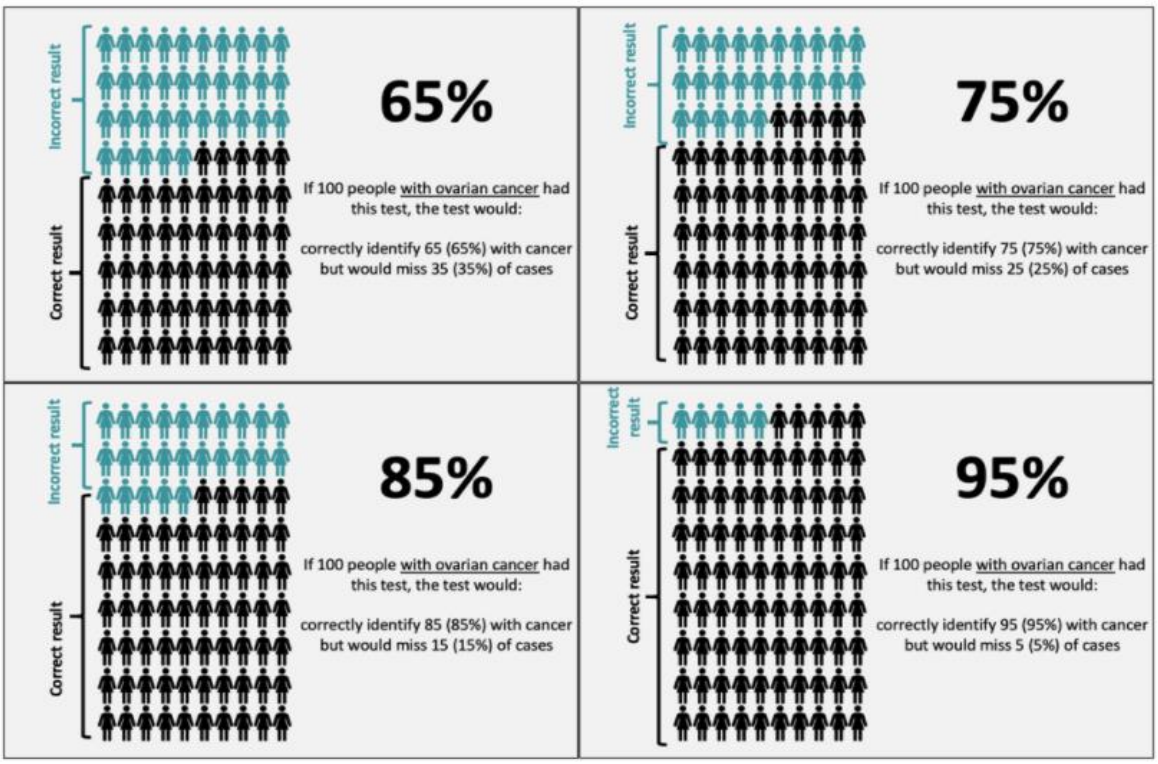
1. Accuracy

The tests for ovarian cancer differ in terms of how 'accurate' or 'reliable' the results are.

The accuracy of a test refers to how often a test correctly gives a positive result for people who has the disease being tested for.

An accurate test will identify almost everyone who has ovarian cancer and not miss many cases.

A less accurate test will generate more "false-negative" results. This is where a person with cancer receives a negative test result.



2. Identifiable conditions

The tests also vary in the conditions they can find and diagnose. Most women who are tested will not have ovarian cancer. Some tests can only find ovarian cancer. Other tests can identify ovarian cancer or other causes of the symptoms, such as ovarian cysts or fibroids (non-cancerous growths).

To summarise, in this survey you will be asked to choose between tests that can detect:

1. Ovarian cancer only
2. Ovarian cancer and alternative conditions that explain your symptoms

Versions where time to diagnosis has no impact on survival:

3. Time to diagnosis

This characteristic considers the total length of time from the test being ordered by your GP to being given a final diagnosis if you have cancer. This includes the waiting time the first test, the time waiting for the results and any follow up tests or appointments required. This may differ substantially between tests. In this survey we consider four lengths of time:

- 1 month
- 2 months
- 3 months
- 4 months

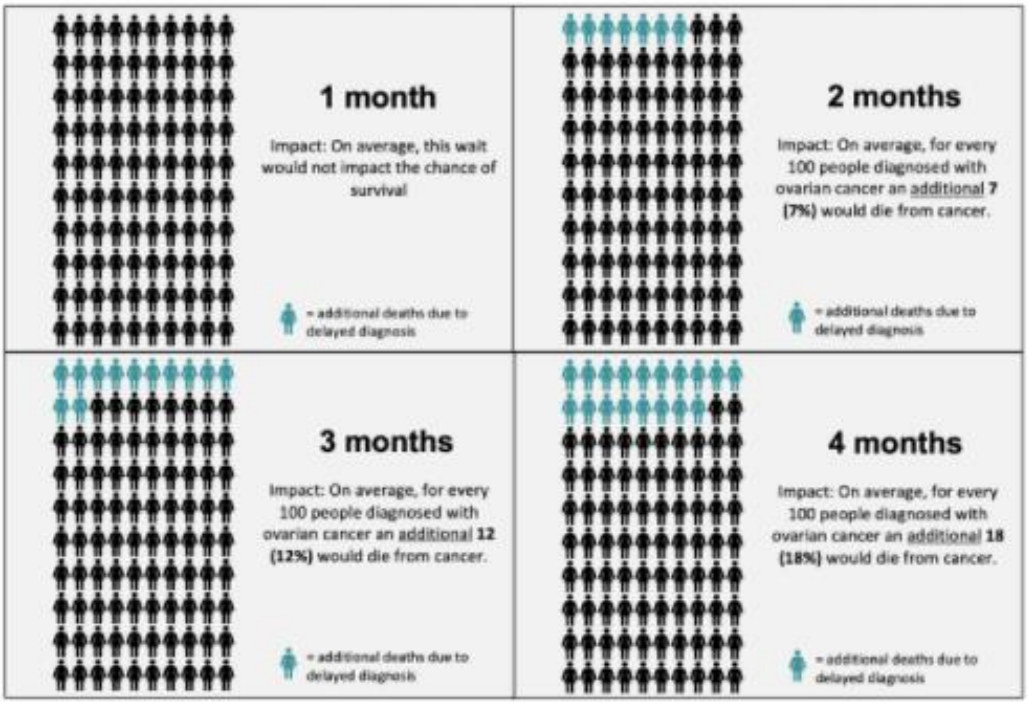
The impact of delays in diagnosis are not fully understood at present. For this survey, please imagine after one month, **each additional month waiting for a diagnosis results in a 4% decrease in the chance of surviving 5 years after being diagnosed with ovarian cancer, on average (or please imagine that delays in diagnosis have no impact on the chance of surviving if you are diagnosed with cancer. Additional time waiting for results may lead to uncertainty and worry).**

Versions where time to diagnosis has an age-adjusted impact on survival (associated 10-year survival probabilities shown were changed based on the respondent's age:

3. Time to diagnosis

This characteristic considers the total length of time from the test being ordered by your GP to being given a final diagnosis if you have cancer. This includes the waiting time for the first test, the time waiting for the results and any follow up tests or appointments required. This may differ substantially between tests.

In April 2020, NHS England introduced the Faster Diagnosis Standard which aims to ensure all patients referred for investigation of suspected cancer find out, within 1 month (28 days), if they do or do not have cancer. Additional waiting times beyond 1 month can be considered a delay in diagnosis and may mean the chance of a person surviving their cancer is reduced.
In this survey we will ask you to consider four lengths of time, each with an associated reduction in the chance of survival:



4. Communication throughout the testing process

Communication throughout the testing process

This describes the quality of communication throughout the testing process. This means from the time you first speak to your GP about your symptoms, undergoing the test, waiting and receiving your results. There are three possible levels of communication within this study:

1. Good
2. Fair
3. Poor

Communication can include many different aspects such as:

- Ability of doctors and nurses to listen and explain things clearly
- Continuity of care (refers to the communication and coordination between the different health care providers you encounter during testing e.g. your GP, nurses and specialists)
- Being treated as an individual
- Availability of support

The preferred communication style may differ from person to person.

Please use your own view of what you would consider “good”, "fair" and “poor” communication when thinking about this characteristic.

Please briefly explain, what 'good' communication from a doctor or health care provider means to you:

Part 2: Preferences towards ovarian cancer testing

We would like you to imagine that you have gone to your GP because you are regularly feeling bloated. Upon examination, your doctor tells you that they think you also have a build-up of gas or fluid in your stomach (abdominal distension). Your doctor explains that 1 in 33 (3%) women with similar symptoms have ovarian cancer. Your doctor suggests you may wish to be tested for ovarian cancer.

We will now ask you a set of 20 questions. In each question, you will be shown two different ovarian cancer tests. Each test will be described by the characteristics you have just read about. We would like to know which test you would prefer to have, if you were given a choice.

Please base your choices on the on the information presented. The tests are identical in all other ways.

There are no right or wrong answers.

DCE questions here: 16 questions from experimental design below. Plus 4 validity check questions. Question formats for each survey version are shown in section C13. Underlying experimental design is the same in all 4 versions.

What characteristics did you base your choices on?

- ☐ Accuracy
- ☐ Identifiable conditions
- ☐ Time to completion of testing
- ☐ Communication skills of health care providers

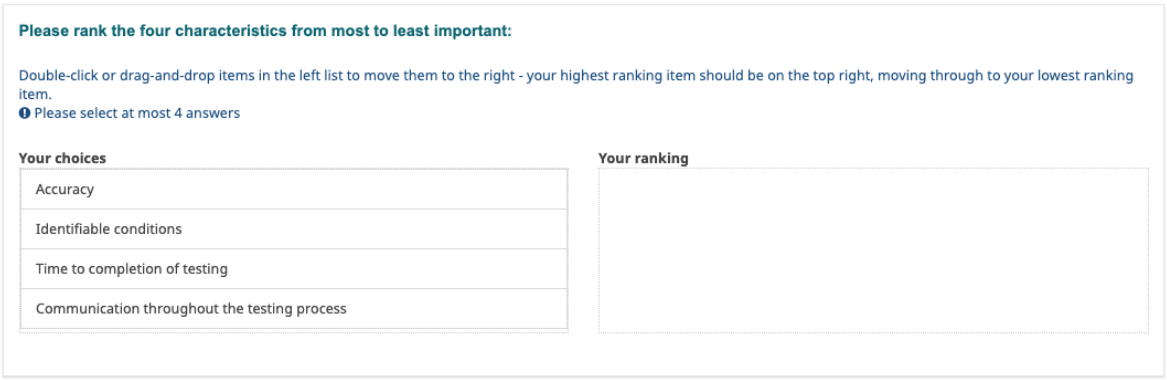
If you based your choices on only some of the characteristics, why?

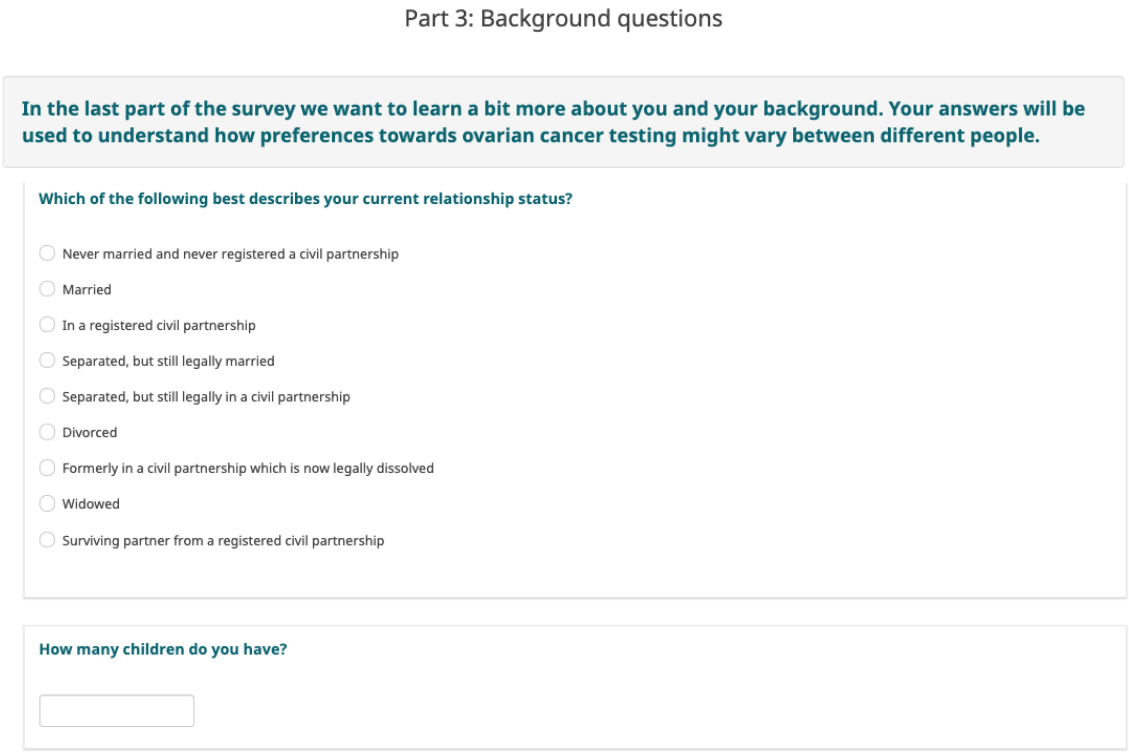
- ☐ I considered all characteristics
- ☐ There were too many characteristics to look at
- ☐ The other characteristics were unclear
- ☐ The other characteristics were not important to me
- ☐ Other: _____

In one or more of the questions you indicated that you liked both tests equally. What did you mean when you chose this option?

- ○ I liked the two options exactly the same
- ○ I thought both options were good or I would be happy to have either test
- ○ I could not tell the difference between the two tests
- ○ Other: _____

Ranking question was randomised to before and after the choice tasks to control for position bias





What is the total annual income of your household (before tax)?

- ○ Prefer not to say
- ○ £0-£9,999
- ○ £10,000-£19,999
- ○ £20,000-£29,999
- ○ £30,000-£39,999
- ○ £40,000-£49,999
- ○ £50,000-£59,999
- ○ £60,000-£69,999
- ○ £70,000 or more

What is the highest level of education you have completed?

Please choose...

Part 3: Background questions

How is your health in general?

- ○ Very good
- ○ Good
- ○ Fair
- ○ Bad
- ○ Very bad
- ○ Prefer not to say

Compared to the average woman of your age, how would you describe your risk of developing ovarian cancer?

- ○ Very high risk
- ○ High risk
- ○ Average risk
- ○ Low risk
- ○ Very low risk
- ○ Prefer not to say
- ○ Don't know

To what extent do you worry about your risk of ovarian cancer?

- ○ Not at all
- ○ A little
- ○ A moderate amount
- ○ A lot
- ○ A great deal
- ○ Prefer not to say

On average, how often do you visit your GP every year? Please enter 'yes' to show that you are paying attention

Have you ever been diagnosed with any type of cancer?

- ○ Yes
- ○ No
- ○ Prefer not to say

What type of cancer were you diagnosed with?

To your knowledge, have any of your family or friends been diagnosed with ovarian cancer?

- ○ Yes
- ○ No
- ○ Prefer not to say

Was this person your:

ℹ Check all that apply

- ☐ Blood relative
- ☐ Non-blood relative
- ☐ Friend
- ☐ Acquaintance/work colleague
- ☐ Prefer not to say
- ☐ Other:

Next

Part 3: Background questions

I see myself as someone who...

	Disagree strongly	Disagree a little	Neither agree or disagree	Agree a little	Agree strongly
...is reserved	○	○	○	○	○
...is generally trusting	○	○	○	○	○
...tends to be lazy	○	○	○	○	○
...is relaxed, handles stress well	○	○	○	○	○
...please choose "agree strongly"	○	○	○	○	○
...has few artistic interests	○	○	○	○	○
...is outgoing, sociable	○	○	○	○	○
...tends to find fault in others	○	○	○	○	○
...does a thorough job	○	○	○	○	○
...gets nervous easily	○	○	○	○	○
...has an active imagination	○	○	○	○	○

How much confidence and trust do you have in general practitioners (GPs)?

- ○ A great deal
- ○ A lot
- ○ A moderate amount
- ○ A little
- ○ None at all

When seeking help for medical issues, how much do you *wish to* be able to be involved in decisions about the medical process?

- ○ A great deal
- ○ A lot
- ○ A moderate amount
- ○ A little
- ○ Not at all

When seeking help for medical issues, how much do you *feel able* to be involved in medical decisions about the medical process?

- ○ A great deal
- ○ A lot
- ○ A moderate amount
- ○ A little
- ○ Not at all

In general, how willing are you to take risks?

Not confident at all — 5 — Extremely confident

1 10

Submit

Thank you for completing this survey.

Important: Please click here to return to prolific and confirm your submission

---

Your responses to this survey will add to a body of research which we hope will help to improve the experiences of women being tested and diagnosed with cancer in primary care.

We know that cancer is a sensitive subject and being asked to imagine the scenarios we have shown during the survey may have cause some people to feel anxious or concerned. Please speak to your GP if you are concerned about your risk of ovarian cancer.

The true impact of delays is not fully understood and there is a lot of mixed evidence. When answering the questions we asked you to imagine that a 1 week delay in diagnosis would reduce survival by 1% on average- this figure is among the more extreme estimates.

As with most cancers, early recognition of symptoms will help increase the chances of sucessful treatment. Being of aware of the symptoms will help you to spot them more easily.

Common symptoms of ovarian cancer include:

- feeling constantly bloated
- a swollen tummy
- discomfort in your tummy or pelvic area
- feeling full quickly when eating
- needing to pee more often than usual

Please contact your GP if you have any of these symptoms and do not go away. More information on ovarian cancer can be found on the NHS website: https://www.nhs.uk/be-clear-on-cancer/symptoms/ovarian-cancer

More information about ovarian cancer and the tests available can be found at :

Cancer Research UK: https://www.cancerresearchuk.org/about-cancer/ovarian-cancer

Target ovarian cancer: https://www.targetovariancancer.org.uk

The Eve Appeal: https://eveappeal.org.uk/gynaecological-cancers/ovarian-cancer/

If you are currently living with ovarian cancer and have been affected by any of the issues in this study, please speak to your clinical nurse specialist.

If you have any questions or concerns about the survey please contact Rebekah Hall at rh591@exeter.ac.uk or Prof Anne Spencer at A.E.Spencer@exeter.ac.uk

Postal address: University of Exeter Medical School, Room 1.15, South Cloisters, St Luke's Campus, Magdalen Road, City, Exeter, EX1 2LU

If you would like to be kept informed of the results from this study, please contact Rebekah Hall via your prolific account.

Appendix 7.1: Summary of questions relating to health attitudes and behaviours

	Version 1: 1%	Version 2: 2%	Version 3: 3%	Version 4: Timing-survival	P-value (ANOVA)
Self-reported overall health, n (%)					
Very good	32 (21%)	25 (17%)	31 (21%)	21 (14%)	0.55
Good	73 (49%)	78 (52%)	71 (47%)	81 (54%)	
Fair	40 (27%)	39 (26%)	38 (25%)	37 (25%)	
Poor	6 (3.3%)	8 (5%)	8 (5%)	11 (7%)	
Very poor	0 (0%)	0 (0%)	2 (1%)	0 (0%)	
Perceived risk of OC, n (%)					
Very low	5 (3%)	2 (1%)	4 (3%)	0 (0%)	0.84
Low	10 (7%)	15 (10%)	16 (11%)	13 (9%)	
Average	94 (63%)	96 (64%)	91 (61%)	95 (63%)	
High	10 (7%)	5 (3%)	11 (7%)	6 (4%)	
Very high	1 (1%)	2 (1%)	1 (1%)	1 (1%)	
Don't know	30 (20%)	29 (19%)	27 (18%)	35 (23%)	
Prefer not to say	-	1 (1%)	-	-	
OC-related worry, n (%)					
A great deal	5 (3%)	2 (1%)	2 (1%)	2 (1%)	0.63
A lot	6 (4%)	7 (5%)	5 (3%)	6 (4%)	
A moderate amount	20 (13%)	20 (13%)	25 (17%)	34 (23%)	
A little	72 (48%)	83 (55%)	69 (46%)	68 (45%)	
Not at all	47 (31%)	38 (25%)	49 (33%)	40 (27%)	
Personal history of cancer, n (%)					
Yes	9 (6%)	11 (7%)	14 (9%)	14 (9%)	0.18
No	141 (94%)	139 (93%)	136 (91%)	136 (91%)	
Know person diagnosed with OC, n (%)					
Yes	27 (18%)	23 (15%)	23 (15%)	36 (24%)	<0.001
No	123 (82%)	127 (85%)	127 (85%)	114 (76%)	
Previously tested for OC, n (%)					
Yes	13 (9%)	13 (9%)	8 (5%)	15 (10%)	0.20
No	137 (91%)	137 (91%)	142 (95%)	135 (90%)	
Previously TVUS (any reason), n (%)					
Yes	59 (39%)	47 (31%)	48 (32%)	51 (34%)	0.45
No	91 (61%)	103 (69%)	102 (68%)	99 (66%)	
Level of confidence/trust in GPs, n (%)					
A great deal	18 (12%)	22 (15%)	17 (11%)	11 (7%)	0.095
A lot	53 (35%)	58 (39%)	50 (33%)	46 (31%)	
A moderate amount	59 (39%)	49 (33%)	52 (35%)	71 (47%)	
A little	19 (13%)	19 (13%)	27 (18%)	20 (13%)	
None at all	1 (1%)	2 (1%)	4 (3%)	2 (1%)	
Willingness to take risks					
Mean (SD)	4.85 (2.2)	4.6 (2.2)	5.0 (2.2)	4.7 (1.9)	0.35

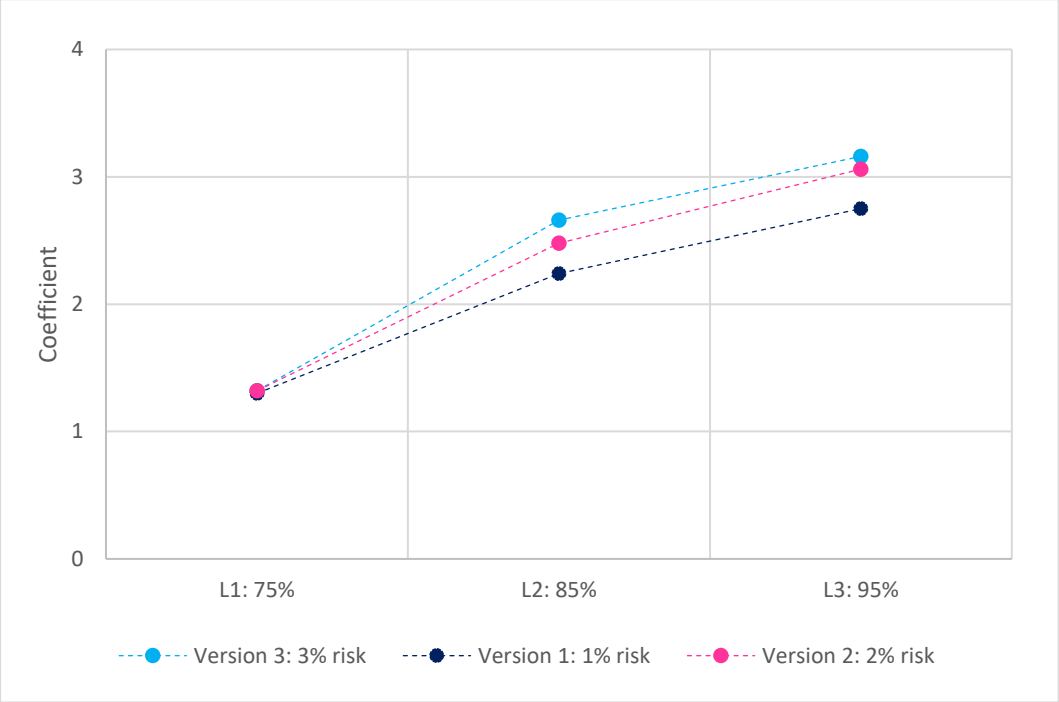
Appendix 7.2: Multinomial results for versions 1-3 using categorical dummy-coding parameters

	Version1: 1% risk of cancer		Version 2: 2% risk of cancer		Version 3: 3% risk of cancer	
	Coeff	Std error	Coeff	Std error	Coeff	Std error
Accuracy						
65%	Ref	-	Ref	-	Ref	-
75%	1.30***	0.16	1.32***	0.17	1.32***	0.18
85%	2.24***	0.21	2.48***	0.22	2.66***	0.22
95%	2.75***	0.19	3.06***	0.20	3.16***	0.22
Timing						
1 month	Ref	-	Ref	-	Ref	-
2 months	-0.17	0.14	-0.57***	0.15	-0.75***	0.15
3 months	-0.70***	0.18	-0.66***	0.16	-0.93***	0.16
4 months	-1.26***	0.13	-1.26***	0.13	-1.57***	0.14
Identifiable conditions						
Cancer only	Ref	-	Ref	-	Ref	-
Cancer plus additional related conditions	0.83***	0.09	0.61***	0.09	0.62***	0.10
Communication						
Poor	Ref	-	Ref	-	Ref	-
Fair	0.74***	0.08	0.62***	0.07	0.70***	0.08
Good	0.89***	0.11	0.94***	0.11	1.00***	0.10
Neither test	0.04	0.27	-0.28	0.28	0.23	0.29
Model fit statistics						
LL	-1521.37		-1445.99		-1453.40	
LR test (Continuous vs categorical level coding)	-15.53		-11.73		-17.21	
Observations	6,840		6,918		6,840	
N	150		150		150	
Key: ***significant at 99% confidence level; **significant at 95% confidence level; *significant at 90% confidence level						

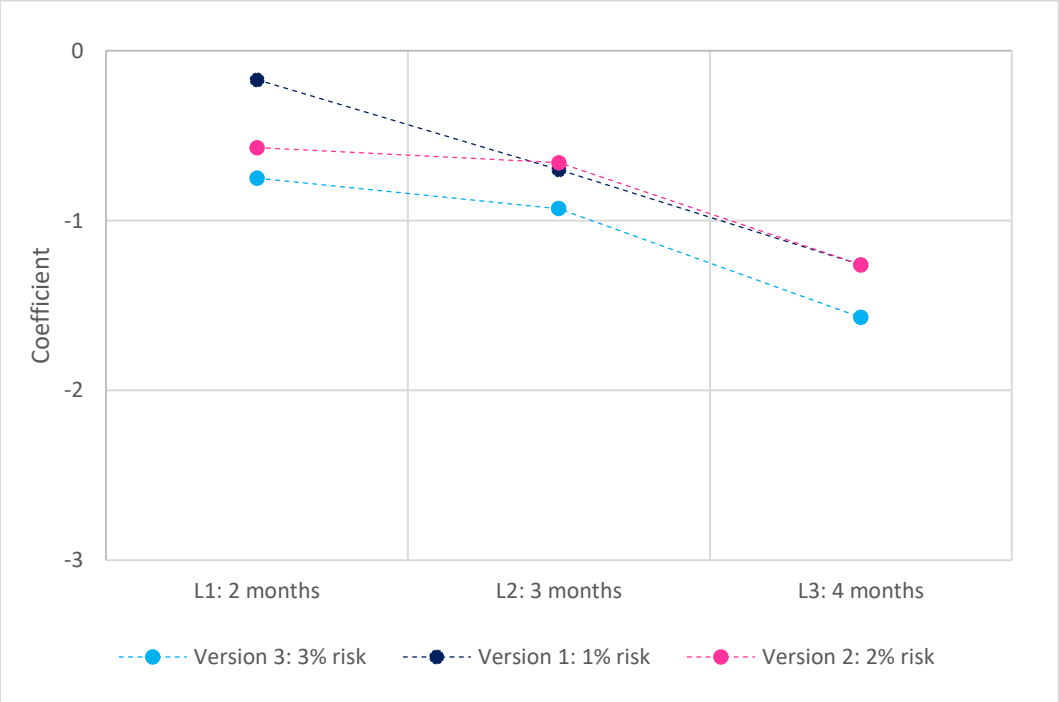
Appendix 7.3: Visualisations of Accuracy and Time to diagnosis categorically coded parameter estimates.

Estimates were used to assess the functional form of the final model

Accuracy:



Time to diagnosis:



Appendix 7.4: Interaction conditional logit model used to compare estimates from versions 1-3 (Differences in preferences based on risk of cancer)

	Coeff	Std error
Accuracy		
Per %	0.09***	0.01
Timing		
Per month	-0.37***	0.04
Identifiable conditions		
Cancer only	Ref	-
Cancer plus additional related conditions	0.77***	0.07
Communication		
Poor	Ref	-
Fair	0.73***	0.07
Good	0.81***	0.10
Neither test	-0.64***	0.25
Interaction terms- comparing preferences across risk levels		
accuracy_percentage2	0.01	0.01
timing_ percentage2	0.04	0.05
conditions_related_ percentage2	-0.01	0.10
communication_fair_ percentage2	-0.08	0.10
communication_good_ percentage2	0.13	0.14
accuracy_percentage3	0.01	0.01
timing_ percentage3	-0.03	0.05
conditions_related_ percentage3	-0.06	0.10
communication_fair_ percentage3	-0.03	0.11
communication_good_ percentage3	-0.10	0.13
Model fit statistics		
LL	-4445.08	
Observations	20,598	
N	450	

Appendix 7.5: Sensitivity analysis- 10 minutes or less

Interaction model used to check for differences in preferences between those who completed the survey under 10 minutes and those who took longer

	Coeff	Std error
Accuracy		
Per %	0.10***	0.00
Timing		
Per month	-0.36***	0.02
Identifiable conditions		
Cancer only	Ref	-
Cancer plus additional related conditions	0.75***	0.04
Communication		
Poor	Ref	-
Fair	0.69***	0.04
Good	0.89***	0.06
Neither test	-0.54***	0.16
Interaction terms- time below 10 minutes		
accuracy_under10	-0.00	0.02
timing_under10	-0.20	0.14
communication_fair_under10	0.12	0.18
communication_good_under10	-0.23	0.19
conditions_related_under10	-0.24	0.18
neither_under10	-0.89	0.89
Model fit statistics		
LL	-4449.67	
Observations	20,698	
N	450	

Appendix 7.6: Sensitivity analysis- rationality failures

Interaction model used to check for differences in preferences between those who failed one or more rationality check and those who did not

	Coeff	Std error
Accuracy		
Per %	0.10***	0.00
Timing		
Per month	-0.37***	0.03
Identifiable conditions		
Cancer only	Ref	-
Cancer plus additional related conditions	0.78***	0.05
Communication		
Poor	Ref	-
Fair	0.77***	0.05
Good	0.94***	0.07
Neither test	-0.59***	0.15
Interaction terms- failing 1 or more rationality checks		
accuracy_rationalityfail	-0.03***	0.01
timing_rationalityfail	0.00	0.05
conditions_related_rationalityfail	-0.13	0.10
communication_fair_rationalityfail	-0.23**	0.10
communication_good_rationalityfail	-0.16	0.13
Model fit statistics		
LL	-4410.83	
Observations	20,598	
N	450	

Appendix 7.7: Logistic regression investigating the probability of failing one or more rationality check question

	Odds ratio	Std error
Task difficulty	0.90	0.21
Completion time	0.99	0.00
Attended university	0.55***	0.13
Certainty of choices for rationality questions	0.84***	0.06
Model fit statistics		
LL	-242	
Pseudo R^2	0.03	
N	450	

Appendix 7.8: Model fit statistics used to identify the appropriate number of classes for latent class logit analysis

Number of classes	Log-likelihood	BIC	AIC	CAIC	Lowest class probability
2	-3999.48	8090.57	8028.97	8105.57	16%
3	-3627.42	7395.30	7300.84	7418.3	16%
4	-3566.30	7321.91	7194.59	7352.91	6%
5	-3430.31	7098.80	6938.63	7137.80	6%
6	-3363.14	7013.30	6820.27	7060.30	<5%

Appendix 7.9: Logistic regression to investigate the relationship between key characteristics and the probability of choosing the "neither test" alternative at least once

	Full model OR (SE)	Reduced model OR (SE)
Age	1.00 (0.01)	
Attended university	0.89 (0.22)	
Found task difficult/very difficult	0.89 (0.22)	
Previously tested for ovarian cancer	0.96 (0.31)	
Previously undergone a TVUS	0.96 (0.22)	
Employed	0.62* (0.16)	0.68* (0.16)
Risk averse	0.82 (0.19)	
Active role in medical decision making	0.85 (0.10)	
Wishes to have an active role in medical decision making	1.00 (0.14)	
Knows someone diagnosed with ovarian cancer	1.15 (0.37)	
Worried about ovarian cancer	0.56** (0.15)	0.60* (0.18)
In good/very good health	0.56** (0.15)	0.60** (0.14)
Aware of ovarian cancer symptoms	1.16 (0.30)	
Help-seeking	1.18 (0.26)	
White	0.64 (0.26)	
Children	0.86 (0.09)	
Confidence in choices	0.87 (0.08)	
Total completion time	1.00 (0.00)	
Constant	0.39 (0.45)	
Model fit statistics		
LL	-2035	-2053
Pseudo R^2	0.02	0.01
N	450	450

Key: ***significant at 99% confidence level
**significant at 95% confidence level
* significant at 90% confidence level

Appendix 8.1: Logistic regression examining the relationship between unstable preferences and the position of the stability check question within the survey.

Dummy variables represent the odds of failing the stability rationality check relative to instances where the stability check was positioned in the late stage of the DCE tasks.

	Odds ratio	Std error
Early choice stability dummy	1.56	0.45
Cross survey choice stability dummy	1.45	0.42
Constant	0.21***	0.05
Model fit statistics		
LL	-11253	
Pseudo R^2	0.01	
N	450	

Appendix 8.2: Regression analysis examining the relationship between the scale parameter and question order.

The model also includes average choice confidence and utility balance between alternatives for each choice task as independent variables to control for confounding effects.

	Odds ratio	Std error
Question order	0.01	0.03
Question order squared	0.00	0.00
Choice certainty	0.20**	0.07
Utility balance	-0.06**	0.02
Constant	0.46	0.45
Model fit statistics		
R^2	0.40	
N	450	
Key: ***significant at 99% confidence level; **significant at 95% confidence level; *significant at 90% confidence level		

Appendix 8.3: Mixed logit models analysing early and late stage responses separately.

Model outputs were used to calculate marginal rates of substitution

	Early responses		Late responses	
	Mean (SE)	Std dev	Mean (SE)	Std dev
Accuracy				
Per %	0.17*** (0.01)	0.09***	**0.24*** (0.02)**	0.16***
Timing				
Per month	**-0.62*** (0.05)**	0.52***	**-0.84*** (0.07)**	0.72***
Identifiable conditions				
Cancer only	Ref	-	Ref	-
Cancer plus additional related conditions	**1.11*** (0.10)**	0.91***	**1.18*** (0.11)**	1.10***
Communication				
Poor	Ref	-	Ref	-
Fair	**1.11*** (0.10)**	0.34**	**1.09*** (0.10)**	0.10
Good	**1.19*** (0.09)**	0.93***	**1.45*** (0.12)**	0.57***
Neither test	**-3.13*** (0.45)**	3.99***	**-3.06*** (0.42)**	4.37***
Model fit statistics				
LL	-1837.03		-1572.58	
Observations	10,266		10,266	
N	450		450	

Key: ***significant at 99% confidence level; **significant at 95% confidence level; *significant at 90% confidence level

Appendix 8.4: Description of socio-demographic and experimental characteristics included in the logit model investigating associations with indifferent preferences

Category	Description	Hypothesis
Socio-demographic factors		
Education	• Attended university • Did not attend university	Increased education is associated with reduced selection of indifference option
Self-reported health	• Good health • Average/below good health	Those in poorer health may be more inclined to default to doctor recommendations resulting in more passive preferences and greater likelihood of indifferent responses
Ovarian cancer testing experience	• Previously tested • Never tested	Those with previous test experience are likely to have better formed, pre-existing preferences relating to testing resulting in a reduction in indifferent preferences
Current medical decision-making role	• Active (Report having "a great deal" or "a lot" of input in medical decisions) • Passive (Report having a lesser role in medical decisions)	Those who currently play an active role in medical decisions are likely to have clearer formed preferences and be more confident when making choices resulting in reduced indifferent responses
Desired medical decision-making role	• Active (Report wanting "a great deal" or "a lot" of input in medical decisions) • Passive (Report wanting a lesser role in medical decisions)	As above
Symptom awareness	• Continuous variable indicating the number of OC symptoms that were recognised (0-12)	Increased knowledge of ovarian cancer results in a reduction in indifferent preferences
Experiment-related factors		
Utility difference between alternatives	• Continuous variable representing the absolute utility difference between test alternatives within a task (i.e. Test A and Test B)	Larger differences in estimated utility between alternatives makes it easier to distinguish between alternatives resulting in reduced indifference
Choice confidence	• Continuous variance (1-10)	Less confidence in the choice within a DCE task is associated with increased chance of indifferent responses

Task difficulty	• Easy/neutral • Very difficult/difficult	Unclear- Increased difficultly may be associated with increased indifferent preferences due to difficulty in differentiating between alternative. Alternative, increased perceived difficulty may indicate greater deliberation and effort to choose a most preferred option
Total opt-out selections	• Continuous variable (0-16) representing the number of times the “neither test” alternative was chosen by a respondent throughout the choice tasks	Increased selection of the opt-out alternative is associated with reduced selection of indifference alternative

Appendix 8.5: Results of the HCL model used to examine scale differences between the DCE with and without an indifference alternative

Attribute	Coefficient	SE
Accuracy		
Per %	0.10***	.01
Time to diagnosis		
Per 1 month wait	-0.45***	0.03
Identifiable conditions		
Cancer only	-	-
Additional conditions	0.80***	0.06
Communication		
Poor	-	-
Fair	0.81***	0.07
Good	1.07***	0.08
Interactions		
No test	-0.75***	0.21
Scale covariates		
Indifference alternative included	-0.07	0.08
Model fit statistics		
Log-likelihood	-2915.46	
Observations	14,040	
N	300	
***=1% significant; **= 5% significant; *=10% significant		

Appendix 8.6: Sociodemographic characteristics for version 5- no indifference alternative

Characteristic	N (%)	p-value*
Age Mean (SD) Range	 51.8 (9.3) 40-81	0.67
Ethnicity, n (%) White Other Prefer not to say	 142 (95%) 8 (5%) -	0.11
Children, n (%) Mean (SD) Range	 1.5 (1.3) 0-8	0.58
Relationship status, n (%) Single In a relationship Married/civil partnership Separated/divorce Widowed Prefer not to say	 26 (17%) 25 (17%) 81 (54%) 16 (11%) 3 (1%) -	0.46
Education, n (%) No qualifications GCSE A-Level/ College Undergraduate Post-graduate/professional quals Other Prefer not to say	 1 (1%) 28 (19%) 34 (23%) 46 (31%) 39 (26%) 2 (1%) -	0.98
Employment, n (%) Employed, full-time Employed, part-time Self-employed Not employed Retired Other Prefer not to say	 50 (33%) 26 (17%) 22 (15%) 14 (9%) 18 (12%) 38 (25%) -	0.38
Household income, n (%) £0-9,999 £10,000-19,999 £20,000-29,999 £30,000-39,999 £40,000- 49,999 £50,000- 59,999 £60,000-69,999 £70,000+ Prefer not to say	 13 (9%) 23 (15%) 25 (17%) 23 (15%) 16 (11%) 14 (9%) 10 (7%) 13 (9%) 13 (9%)	<0.001
*p-value of chi2 comparing characteristics between survey versions with and without an indifference alternative		

Characteristic	N (%)	p-value
Self-reported overall health		0.24
Very good	19 (13%)	
Good	75 (50%)	
Fair	48 (32%)	
Poor	7 (5%)	
Very poor	1 (1%)	
Perceived risk of OC		0.57
Very low	-	
Low	11 (7%)	
Average	103 (69%)	
High	4 (3%)	
Very high	-	
Don't know	32 (21%)	
Prefer not to say	-	
OC-related worry		0.58
A great deal	1 (1%)	
A lot	5 (3%)	
A moderate amount	19 (13%)	
A little	77 (51%)	
Not at all	48 (32%)	
Personal history of cancer		0.20
Yes	9 (6%)	
No	141 (94%)	
Prefer not to say	1 (1%)	
Know person diagnosed with OC		<0.001
Yes	30 (20%)	
No	120 (80%)	
Previously tested for OC		<0.001
Yes	10 (7%)	
No	140 (93%)	
Current level of input in medical decisions		0.57
A great deal	10 (7%)	
A lot	32 (21%)	
A moderate amount	57 (38%)	
A little	40 (27%)	
Not at all	11 (7%)	
Desired level of input in medical decisions		0.49
A great deal	66 (344%)	
A lot	48 (32%)	
A moderate amount	32 (21%)	
A little	4 (3%)	
Not at all	0 (0%)	
Willingness to take risks		
Mean (SD)	4.59 (2.3)	0.13
*p-value of chi2 comparing characteristics between survey versions with and without an indifference alternative		

Appendix 8.7: Mixed logit results for the non-indifference survey version

	Mean (SE)	Std dev
Accuracy		
Per %	0.19*** (0.01)	0.11***
Timing		
Per month	**-0.86*** ** (0.08)	0.06***
Identifiable conditions		
Cancer only	Ref	-
Cancer plus additional related conditions	**1.29*** ** (0.11)	0.91***
Communication		
Poor	Ref	-
Fair	**1.43*** ** (0.12)	0.26
Good	**1.82*** ** (0.15)	0.76***
Neither test	**-4.47*** ** (072)	3.68***
Model fit statistics		
LL	-1117.32	
Observations	7,200	
N	150	

Key: ***significant at 99% confidence level; **significant at 95% confidence level; *significant at 90% confidence level

Appendix 9.1: Detailed BWS methods and results assessing ovarian screening priorities

This appendix provides an extended description of the methods and results of the online best-worst scaling survey conducted as part of attribute development for the ovarian cancer screening DCE developed in Chapter 9.

Methods

Object-case best-worst scaling (BWS) was used to understand the relative importance of the remaining potential attributes to the target population.

Experimental design

A BIBD was generated in SAS 9.4. Nine potential designs consisting of 13-39 tasks each with 3-12 attributes per task were generated and compared. The design was selected based on a balance of efficiency and perceived respondent burden. The final design consisted of 13 choice tasks. Each choice task included a sub-set of four attributes from which participants were asked to select the "most important" and least important". Each attribute appeared four times across the choice tasks. The order of choice sets was randomised between respondents to control for ordering effects. The position of attributes was optimised so each attribute appeared in every position once throughout the survey. Overall, the experimental design had a d-efficiency of 81.3%.

Survey instrument

The best-worst scaling study was embedded into an online survey hosted on Limesurvey (Limesurvey.com).. Alongside the best-worst questions, the survey also included socio-demographic questions (e.g. age, employment status) and health behaviour questions (e.g. ovarian cancer knowledge, family history and previous screening uptake).

To ensure no relevant attributes were missed and mediate the influence of the internal reduction exercise, an open-ended question was included asking participants if there were any additional attributes they would consider important when making a decision about ovarian cancer screening.

Participants

Participants were recruited using Prolific (Prolific.co). Participation was limited to women and people with ovaries (restricted using the 'sex' filter) over the age of 40 (no upper limit), living in England and Wales. The minimum sample size estimated alongside the BIBD was 52, given

the uncertainty surrounding this estimated a target sample size of 100 was chosen[11]. Participants received a payment of £2.00 via prolific account as compensation for their time.

Statistical analysis

Sociodemographic and health-related background questions were reported using means, standard deviations, and ranges.

BWS data were analysed using two methods: counting approach and multinomial logit.

Counting approach

Counting scores were calculated by subtracting the number of times an item was selected as "least important" from the number of times the attribute was selected as "most important" across all choice tasks and all respondents. Raw counting scores ranged between -400 and +400 (each attribute appeared 4 time and there were 100 respondents), with a higher score indicating greater importance. To aid interpretation, scores were standardised to between -1 and +1.

Multinomial logit (MNL)

A MNL model was estimated using the lowest scoring attribute from the counting analysis as the reference item for analysis. The level of importance of all remaining attributes were therefore estimated relative to this attribute.

Results

Table 1 summarises the demographic and health-related characteristics of respondents. The average age of respondents was 50 years old (range: 40-72). The majority of respondents were white (93%), married (65%), university-educated (53%) and employed (56%).

Respondents most commonly reported being in good or very good health (69%) (Table 2). Anxiety surrounding personal-risk of ovarian cancer was generally low, with 57% of respondents reporting feeling little-to-no ovarian-cancer related worry. Previous testing for ovarian cancer was low (9%). Screening attendance across respondents was high with 81% reporting always attending cervical screening when invited (or always attending before

[11] When generating BIBD in SAS an estimated minimum sample size is provided, however, it is unclear how this estimate is derived. Currently there is no guidance on appropriate sample size for best-worst scaling studies.

reaching 65 years old). Breast cancer screening attendance was lower, with 70% (33/47) of eligible participants reporting routinely attending.

Table 1: Sociodemographic characteristics of BWS respondents

Characteristic	n
Age	
Mean (SD)	50.3 (8.7)
Range	40-72
Ethnicity	
White	93
Asian	2
Black	2
Mixed	2
Not reported	1
Relationship status	
Single	10
In a relationship	7
Married/civil partnership	65
Divorced/separated	15
Widowed	3
Level of education	
GCSE	27
A-Level/ College	20
Undergraduate	27
Post-graduate	26
Employment status	
Employed, full-time	30
Employed, part-time	26
Self-employed	13
Not employed	3
Retired	13
Other	12
Income	
£0-£9,999	5
£9,999-£19,999	9
£20,000-£29,999	22
£30,000-£39,999	15
£40,000-£49,999	8
£50,000-£59,999	9
£60,000-£69,999	9
£70,000+	10
Not reported	13

Table 2: Health related characteristics of BWS respondents (n=100)

Characteristic	n (%)
Self-reported overall health	
Very good	14
Good	55
Fair	24
Poor	5
Very poor	2
Ovarian cancer-related worry	
A great deal	3
A lot	9
A moderate amount	27
A little	47
Not at all	10
Not reported	4
Personal history of cancer	10
Knew someone who was diagnosed with ovarian cancer	21
Previously tested for ovarian cancer	9
Previously undergone a TVUS (any reason)	37
Cervical cancer screening behaviour	
Always attends/attended	81
Irregularly attends/attended	11
Never attended/stopped attending	8
Breast cancer screening behaviour	
Always attends/attended	33
Irregularly attends/attended	6
Never attended/stopped attending	8
Not eligible	47
How much confidence and trust do you have in GPs	
A great deal	12
A lot	40
A moderate amount	40
A little	6
None at all	2
How much do you feel able to be involved in medical decisions	
A great deal	10
A lot	31
A moderate amount	34
A little	21
Not at all	4
How much do you wish to be involved in medical decisions	
A great deal	38
A lot	42
A moderate amount	14
A little	5
Not at all	0

Best worst scaling results

A summary of aggregate best and worst selections for each attribute are shown in Table 3

Table 3: Best-worst counting scores

Attribute	Aggregate population				Individual mean	
	Rank	Most	Least	M-L	Standardised M-L score	95% CI
Chance of dying from ovarian cancer	1	273	15	258	0.65	0.53 –0.76
Accuracy: chance of **false-negative** result	2	186	12	174	0.44	0.36 –0.51
Chance of being unnecessarily diagnosed and treated for a cancer that would never have caused symptoms or death	3	176	9	167	0.42	0.33 –0.51
Chance of cancer diagnosis	4	157	23	134	0.34	0.26 –0.41
Accuracy: chance of a **false-positive** result	5	137	10	127	0.32	0.27 –0.37
Waiting time for the result	6	74	86	-12	-0.03	-0.03 – (-0.03)
Chance of needing a follow up test	7	59	95	-36	-0.09	-0.16 – (-0.02)
Waiting time for the test	8	60	105	-45	-0.11	-0.12 – (-0.10)
Screening interval	9	28	101	-73	-0.18	-0.25 – (-0.11)
Type of test	10	55	168	-113	-0.28	-0.40 – (-0.16)
Test location	11	41	229	-188	-0.47	-0.58 – (-0.36)
Who performs the test?	12	28	217	-189	-0.47	-0.61 – (-0.33)
Action required by you to arrange the result	13	26	230	-204	-0.51	-0.62 – (-0.40)

Counting results

"Chance of dying of ovarian cancer" was deemed most important on average with a standardised score of 0.65 (95% CI: 0.53-0.76) and "Action required by you to arrange the test" was deemed least important (-0.51, 95% CI:-0.62 – [-0.40]). The distance between each attribute spatially represents the relative importance of each attribute. Heterogeneity in best and worst selections across respondents varied between attributes Figure 1). However, overall heterogeneity was lower than observed in the diagnostic survey, as reflected by the smaller confidence intervals associated with importance scores.

Importance scores formed two distinct groups. Attributes in cluster 1 were the most important to respondents are related to the performance characteristics of tests. Attributes in the second cluster were distinctly less important to respondents and related to service delivery aspects of screening (except for "chance of needing a follow up test").

Multinomial logit results

MNL results are shown in Table 4. As the lowest scoring attribute during the counting analysis, "Action required by you to arrange the test" was selected as the base attribute and set to zero. With the exception of "Test location" and "Who performs the test", all attributes were significantly more important to respondents than the base attribute. The order of attribute importance remained largely unchanged, however, "Accuracy: chance of a **false positive** result" and "Chance of cancer diagnosis" alternated positions. When mapped, MNL coefficients and counting coefficients were highly correlated Figure 2.

Figure 1: Range of individual scores from the best-worst scaling study

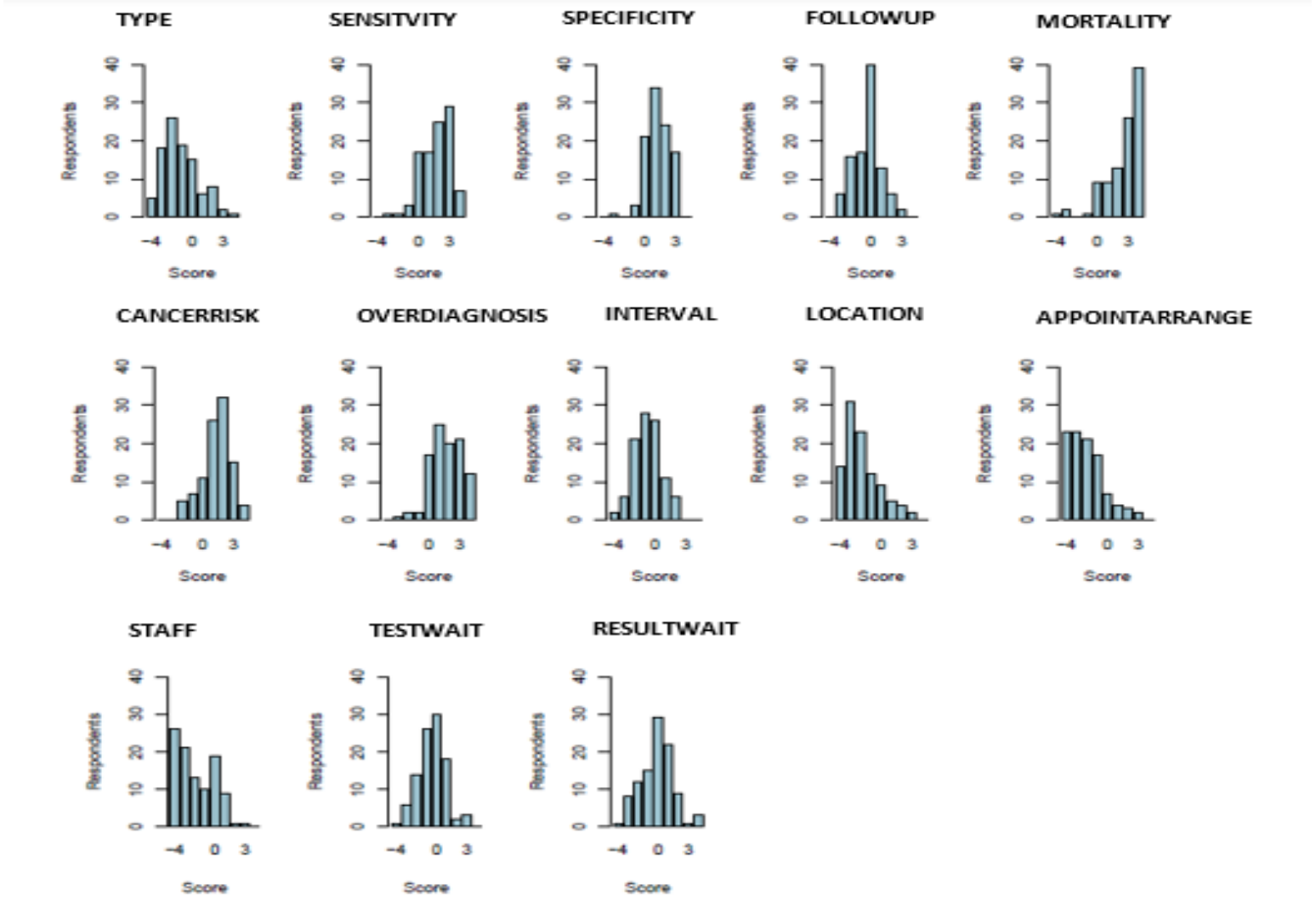
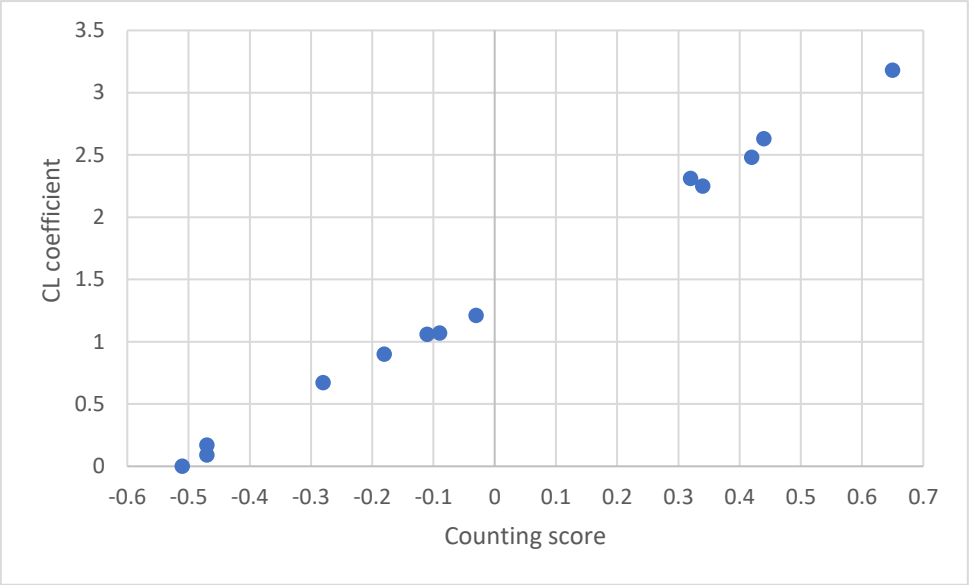


Table 4: Conditional logit BWS results using “Action required by you to arrange the test” as the base attribute

Attribute	Coefficient	SE
Chance of dying from ovarian cancer	3.18***	0.14
Accuracy: chance of **false-negative** result	2.63***	0.13
Chance of being unnecessarily diagnosed and treated for a cancer that would never have caused symptoms or death	2.48***	0.13
Accuracy: chance of a **false-positive** result	2.31***	0.13
Chance of cancer diagnosis	2.25***	0.13
Waiting time for the result	1.21***	0.12
Chance of needing a follow up test	1.07***	0.12
Waiting time for the test	1.06***	0.12
Screening interval	0.90***	0.12
Type of test	0.67***	0.12
Test location	0.17	0.12
Who performs the test?	0.09	0.12
***=p-value<0.01		

Figure 2: Plot demonstrating the correlation between BWS responses analysed using the counting method and conditional logit modelling



Self-reported importance of attributes

Following completion of the best-worst task, respondents were asked to indicate the attributes that they would consider "always important" when making ovarian cancer screening decisions and any that they would consider as "never important". Figure 3 shows the responses for each attribute plotted against the total number of "most" and "least" selections from the BWS exercise for comparison. Self-reported importance generally followed the same pattern and re-demonstrate the strong prioritisation of attributes falling in to "cluster 1" in the best-worst task.

Figure 3: Number of respondents that indicated each attribute was "always important" or "never important" to them when considering ovarian cancer screening

Adamowicz, W., Boxall, P., Williams, M., & Louviere, J. (1998). Stated preference approaches for measuring passive use values: choice experiments and contingent valuation. *American Journal of Agricultural Economics, 80*(1), 64-75.

Adamowicz, W., Louviere, J., & Williams, M. (1994). Combining revealed and stated preference methods for valuing environmental amenities. *Journal of environmental economics and management, 26*(3), 271-292.

Aizaki, H. (2014). support. BWS: Basic functions for supporting an implementation of best–worst scaling, R package version 0.1-1. In: URL http://CRAN. R-project. org/package= support. BWS.

Akintomide, A. O., & Obasi, U. O. (2019). Intimate patient examinations: The awareness, acceptance and practice preference of transvaginal ultrasound scan among women in a South-southern State of Nigeria. *Journal of Family Medicine and Primary Care, 8*(1), 109.

Akl, E. A., Oxman, A. D., Herrin, J., Vist, G. E., Terrenato, I., Sperati, F., . . . Schünemann, H. (2011). Using alternative statistical formats for presenting risks and risk reductions. *Cochrane Database of Systematic Reviews*(3).

Alemu, M. H., Mørkbak, M. R., Olsen, S. B., & Jensen, C. L. (2013). Attending to the reasons for attribute non-attendance in choice experiments. *Environmental and resource economics, 54*(3), 333-359.
Ali, S., & Ronaldson, S. (2012). Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. *British Medical Bulletin, 103*(1), 21-44. doi:10.1093/bmb/lds020
Ancker, J. S., Senathirajah, Y., Kukafka, R., & Starren, J. B. (2006). Design features of graphs in health risk communication: a systematic review. *Journal of the American Medical Informatics Association, 13*(6), 608-618.
Andermann, A., Blancquaert, I., Beauchamp, S., & Déry, V. (2008). Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bulletin of the World Health Organization, 86*, 317-319.
Arana, J. E., Leon, C. J., & Quevedo, J. L. (2006). The effect of medical experience on the economic evaluation of health policies. A discrete choice experiment. *Social Science and Medicine, 63*(2), 512-524. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1016/j.socscimed.2005.12.008
Armstrong, N., Herbert, G., Aveling, E. L., Dixon-Woods, M., & Martin, G. (2013). Optimizing patient involvement in quality improvement. *Health Expect, 16*(3), e36-47. doi:10.1111/hex.12039
Arora, N., Quaife, M., Hanson, K., Lagarde, M., Woldesenbet, D., Seifu, A., & Crastes dit Sourd, R. (2022). Discrete choice analysis of health worker job preferences in Ethiopia: Separating attribute non-attendance from taste heterogeneity. *Health economics, 31*(5), 806-819.
Atalabi, O. M., Morhason-Bello, I. O., Adekanmi, A. J., Marinho, A. O., Adedokun, B. O., Kalejaiye, A. O., . . . Gbadamosi, S. A. (2012). Transvaginal ultrasonography: a survey of the acceptability and its predictors among a native African women population. *International Journal of Women's Health, 4*, 1.
Bahamonde-Birke, F. J., Navarro, I., & de Dios Ortúzar, J. (2017). If you choose not to decide, you still have made a choice. *Journal of choice modelling, 22*, 13-23.
Baldwin, L.-M., Trivers, K. F., Matthews, B., Andrilla, C. H. A., Miller, J. W., Berry, D. L., . . . Goff, B. A. (2012). Vignette-based study of ovarian cancer screening: do US physicians report adhering to evidence-based recommendations? *Annals of internal medicine, 156*(3), 182-194.
Banks, J., Hollinghurst, S., Bigwood, L., Peters, T. J., Walter, F. M., & Hamilton, W. (2014). Preferences for cancer investigation: a vignette-based study of primary-care attendees. *The Lancet Oncology, 15*(2), 232-240.
Bansal, N., Bhopal, R., Steiner, M., & Brewster, D. H. (2012). Major ethnic group differences in breast cancer screening uptake in Scotland are not extinguished by adjustment for indices of geographical residence, area deprivation, long-term illness and education. *British Journal of Cancer, 106*(8), 1361-1366.
Bansback, N., Brazier, J., Tsuchiya, A., & Anis, A. (2012). Using a discrete choice experiment to estimate health state utility values. *Journal of health economics, 31*(1), 306-318.
Barnett, K., Mercer, S. W., Norbury, M., Watt, G., Wyke, S., & Guthrie, B. (2012). Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *The Lancet, 380*(9836), 37-43. doi:10.1016/S0140-6736(12)60240-2
Bateman, I. J., Burgess, D., Hutchinson, W. G., & Matthews, D. I. (2008). Learning design contingent valuation (LDCV): NOAA guidelines, preference learning and coherent arbitrariness. *Journal of environmental economics and management, 55*(2), 127-141.
Bech, M., Kjaer, T., & Lauridsen, J. (2011). Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. *Health economics, 20*(3), 273-286.
Bechel, D. L., Myers, W. A., & Smith, D. G. (2000). Does patient-centered care pay off? *Jt Comm J Qual Improv, 26*(7), 400-409. doi:10.1016/s1070-3241(00)26033-1
Bennett, K. F., Waller, J., Chorley, A. J., Ferrer, R. A., Haddrell, J. B., & Marlow, L. A. (2018). Barriers to cervical screening and interest in self-sampling among women who actively decline screening. *Journal of Medical Screening, 25*(4), 211-217.
Benning, T. M., Dellaert, B. G. C., Dirksen, C. D., & Severens, J. L. (2014a). Preferences for potential innovations in non-invasive colorectal cancer screening: A labeled discrete choice experiment for a Dutch screening campaign. *Acta Oncologica, 53*(7), 898-908. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.3109/0284186X.2013.877159
Benning, T. M., Dellaert, B. G. C., Severens, J. L., & Dirksen, C. D. (2014b). The effect of presenting information about invasive follow-up testing on individuals' noninvasive colorectal cancer screening participation decision: Results from a discrete choice experiment. *Value in health, 17*(5), 578-587. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1016/j.jval.2014.04.007

Berchi, C., Dupuis, J.-M., & Launoy, G. (2006). The Reasons of General Practitioners for Promoting Colorectal Cancer Mass Screening in France. *European Journal of Health Economics, 7*(2), 91-98. doi:https://link.springer.com/journal/volumesAndIssues/10198
Bien, D. R., Danner, M., Vennedey, V., Civello, D., Evers, S. M., & Hiligsmann, M. (2017). Patients' Preferences for Outcome, Process and Cost Attributes in Cancer Treatment: A Systematic Review of Discrete Choice Experiments. *Patient, 10*(5), 553-565. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1007/s40271-017-0235-y
Bilger, M., Özdemir, S., & Finkelstein, E. A. (2020). Demand for Cancer Screening Services: Results From Randomized Controlled Discrete Choice Experiments. *Value in health, 23*(9), 1246-1255.
Birch, S., Melnikow, J., & Kuppermann, M. (2003). Conservative versus aggressive follow up of mildly abnormal Pap smears: testing for process utility. *Health Economics, 12*(10), 879-884.
Bliemer, M. C., Rose, J. M., & Hensher, D. A. (2009). Efficient stated choice experiments for estimating nested logit models. *Transportation Research Part B: Methodological, 43*(1), 19-35.
Bliemer, M. C., Rose, J. M., & Hess, S. (2008). Approximation of Bayesian efficiency in experimental choice designs. *Journal of choice modelling, 1*(1), 98-126.
Board, S. (2009). Preferences and utility. *UCLA, Oct*.
Boland, A., Cherry, G., & Dickson, R. (2017). *Doing a systematic review: A student's guide*: Sage.
Boone, D., Mallett, S., Zhu, S., Yao, G. L., Bell, N., Ghanouni, A., . . . Halligan, S. (2013). Patients' & healthcare professionals' values regarding true- & false-positive diagnosis when colorectal cancer screening by CT colonography: Discrete choice experiment. *PLoS ONE, 8*(12), e80767. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1371/journal.pone.0080767
Boyle, K. J. (2003). Introduction to revealed preference methods. In *A primer on nonmarket valuation* (pp. 259-267): Springer.
Boyle, K. J., Holmes, T. P., Teisl, M. F., & Roe, B. (2001). A comparison of conjoint analysis response formats. *American journal of agricultural economics, 83*(2), 441-454.
Bradley, M., & Daly, A. (1994). Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. *Transportation, 21*(2), 167-184. doi:10.1007/BF01098791
Brazier, J., Akehurst, R., Brennan, A., Dolan, P., Claxton, K., McCabe, C., . . . Tsuchyia, A. (2005). Should patients have a greater role in valuing health states? *Applied health economics and health policy, 4*(4), 201-208.
Brazier, J., Ratcliffe, J., Saloman, J., & Tsuchiya, A. (2017). *Measuring and valuing health benefits for economic evaluation*: OXFORD university press.
Brenner, A., Howard, K., Lewis, C., Sheridan, S., Crutchfield, T., Hawley, S., . . . Pignone, M. (2014). Comparing 3 values clarification methods for colorectal cancer screening decision-making: A randomized trial in the US and Australia. *Journal of General Internal Medicine, 29*(3), 507-513. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1007/s11606-013-2701-0
Brewer, A. M., & Hensher, D. A. (2000). Distributed work and travel behaviour: the dynamics of interactive agency choices between employers and employees. *Transportation, 27*(1), 117-148.
Bridges, J. (2003). Stated preference methods in health care evaluation: an emerging methodological paradigm in health economics. *Applied health economics and health policy, 2*(4), 213-224.
Bridges, J. F., Hauber, A. B., Marshall, D., Lloyd, A., Prosser, L. A., Regier, D. A., . . . Mauskopf, J. (2011). Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value in health, 14*(4), 403-413.
Brouwer, W. B., Culyer, A. J., van Exel, N. J. A., & Rutten, F. F. (2008). Welfarism vs. extra-welfarism. *Journal of health economics, 27*(2), 325-338.
Brouwer, W. B., van Exel, N. J. A., van den Berg, B., van den Bos, G. A., & Koopmanschap, M. A. (2005). Process utility from providing informal care: the benefit of caring. *Health policy, 74*(1), 85-99.
Buchanan, J., Hill, M., Vass, C. M., Hammond, J., Riedijk, S., Klapwijk, J. E., . . . Hui, L. (2022). Factor's that impact on women's decision-making around prenatal genomic tests: An international discrete choice survey. *Prenatal Diagnosis*.
Bulliard, J. L., Beau, A. B., Njor, S., Wu, W. Y. Y., Procopio, P., Nickson, C., & Lynge, E. (2021). Breast cancer screening and overdiagnosis. *International Journal of Cancer, 149*(4), 846-853.
Buys, S. S., Partridge, E., Black, A., Johnson, C. C., Lamerato, L., Isaacs, C., . . . Kessel, B. (2011). Effect of screening on ovarian cancer mortality: the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening randomized controlled trial. *Jama, 305*(22), 2295-2303.

Bywall, K. S., Veldwijk, J., Hansson, M. G., Baecklund, E., Raza, K., Falahee, M., & Kihlbom, U. (2021). Does being exposed to an educational tool influence patient preferences? The influence of an educational tool on patient preferences assessed by a discrete choice experiment. *Patient Education and Counseling, 104*(10), 2577-2585. doi:https://doi.org/10.1016/j.pec.2021.03.013

Campbell, D., Boeri, M., Doherty, E., & Hutchinson, W. G. (2015). Learning, fatigue and preference formation in discrete choice experiments. *Journal of Economic Behavior & Organization, 119*, 345-363.

Campbell, D., & Erdem, S. (2019). Including opt-out options in discrete choice experiments: issues to consider. *The Patient-Patient-Centered Outcomes Research, 12*(1), 1-14.

Campbell, D., Hutchinson, W. G., & Scarpa, R. (2008). Incorporating discontinuous preferences into the analysis of discrete choice experiments. *Environmental and resource economics, 41*(3), 401-417.

Cancer Research UK. (2019). Ovarian Cancer Statistics Retrieved from https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/ovarian-cancer#heading-Two

Cancer Research UK. (2022). Ultrasound scan for ovarian cancer. Retrieved from https://www.cancerresearchuk.org/about-cancer/ovarian-cancer/getting-diagnosed/tests/ultrasound

Cantillo, V., Amaya, J., & Ortúzar, J. d. D. (2010). Thresholds and indifference in stated choice surveys. *Transportation Research Part B: Methodological, 44*(6), 753-763.

Carethers, J. M., Sengupta, R., Blakey, R., Ribas, A., & D'Souza, G. (2020). Disparities in Cancer Prevention in the COVID-19 EraCOVID-19 and Cancer Prevention. *Cancer prevention research, 13*(11), 893-896.

Carlsson, F., Kataria, M., & Lampi, E. (2010). Dealing with ignored attributes in choice experiments on valuation of Sweden's environmental quality objectives. *Environmental and resource economics, 47*(1), 65-89.

Carter, J. L., Coletti, R. J., & Harris, R. P. (2015). Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. *Bmj, 350*.

Castleberry, A., & Nolen, A. (2018). Thematic analysis of qualitative research data: Is it as easy as it sounds? *Currents in pharmacy teaching and learning, 10*(6), 807-815.

Caussade, S., de Dios Ortúzar, J., Rizzi, L. I., & Hensher, D. A. (2005). Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B: Methodological, 39*(7), 621-640.

Chamot, E., Mulambia, C., Kapambwe, S., Shrestha, S., Parham, G. P., Macwan'gi, M., & Mwanahamuntu, M. H. (2015). Preference for human papillomavirus-based cervical cancer screening: Results of a choice-based conjoint study in Zambia. *Journal of Lower Genital Tract Disease, 19*(2), 119-123. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1097/LGT.0000000000000081

Charvin, M., Launoy, G., & Berchi, C. (2020). The effect of information on prostate cancer screening decision process: a discrete choice experiment. *BMC health services research, 20*, 1-10.

Cheraghi-Sohi, S., Bower, P., Mead, N., McDonald, R., Whalley, D., & Roland, M. (2007). Making sense of patient priorities: applying discrete choice methods in primary care using 'think aloud' technique. *Fam Pract, 24*(3), 276-282. doi:10.1093/fampra/cmm007

Cheung, K. L., Wijnen, B. F., Hollin, I. L., Janssen, E. M., Bridges, J. F., Evers, S. M., & Hiligsmann, M. (2016). Using best–worst scaling to investigate preferences in health care. *Pharmacoeconomics, 34*(12), 1195-1209.

Chornokur, G., Amankwah, E. K., Schildkraut, J. M., & Phelan, C. M. (2013). Global ovarian cancer health disparities. *Gynecologic oncology, 129*(1), 258-264.

Coast, J., Al-Janabi, H., Sutton, E. J., Horrocks, S. A., Vosper, A. J., Swancutt, D. R., & Flynn, T. N. (2012). Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. *Health economics, 21*(6), 730-741.

Coast, J., Flynn, T., Sutton, E., Al-Janabi, H., Vosper, J., Lavender, S., . . . Peters, T. (2008). Investigating Choice Experiments for Preferences of Older People (ICEPOP): evaluative spaces in health economics. *Journal of health services research & policy, 13*(3_suppl), 31-37.

Coast, J., & Horrocks, S. (2007). Developing attributes and levels for discrete choice experiments using qualitative methods. *Journal of health services research & policy, 12*(1), 25-30.

Collins, A. (2012). Attribute nonattendance in discrete choice models: measurement of bias, and a model for the inference of both nonattendance and taste heterogeneity.

Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). Mathematical psychology: An elementary introduction.
Corso, P. S., Hammitt, J. K., & Graham, J. D. (2001). Valuing mortality-risk reduction: using visual aids to improve the validity of contingent valuation. *Journal of Risk and Uncertainty, 23*(2), 165-184.
Cowan Bennett, C., & Richards, D. (2000). Patient acceptance of endovaginal ultrasound. *Ultrasound in Obstetrics and Gynecology, 15*(1), 52-55.
Cramer, D. W., Bast, R. C., Berg, C. D., Diamandis, E. P., Godwin, A. K., Hartge, P., . . . Mor, G. (2011). Ovarian cancer biomarker performance in prostate, lung, colorectal, and ovarian cancer screening trial specimens. *Cancer prevention research, 4*(3), 365-374.
Croissant, Y. (2012). Estimation of multinomial logit models in R: The mlogit Packages. *R package version 0.2-2. URL: http://cran. r-project. org/web/packages/mlogit/vignettes/mlogit. pdf*.
Cronin, R. M., Mayo-Gamble, T. L., Stimpson, S.-J., Badawy, S. M., Crosby, L. E., Byrd, J., . . . Murry, V. M. (2018). Adapting medical guidelines to be patient-centered using a patient-driven process for individuals with sickle cell disease and their caregivers. *BMC hematology, 18*(1), 1-13.
Culyer, A. J. (1989). The normative economics of health care finance and provision. *Oxford review of economic policy, 5*(1), 34-58.
Davey, H. M., Lim, J., Butow, P. N., Barratt, A. L., & Redman, S. (2004). Women's preferences for and views on decision-making for diagnostic tests. *Social science & medicine, 58*(9), 1699-1707. doi:https://doi.org/10.1016/S0277-9536(03)00339-3
Day, B., Bateman, I. J., Carson, R. T., Dupont, D., Louviere, J. J., Morimoto, S., . . . Wang, P. (2012). Ordering effects and choice set awareness in repeat-response stated preference studies. *Journal of environmental economics and management, 63*(1), 73-91. doi:https://doi.org/10.1016/j.jeem.2011.09.001
de Bekker-Grob, E., Rose, J., Donkers, B., Essink-Bot, M.-L., Bangma, C., & Steyerberg, E. (2013a). Men's preferences for prostate cancer screening: a discrete choice experiment. *British Journal of Cancer, 108*(3), 533-541.
de Bekker-Grob, E., Rose, J. M., Donkers, B., Essink-Bot, M. L., Bangma, C. H., & Steyerberg, E. W. (2013b). Men's preferences for prostate cancer screening: A discrete choice experiment. *British Journal of Cancer, 108*(3), 533-541. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1038/bjc.2013.5
de Bekker-Grob, E. W., Donkers, B., Jonker, M. F., & Stolk, E. A. (2015). Sample size requirements for discrete-choice experiments in healthcare: a practical guide. *The Patient-Patient-Centered Outcomes Research, 8*(5), 373-384.
de Bekker-Grob, E. W., Donkers, B., Veldwijk, J., Jonker, M. F., Buis, S., Huisman, J., & Bindels, P. (2020). What Factors Influence Non-Participation Most in Colorectal Cancer Screening? A Discrete Choice Experiment. *The Patient-Patient-Centered Outcomes Research*, 1-13.
De Bekker-Grob, E. W., Rose, J. M., Donkers, B., Essink-Bot, M. L., Bangma, C. H., & Steyerberg, E. W. (2013c). Men's preferences for prostate cancer screening: A discrete choice experiment. *British Journal of Cancer, 108*(3), 533-541. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1038/bjc.2013.5
de Bekker-Grob, E. W., Ryan, M., & Gerard, K. (2012). Discrete choice experiments in health economics: a review of the literature. *Health Econ, 21*(2), 145-172. doi:10.1002/hec.1697
de Bekker-Grob, E. W., Ryan, M., & Gerard, K. (2012). Discrete choice experiments in health economics: a review of the literature. *Health Economics, 21*(2), 145-172.
De Brún, A., Flynn, D., Ternent, L., Price, C. I., Rodgers, H., Ford, G. A., . . . Teah, J. (2018). A novel design process for selection of attributes for inclusion in discrete choice experiments: case study exploring variation in clinical decision-making about thrombolysis in the treatment of acute ischaemic stroke. *BMC health services research, 18*(1), 1-14.
Debreu, G. (1959). *Theory of value: An axiomatic analysis of economic equilibrium*: Yale University Press.
Deed, K., Childs, J., & Thoirs, K. (2014). What are the perceptions of women towards transvaginal sonographic examinations? *Sonography, 1*(2), 33-38.
Dekker, T. (2014). Indifference based value of time measures for Random Regret Minimisation models. *Journal of choice modelling, 12*, 10-20.
Department of Health. (2010). Equity and excellence: Liberating the NHS. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213823/dh_117794.pdf

Department of Health. (2015). The NHS Constitution. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/fi le/480482/NHS_Constitution_WEB.pdf
Determann, D., Gyrd-Hansen, D., de Wit, G. A., de Bekker-Grob, E. W., Steyerberg, E. W., Lambooij, M. S., & Bjørnskov Pedersen, L. (2019). Designing unforced choice experiments to inform health care decision making: implications of using opt-out, neither, or status quo alternatives in discrete choice experiments. *Medical Decision Making, 39*(6), 681-692.
Devlin, N. J., Shah, K. K., Feng, Y., Mulhern, B., & van Hout, B. (2018). Valuing health-related quality of life: An EQ-5 D-5 L value set for E ngland. *Health Economics, 27*(1), 7-22.
Dhar, R. (1997). Consumer preference for a no-choice option. *Journal of consumer research, 24*(2), 215-231.
Dhillon, N., Oliffe, J. L., Kelly, M. T., & Krist, J. (2020). Bridging barriers to cervical cancer screening in transgender men: a scoping review. *American journal of men's health, 14*(3), 1557988320925691.
Dochez, V., Caillon, H., Vaucel, E., Dimet, J., Winer, N., & Ducarme, G. (2019). Biomarkers and algorithms for diagnosis of ovarian cancer: CA125, HE4, RMI and ROMA, a review. *Journal of ovarian research, 12*(1), 1-9.
Doherty, E., Hobbins, A., Whitehurst, D. G., & O'Neill, C. (2021). An exploration on attribute non-attendance using discrete choice experiment data from the Irish EQ-5D-5L National Valuation Study. *PharmacoEconomics-Open, 5*(2), 237-244.
Dolan, P. (1999). Whose preferences count? *Medical Decision Making, 19*(4), 482-486.
Dolan, P., Gudex, C., Kind, P., & Williams, A. (1996). The time trade-off method: results from a general population study. *Health Economics, 5*(2), 141-154.
Donaldson, C., & Shackley, P. (1997). Does "process utility" exist? A case study of willingness to pay for laparoscopic cholecystectomy. *Social Science & Medicine, 44*(5), 699-707.
Doubeni, C. A., Doubeni, A. R., & Myers, A. E. (2016). Diagnosis and management of ovarian cancer. *American family physician, 93*(11), 937-944.
Drescher, C., Holt, S. K., Andersen, M. R., Anderson, G., & Urban, N. (2000). Reported ovarian cancer screening among a population-based sample in Washington State. *Obstetrics & Gynecology, 96*(1), 70-74.
Drescher, C. W., Nelson, J., Peacock, S., Andersen, M. R., McIntosh, M. W., & Urban, N. (2004). Compliance of average-and intermediate-risk women to semiannual ovarian cancer screening. *Cancer Epidemiology and Prevention Biomarkers, 13*(4), 600-606.
Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes*: Oxford university press.
Duffy, S. W., Vulkan, D., Cuckle, H., Parmar, D., Sheikh, S., Smith, R. A., . . . Moss, S. M. (2020). Effect of mammographic screening from age 40 years on breast cancer mortality (UK Age trial): final results of a randomised, controlled trial. *The Lancet Oncology, 21*(9), 1165-1172. doi:10.1016/S1470-2045(20)30398-3
Ellimoottil, C., Marcotte, M., Grace, D., Krasnikov, A., Phillips, J. M., Quek, M. L., . . . Gupta, G. N. (2018). Examining Patient Willingness to Pay for Magnetic Resonance Imaging Guided Prostate Biopsy: Implications in the Era of Health Savings Accounts. *Urology Practice*. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1016/j.urpr.2017.03.003
Ells, C., & Thombs, B. D. (2014). The ethics of how to manage incidental findings. *CMAJ, 186*(9), 655-656.
Elwyn, G., Edwards, A., & Kinnersley, P. (1999). Shared decision-making in primary care: the neglected second half of the consultation. *British Journal of General Practice, 49*(443), 477-482.
Erdem, S., Campbell, D., & Hole, A. R. (2015). Accounting for attribute-level non-attendance in a health choice experiment: does it matter? *Health economics, 24*(7), 773-789.
Fagerlin, A., Zikmund-Fisher, B. J., & Ubel, P. A. (2011). Helping patients decide: ten steps to better risk communication. *Journal of the National Cancer Institute, 103*(19), 1436-1443.
Fallowfield, L., Fleissig, A., Barrett, J., Menon, U., Jacobs, I., Kilkerr, J., & Farewell, V. (2010). Awareness of ovarian cancer risk factors, beliefs and attitudes towards screening: baseline survey of 21 715 women participating in the UK Collaborative Trial of Ovarian Cancer Screening. *British Journal of Cancer, 103*(4), 454-461.
Farghaly, S. (2014). Current diagnosis and management of ovarian cysts. *Clinical and experimental obstetrics & gynecology, 41*(6), 609-612.
Fenichel, E. P., Lupi, F., Hoehn, J. P., & Kaplowitz, M. D. (2009). Split-sample tests of "no opinion" responses in an attribute-based choice model. *Land Economics, 85*(2), 348-362.

Fiebig, D. G., Haas, M., Hossain, I., Street, D. J., & Viney, R. (2009). Decisions about Pap tests: What influences women and providers? *Social Science and Medicine, 68*(10), 1766-1774. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1016/j.socscimed.2009.03.002
Finn, A., & Louviere, J. J. (1992). Determining the appropriate response to evidence of public concern: the case of food safety. *Journal of Public Policy & Marketing, 11*(2), 12-25.
Fitch, M. I., Deane, K., Howell, D., & Gray, R. E. (2002). Women's experiences with ovarian cancer: reflections on being diagnosed. *Canadian Oncology Nursing Journal/Revue canadienne de soins infirmiers en oncologie, 12*(3), 152-159.
Florin, D., & Dixon, J. (2004). Public involvement in health care. *Bmj, 328*(7432), 159-161. doi:10.1136/bmj.328.7432.159
Flynn, T. N., Louviere, J. J., Peters, T. J., & Coast, J. (2007). Best–worst scaling: what it can do for health care research and how to do it. *Journal of health economics, 26*(1), 171-189.
Forman, D. (2009). Cancer Incidence and Survival by Major Ethnic Group, England, 2002-2006. *National Cancer Intelligence Network*.
Fremont, A. M., Cleary, P. D., Hargraves, J. L., Rowe, R. M., Jacobson, N. B., & Ayanian, J. Z. (2001). Patient-centered processes of care and long-term outcomes of myocardial infarction. *J Gen Intern Med, 16*(12), 800-808. doi:10.1111/j.1525-1497.2001.10102.x
Fruchter, R. G., & Boyce, J. (1981). Delays in diagnosis and stage of disease in gynecologic cancer. *Cancer Detect Prev, 4*(1-4), 481-486.
Funston, G., Hamilton, W., Abel, G., Crosbie, E. J., Rous, B., & Walter, F. M. (2020a). The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: A population-based cohort study. *PLoS medicine, 17*(10), e1003295.
Funston, G., Hardy, V., Abel, G., Crosbie, E. J., Emery, J., Hamilton, W., & Walter, F. M. (2020b). Identifying ovarian cancer in symptomatic women: a systematic review of clinical tools. *Cancers, 12*(12), 3686.
Funston, G., Mounce, L. T., Price, S., Rous, B., Crosbie, E. J., Hamilton, W., & Walter, F. M. (2021). CA125 test result, test-to-diagnosis interval, and stage in ovarian cancer at diagnosis: a retrospective cohort study using electronic health records. *British Journal of General Practice, 71*(707), e465-e472.
Funston, G., Van Melle, M., Baun, M.-L. L., Jensen, H., Helsper, C., Emery, J., . . . Walter, F. M. (2019). Variation in the initial assessment and investigation for ovarian cancer in symptomatic women: a systematic review of international guidelines. *BMC cancer, 19*(1), 1028.
Garcia-Retamero, R., & Cokely, E. T. (2017). Designing visual aids that promote risk literacy: a systematic review of health research and evidence-based design heuristics. *Human factors, 59*(4), 582-627.
Garcia-Retamero, R., & Galesic, M. (2009). Communicating treatment risk reduction to people with low numeracy skills: a cross-cultural comparison. *American journal of public health, 99*(12), 2196-2202.
Garcia-Retamero, R., Galesic, M., & Gigerenzer, G. (2010). Do icon arrays help reduce denominator neglect? *Medical Decision Making, 30*(6), 672-684.
Gatos, K. C. (2018). A Literature Review of Cervical Cancer Screening in Transgender Men. *Nursing for Women's Health, 22*(1), 52-62. doi:https://doi.org/10.1016/j.nwh.2017.12.008
Gelhorn, H., Ross, M. M., Kansal, A. R., Fung, E. T., Seiden, M. V., Krucien, N., & Chung, K. C. (2022). Patient Preferences for Multi-Cancer Early Detection (MCED) Screening Tests. *The Patient-Patient-Centered Outcomes Research*, 1-14.
Gentry-Maharaj, A., Sharma, A., Burnell, M., Ryan, A., Amso, N. N., Seif, M., . . . Rangar, R. (2013). Acceptance of transvaginal sonography by postmenopausal women participating in the United Kingdom Collaborative Trial of Ovarian Cancer Screening. *Ultrasound in obstetrics & gynecology, 41*(1), 73-79.
Georgescu-Roegen, N. (1958). Threshold in Choice and the Theory of Demand. *Econometrica: Journal of the Econometric Society*, 157-168.
Gerard, K., Shanahan, M., & Louviere, J. (2003). Using stated preference discrete choice modelling to inform health care decisionmaking: A pilot study of breast screening participation. *Applied Economics, 35*(9), 1073-1085. doi:10.1080/0003684032000081348
Ghanouni, A., Halligan, S., Taylor, S. A., Boone, D., Plumb, A., Stoffel, S., . . . Von Wagner, C. (2014). Quantifying public preferences for different bowel preparation options prior to screening CT colonography: A discrete choice experiment. *Bmj Open, 4*(4), e004327. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1136/bmjopen-2013-004327

Ghanouni, A., Sanderson, S. C., Pashayan, N., Renzi, C., Von Wagner, C., & Waller, J. (2020a). Attitudes towards risk-stratified breast cancer screening among women in England: A cross-sectional survey. *Journal of Medical Screening, 27*(3), 138-145.
Ghanouni, A., Smith, S. G., Halligan, S., Plumb, A., Boone, D., Yao, G. L., . . . Von Wagner, C. (2013). Public preferences for colorectal cancer screening tests: a review of conjoint analysis studies. *Expert review of medical devices, 10*(4), 489-499.
Ghanouni, A., Waller, J., Stoffel, S. T., Vlaev, I., & von Wagner, C. (2020b). Acceptability of risk-stratified breast screening: Effect of the order of presenting risk and benefit information. *Journal of Medical Screening, 27*(1), 52-56. doi:10.1177/0969141319877669
Gigerenzer, G., Mata, J., & Frank, R. (2009). Public knowledge of benefits of breast and prostate cancer screening in Europe. *Journal of the National Cancer Institute, 101*(17), 1216-1220.
Gilbert, L., Basso, O., Sampalis, J., Karp, I., Martins, C., Feng, J., . . . Takefman, J. (2012). Assessment of symptomatic women for early diagnosis of ovarian cancer: results from the prospective DOvE pilot project. *The Lancet Oncology, 13*(3), 285-291.
Gowdy, J. M., & Mayumi, K. (2001). Reformulating the foundations of consumer choice theory and environmental valuation. *Ecological Economics, 39*(2), 223-237.
Green, P. E., Krieger, A. M., & Agarwal, M. K. (1991). Adaptive conjoint analysis: Some caveats and suggestions. *Journal of Marketing Research, 28*(2), 215-222.
Greene, W. H., & Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B: Methodological, 37*(8), 681-698.
Groothuis-Oudshoorn, C. G., Fermont, J. M., van Til, J. A., & Ijzerman, M. J. (2014). Public stated preferences and predicted uptake for genome-based colorectal cancer screening. *BMC medical informatics and decision making, 14*, 18. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1186/1472-6947-14-18
Guerriero, S., Mallarini, G., Ajossa, S., Risalvato, A., Satta, R., Mais, V., . . . Melis, G. B. (1997). Transvaginal ultrasound and computed tomography combined with clinical parameters and CA-125 determinations in the differential diagnosis of persistent ovarian cysts in premenopausal women. *Ultrasound in Obstetrics and Gynecology: The Official Journal of the International Society of Ultrasound in Obstetrics and Gynecology, 9*(5), 339-343.
Gupta, K. K., Gupta, V. K., & Naumann, R. W. (2019). Ovarian cancer: screening and future directions. *International Journal of Gynecologic Cancer, 29*(1).
Hall, R., Medina-Lara, A., Hamilton, W., & Spencer, A. E. (2021). Attributes Used for Cancer Screening Discrete Choice Experiments: A Systematic Review. *The Patient-Patient-Centered Outcomes Research*, 1-17.
Hamilton, W. (2010). Cancer diagnosis in primary care. *British Journal of General Practice, 60*(571), 121. doi:10.3399/bjgp10X483175
Hamilton, W., Peters, T. J., Bankhead, C., & Sharp, D. (2009). Risk of ovarian cancer in women with symptoms in primary care: population based case-control study. *Bmj, 339*, b2998. doi:10.1136/bmj.b2998
Hanley, N., Mourato, S., & Wright, R. E. (2001). Choice modelling approaches: a superior alternative for environmental valuatioin? *Journal of economic surveys, 15*(3), 435-462.
Harker, N. (2017). How GPs can offer cancer patients the support they need. Retrieved from https://www.gponline.com/gps-offer-cancer-patients-support-need/cancer/article/1441477
Harrison, M., Rigby, D., Vass, C., Flynn, T., Louviere, J., & Payne, K. (2014). Risk as an attribute in discrete choice experiments: a systematic review of the literature. *Patient, 7*(2), 151-170. doi:10.1007/s40271-014-0048-1
Hauber, A. B., González, J. M., Groothuis-Oudshoorn, C. G., Prior, T., Marshall, D. A., Cunningham, C., . . . Bridges, J. F. (2016). Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value in health, 19*(4), 300-315.
Hauser, D. J., & Schwarz, N. (2015). It'sa trap! Instructional manipulation checks prompt systematic thinking on "tricky" tasks. *SAGE Open, 5*(2), 2158244015584617.
Hausman, D. M. (2011). *Preference, value, choice, and welfare*: Cambridge University Press.
He, W., Li, M., Zuo, L., Wang, M., Jiang, L., Shan, H., . . . Han, X. (2019). Acupuncture for treatment of insomnia: An overview of systematic reviews. *Complementary therapies in medicine, 42*, 407-416.
Heidenreich, S., Watson, V., Ryan, M., & Phimister, E. (2018). Decision heuristic or preference? Attribute non-attendance in discrete choice problems. *Health Economics, 27*(1), 157-171.

Helter, T. M., & Boehler, C. E. H. (2016). Developing attributes for discrete choice experiments in health: a systematic literature review and case study of alcohol misuse interventions. *Journal of substance use, 21*(6), 662-668.
Henderson, J. T., Webber, E. M., & Sawaya, G. F. (2018). Screening for ovarian cancer: updated evidence report and systematic review for the US preventive services task force. *Jama, 319*(6), 595-606.
Hendrix, N., Hauber, B., Lee, C. I., Bansal, A., & Veenstra, D. L. (2020). Artificial intelligence in breast cancer screening: primary care provider preferences. *Journal of the American Medical Informatics Association*.
Hensher, D., Louviere, J., & Swait, J. (1998). Combining sources of preference data. *Journal of Econometrics, 89*(1-2), 197-221.
Hensher, D. A. (2006). Revealing differences in willingness to pay due to the dimensionality of stated choice designs: an initial assessment. *Environmental and resource economics, 34*(1), 7-44.
Hensher, D. A. (2010). Hypothetical bias, choice experiments and willingness to pay. *transportation research part B: methodological, 44*(6), 735-752.
Hensher, D. A., & Greene, W. H. (2010). Non-attendance and dual processing of common-metric attributes in choice analysis: a latent class specification. *Empirical economics, 39*(2), 413-426.
Hensher, D. A., Rose, J., & Greene, W. H. (2005a). The implications on willingness to pay of respondents ignoring specific attributes. *Transportation, 32*(3), 203-222.
Hensher, D. A., Rose, J. M., & Black, I. (2008). Interactive agency choice in automobile purchase decisions: the role of negotiation in determining equilibrium choice outcomes. *Journal of Transport Economics and Policy (JTEP), 42*(2), 269-296.
Hensher, D. A., Rose, J. M., & Greene, W. H. (2005b). *Applied choice analysis: a primer*: Cambridge University Press.
Hensher, D. A., Rose, J. M., Rose, J. M., & Greene, W. H. (2005c). *Applied choice analysis: a primer*: Cambridge university press.
Hersch, J., Jansen, J., Barratt, A., Irwig, L., Houssami, N., Howard, K., . . . McCaffery, K. (2013). Women's views on overdiagnosis in breast cancer screening: a qualitative study. *Bmj, 346*.
Hess, S., & Hensher, D. A. (2010). Using conditioning on observed choices to retrieve individual-specific attribute processing strategies. *Transportation Research Part B: Methodological, 44*(6), 781-790.
Hess, S., Hensher, D. A., & Daly, A. (2012). Not bored yet–revisiting respondent fatigue in stated choice experiments. *Transportation Research Part A: Policy and Practice, 46*(3), 626-644.
Hess, S., Stathopoulos, A., Campbell, D., O'Neill, V., & Caussade, S. (2013). It's not that I don't care, I just don't care very much: confounding between attribute non-attendance and taste heterogeneity. *Transportation, 40*(3), 583-607.
Hippisley-Cox, J., & Coupland, C. (2012). Identifying women with suspected ovarian cancer in primary care: derivation and validation of algorithm. *Bmj, 344*.
Hirst, Y., Stoffel, S., Baio, G., McGregor, L., & von Wagner, C. (2018). Uptake of the English Bowel (Colorectal) Cancer Screening Programme: an update 5 years after the full roll-out. *European Journal of Cancer, 103*, 267-273.
Hoehn, R. S., & Zureikat, A. H. (2020). Cancer disparities in the COVID-19 era. *Journal of Surgical Oncology*.
Hoffmann, T. C., & Del Mar, C. (2015). Patients' expectations of the benefits and harms of treatments, screening, and tests: a systematic review. *JAMA internal medicine, 175*(2), 274-286.
Hoffrage, U., Gigerenzer, G., Krauss, S., & Martignon, L. (2002). Representation facilitates reasoning: what natural frequencies are and what they are not. *Cognition, 84*(3), 343-352.
Hol, L., De Bekker-Grob, E. W., Van Dam, L., Donkers, B., Kuipers, E. J., Habbema, J. D. F., . . . Essink-Bot, M. L. (2010). Preferences for colorectal cancer screening strategies: A discrete choice experiment. *British Journal of Cancer, 102*(6), 972-980. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1038/sj.bjc.6605566
Hole, A. R. (2006). Small-sample properties of tests for heteroscedasticity in the conditional logit model. *Economics Bulletin, 3*(18), 1-14.
Hole, A. R., Kolstad, J. R., & Gyrd-Hansen, D. (2013). Inferred vs. stated attribute non-attendance in choice experiments: A study of doctors' prescription behaviour. *Journal of Economic Behavior & Organization, 96*, 21-31.
Hole, A. R., Norman, R., & Viney, R. (2016). Response patterns in health state valuation using endogenous attribute attendance and latent class analysis. *Health economics, 25*(2), 212-224.

Hollinghurst, S., Banks, J., Bigwood, L., Walter, F. M., Hamilton, W., & Peters, T. J. (2016). Using willingness-to-pay to establish patient preferences for cancer testing in primary care. *BMC Medical Informatics and Decision Making, 16*(1), 105. doi:10.1186/s12911-016-0345-9
Holman, L. L., Lu, K. H., Bast Jr, R. C., Hernandez, M. A., Bodurka, D. C., Skates, S., & Sun, C. C. (2014). Risk perception, worry, and test acceptance in average-risk women who undergo ovarian cancer screening. *American journal of obstetrics and gynecology, 210*(3), 257. e251-257. e256.
Holmes, T. P., & Boyle, K. J. (2005). Dynamic learning and context-dependence in sequential, attribute-based, stated-preference valuation questions. *Land Economics, 81*(1), 114-126.
Horsky, D., Nelson, P., & Posavac, S. S. (2004). Stating preference for the ethereal but choosing the concrete: how the tangibility of attributes affects attribute weighting in value elicitation and choice. *Journal of Consumer Psychology, 14*(1-2), 132-140.
Howard, K., & Salkeld, G. (2009). Does attribute framing in discrete choice experiments influence willingness to pay? Results from a discrete choice experiment in screening for colorectal cancer. *Value in Health, 12*(2), 354-363. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1111/j.1524-4733.2008.00417.x
Howard, K., Salkeld, G., McCaffery, K., & Irwig, L. (2008). HPV triage testing or repeat Pap smear for the management of atypical squamous cells (ASCUS) on Pap smear: is there evidence of process utility? *Health Economics, 17*(5), 593-605.
Howard, K., Salkeld, G., Pignone, M., Hewett, P., Cheung, P., Olsen, J., . . . Roberts-Thomson, I. C. (2011). Preferences for CT colonography and colonoscopy as diagnostic tests for colorectal cancer: A discrete choice experiment. *Value in health, 14*(8), 1146-1152. doi:http://dx.doi.org/10.1016/j.jval.2011.07.012
Howard, K., Salkeld, G. P., Patel, M. I., Mann, G. J., & Pignone, M. P. (2015). Men's preferences and trade-offs for prostate cancer screening: a discrete choice experiment. *Health expectations : an international journal of public participation in health care and health policy, 18*(6), 3123-3135. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1111/hex.12301
Hu, W., Adamowicz, W. L., & Veeman, M. M. (2006). Labeling Context and Reference Point Effects in Models of Food Attribute Demand. *American journal of agricultural economics, 88*(4), 1034-1049. doi:https://doi.org/10.1111/j.1467-8276.2006.00914.x
Huber, J., & Zwerina, K. (1996). The importance of utility balance in efficient choice designs. *Journal of Marketing Research, 33*(3), 307-317.
Huber, O. (1980). The influence of some task variables on cognitive operations in an information-processing decision model. *Acta Psychologica, 45*(1-3), 187-196.
Hurley, J. (1998). Welfarism, extra-welfarism and evaluative economic analysis in the health sector. *Health, health care and health economics: Perspectives on distribution*, 373-395.
Hurley, J. (2000). An overview of the normative economics of the health sector. In *Handbook of health economics* (Vol. 1, pp. 55-118): Elsevier.
Hurley, R. (2018). Overdiagnosis and the cancer label. In: British Medical Journal Publishing Group.
Incisive Health. (2014). *Saving lives averting costs: an analysis of the financial implications of achieving earlier diagnosis of colorectal, lung and ovarian cancer. A report prepared for Cancer Research UK.* Retrieved from http://www.incisivehealth.com/uploads/Saving%20lives%20averting%20costs.pdf
Jack, R. H., Møller, H., Robson, T., & Davies, E. A. (2014). Breast cancer screening uptake among women from different ethnic groups in London: a population-based cohort study. *Bmj Open, 4*(10), e005586.
Jacobs, I. J., Menon, U., Ryan, A., Gentry-Maharaj, A., Burnell, M., Kalsi, J. K., . . . Cruickshank, D. (2016). Ovarian cancer screening and mortality in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *The Lancet, 387*(10022), 945-956.
Jelicic, L., Brooker, J., Shand, L., Knight, T., Ricciardelli, L., Denham, G., & Burney, S. (2019). Experiences and health care preferences of women with ovarian cancer during the diagnosis phase. *Psycho-oncology, 28*(2), 379-385.
Jenkins, V., Fallowfield, L., Langridge, C., Barrett, J., Ryan, A., Jacobs, I., . . . Farewell, V. (2015). Psychosocial factors associated with withdrawal from the United Kingdom collaborative trial of ovarian Cancer screening after 1 episode of repeat screening. *International Journal of Gynecologic Cancer, 25*(8).
Jiang, Y., & Punj, G. N. (2010). The effects of attribute concreteness and prominence on selective processing, choice, and search experience. *Journal of the Academy of Marketing Science, 38*(4), 471-489.

Johar, M., Fiebig, D. G., Haas, M., & Viney, R. (2013). Using repeated choice experiments to evaluate the impact of policy changes on cervical screening. *Applied Economics, 45*(14), 1845-1855.
Johnson, F. R., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D. A., . . . Bridges, J. F. (2013). Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value in health, 16*(1), 3-13.
Johnson, F. R., Yang, J.-C., & Reed, S. D. (2019). The internal validity of discrete choice experiment data: a testing tool for quantitative assessments. *Value in health, 22*(2), 157-160.
Jonker, M. F., Donkers, B., de Bekker-Grob, E. W., & Stolk, E. A. (2018). Effect of level overlap and color coding on attribute non-attendance in discrete choice experiments. *Value in health, 21*(7), 767-771.
Jonker, M. F., Donkers, B., de Bekker-Grob, E., & Stolk, E. A. (2019). Attribute level overlap (and color coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. *Health Economics, 28*(3), 350-363.
Katz, S. J., Belkora, J., & Elwyn, G. (2014). Shared Decision Making for Treatment of Cancer: Challenges and Opportunities. *Journal of Oncology Practice, 10*(3), 206-208. doi:10.1200/jop.2014.001434
Kim, C., Armstrong, M. J., Berta, W. B., & Gagliardi, A. R. (2020). How to identify, incorporate and report patient preferences in clinical guidelines: A scoping review. *Health Expectations, 23*(5), 1028-1036.
Kingsley, D. C., & Brown, T. C. (2010). Preference uncertainty, preference learning, and paired comparison experiments. *Land Economics, 86*(3), 530-544.
Kistler, C. E., Hess, T. M., Howard, K., Pignone, M. P., Crutchfield, T. M., Hawley, S. T., . . . Lewis, C. L. (2015). Older adults' preferences for colorectal cancer-screening test attributes and test choice. *Patient Preference and Adherence, 9*, 1005-1016. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.2147/PPA.S82203
Kitchener, H. C., Gittins, M., Rivero-Arias, O., Tsiachristas, A., Cruickshank, M., Gray, A., . . . Roberts, C. (2016). A cluster randomised trial of strategies to increase cervical screening uptake at first invitation (STRATEGIC). *Health Technology Assessment, 20*(68). doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.3310/hta20680
Kjær, T., Bech, M., Gyrd-Hansen, D., & Hart-Hansen, K. (2006). Ordering effect and price sensitivity in discrete choice experiments: need we worry? *Health Economics, 15*(11), 1217-1228.
Kleij, K.-S., Tangermann, U., Amelung, V. E., & Krauth, C. (2017). Patients' preferences for primary health care–a systematic literature review of discrete choice experiments. *BMC health services research, 17*(1), 1-12.
Kløjgaard, M. E., Bech, M., & Søgaard, R. (2012). Designing a stated choice experiment: the value of a qualitative process. *Journal of choice modelling, 5*(2), 1-18.
Kobayashi, H., Yamada, Y., Sado, T., Sakata, M., Yoshida, S., Kawaguchi, R., . . . Tsuji, Y. (2008). A randomized study of screening for ovarian cancer: a multicenter study in Japan. *International Journal of Gynecologic Cancer, 18*(3).
Kohler, R. E., Gopal, S., Lee, C. N., Weiner, B. J., Reeve, B. B., & Wheeler, S. B. (2017). Breast Cancer Knowledge, Behaviors, and Preferences in Malawi: Implications for Early Detection Interventions From a Discrete Choice Experiment. *Journal of global oncology, 3*(5), 480-489. doi:https://0-dx-doi-org.lib.exeter.ac.uk/10.1200/JGO.2016.005371
Krishnan, K. (1977). Incorporating thresholds of indifference in probabilistic choice models. *Management science, 23*(11), 1224-1233.
Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of political economy, 74*(2), 132-157.
Lancsar, E., & Louviere, J. (2006). Deleting 'irrational'responses from discrete choice experiments: a case of investigating or imposing preferences? *Health economics, 15*(8), 797-811.
Lancsar, E., & Louviere, J. (2008). Conducting discrete choice experiments to inform healthcare decision making. *Pharmacoeconomics, 26*(8), 661-677.
Lancsar, E., & Swait, J. (2014). Reconceptualising the external validity of discrete choice experiments. *Pharmacoeconomics, 32*(10), 951-965.
Lee, M. J., & Kang, H. (2018). Designing skin cancer prevention messages: Should we emphasize gains or losses? Message framing, risk type, and prior experience. *American Journal of Health Promotion, 32*(4), 939-948.
Leuraud, K., Jezewski-Serra, D., Viguier, J., & Salines, E. (2013). Colorectal cancer screening by guaiac faecal occult blood test in France: Evaluation of the programme two years after launching. *Cancer epidemiology, 37*(6), 959-967.

Levens, E. D., Wesley, R., Premkumar, A., Blocker, W., & Nieman, L. K. (2009). Magnetic resonance imaging and transvaginal ultrasound for determining fibroid burden: implications for research and clinical care. *American journal of obstetrics and gynecology, 200*(5), 537. e531-537. e537.
Lew, D. K., & Whitehead, J. C. (2020). Attribute non-attendance as an information processing strategy in stated preference choice experiments: Origins, current practices, and future directions. *Marine Resource Economics, 35*(3), 285-317.
Li, S., Liu, S., Ratcliffe, J., Gray, A., & Chen, G. (2019). Preferences for cervical cancer screening service attributes in rural China: a discrete choice experiment. *Patient Preference and Adherence, 13*, 881.
Light, A., Elhage, O., Marconi, L., & Dasgupta, P. (2019). Prostate cancer screening: where are we now? *BJU international, 123*(6), 916-917.
Lim, A., Mesher, D., Gentry-Maharaj, A., Balogun, N., Widschwendter, M., Jacobs, I., . . . Menon, U. (2016). Time to diagnosis of Type I or II invasive epithelial ovarian cancers: a multicentre observational study using patient questionnaire and primary care records. *Bjog, 123*(6), 1012-1020. doi:10.1111/1471-0528.13447
Lim, S. L., Yang, J.-C., Ehrisman, J., Havrilesky, L. J., & Reed, S. D. (2020). Are Videos or Text Better for Describing Attributes in Stated-Preference Surveys? *The Patient - Patient-Centered Outcomes Research, 13*(4), 401-408. doi:10.1007/s40271-020-00416-9
Little, P., Dorward, M., Warner, G., Stephens, K., Senior, J., & Moore, M. (2004). Importance of patient pressure and perceived pressure and perceived medical need for investigations, referral, and prescribing in primary care: nested observational study. *Bmj, 328*(7437), 444.
Liu, R., Lu, Y., Li, Y., Wei, W., Sun, C., Zhang, Q., . . . Zhang, N. (2022). Preference for endoscopic screening of upper gastrointestinal cancer among Chinese rural residents: a discrete choice experiment. *Frontiers in oncology, 12*.
Lockwood, M. (1996). Non-compensatory preference structures in non-market valuation of natural area policy. *Australian journal of Agricultural economics, 40*(2), 85-101.
Louviere, J., Lings, I., Islam, T., Gudergan, S., & Flynn, T. (2013). An introduction to the application of (case 1) best–worst scaling in marketing research. *International Journal of Research in Marketing, 30*(3), 292-303.
Louviere, J. J. (1988). Conjoint analysis modelling of stated preferences: a review of theory, methods, recent developments and external validity. *Journal of transport economics and policy*, 93-119.
Louviere, J. J., Flynn, T. N., & Carson, R. T. (2010). Discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling, 3*(3), 57-72.
Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). *Best-worst scaling: Theory, methods and applications*: Cambridge University Press.
Louviere, J. J., & Lancsar, E. (2009). Choice experiments in health: the good, the bad, the ugly and toward a brighter future. *Health Economics, Policy and Law, 4*(4), 527-546.
Lurie, G., Wilkens, L. R., Thompson, P. J., Matsuno, R. K., Carney, M. E., & Goodman, M. T. (2010). Symptom presentation in invasive ovarian carcinoma by tumor histological type and grade in a multiethnic population: a case analysis. *Gynecol Oncol, 119*(2), 278-284. doi:10.1016/j.ygyno.2010.05.028
Luu, J. M., Sergeant, A. K., Anand, S. S., Desai, D., Schulze, K., Knoppers, B. M., . . . Black, S. E. (2021). The impact of reporting magnetic resonance imaging incidental findings in the Canadian alliance for healthy hearts and minds cohort. *BMC medical ethics, 22*(1), 1-15.
Lyratzopoulos, G., Saunders, C. L., Abel, G. A., McPhail, S., Neal, R. D., Wardle, J., & Rubin, G. P. (2015). The relative length of the patient and the primary care interval in patients with 28 common and rarer cancers. *British Journal of Cancer, 112*(1), S35-S40. doi:10.1038/bjc.2015.40
Macdonald, S., Macleod, U., Campbell, N. C., Weller, D., & Mitchell, E. (2006). Systematic review of factors influencing patient and practitioner delay in diagnosis of upper gastrointestinal cancer. *British Journal of Cancer, 94*(9), 1272-1280.
Maclennan, D., & Williams, N. (1980). Revealed-preference theory and spatial choices: some limitations. *Environment and Planning A, 12*(8), 909-919.
Malagón, T., Yong, J. H., Tope, P., Miller Jr, W. H., Franco, E. L., Control, M. T. F. o. t. I. o. C. o. C., & Care. (2022). Predicted long-term impact of COVID-19 pandemic-related care delays on cancer mortality in Canada. *International Journal of Cancer, 150*(8), 1244-1254.
Mandrik, O., Yaumenenka, A., Herrero, R., & Jonker, M. F. (2019). Population preferences for breast cancer screening policies: Discrete choice experiment in Belarus. *PLoS ONE, 14*(11).

Mansfield, C., Ekwueme, D. U., Tangka, F. K. L., Brown, D. S., Smith, J. L., Guy, G. P., . . . Hauber, B. (2018). Colorectal Cancer Screening: Preferences, Past Behavior, and Future Intentions. *Patient*, 1-13. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1007/s40271-018-0308-6

Mansfield, C., Tangka, F. K., Ekwueme, D. U., Smith, J. L., Guy, G. P., Li, C., & Hauber, A. B. (2016). Stated Preference for Cancer Screening: A Systematic Review of the Literature, 1990-2013. *Preventing chronic disease, 13*, E27. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.5888/pcd13.150433

Marley, A., Flynn, T. N., & Australia, V. (2012). Best worst scaling: theory and practice. *Handbook of Choice Modelling. Edward Elgar Publishing, Leeds (UK)*.

Marshall, D., McGregor, E., & Currie, G. (2010). Measuring Preferences for Colorectal Cancer Screening. *The Patient: Patient-Centered Outcomes Research, 3*(2), 79-89. doi:10.2165/11532250-000000000-00000

Marshall, D. A., Johnson, F. R., Phillips, K. A., Marshall, J. K., Thabane, L., & Kulin, N. A. (2007). Measuring patient preferences for colorectal cancer screening using a choice-format survey. *Value in health, 10*(5), 415-430. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1111/j.1524-4733.2007.00196.x

Marshall, D. A., Johnson, R., Kulin, N. A., Ozdemir, A., Walsh, A., Marshall, J., . . . Phillips, K. (2009). How Do Physician Assessments of Patient Preferences for Colorectal Cancer Screening Tests Differ from Actual Preferences? A Comparison in Canada and the United States Using a Stated-Choice Survey. *Health Economics, 18*(12), 1420-1439. doi:http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291099-1050/issues

Martens, C. E., Crutchfield, T. M., Laping, J. L., Perreras, L., Reuland, D. S., Cubillos, L., . . . Wheeler, S. B. (2016). Why Wait Until Our Community Gets Cancer?: Exploring CRC Screening Barriers and Facilitators in the Spanish-Speaking Community in North Carolina. *Journal of cancer education : the official journal of the American Association for Cancer Education, 31*(4), 652-659. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1007/s13187-015-0890-4

Mbuya Bienge, C., Pashayan, N., Brooks, J. D., Dorval, M., Chiquette, J., Eloy, L., . . . Lévesque, E. (2021). Women's Views on Multifactorial Breast Cancer Risk Assessment and Risk-Stratified Screening: A Population-Based Survey from Four Provinces in Canada. *Journal of personalized medicine, 11*(2), 95.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of public economics, 3*(4), 303-328.

McFadden, D. (1986). The choice theory approach to market research. *Marketing science, 5*(4), 275-297.

Medina-Lara, A., Mujica-Mota, R. E., Kunkwenzu, E. D., & Lalloo, D. G. (2014). Stated preferences for anti-malarial drug characteristics in Zomba, a malaria endemic area of Malawi. *Malaria journal, 13*(1), 1-9.

Menczer, J. (2000). Diagnosis and treatment delay in gynecological malignancies. Does it affect outcome? *International Journal of Gynecological Cancer, 10*(2), 89-94.

Menon, U., Gentry-Maharaj, A., Burnell, M., Singh, N., Ryan, A., Karpinskyj, C., . . . Raikou, M. (2021). Ovarian cancer population screening and mortality after long-term follow-up in the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS): a randomised controlled trial. *The Lancet, 397*(10290), 2182-2193.

Menon, U., Gentry-Maharaj, A., Hallett, R., Ryan, A., Burnell, M., Sharma, A., . . . Lopes, A. (2009). Sensitivity and specificity of multimodal and ultrasound screening for ovarian cancer, and stage distribution of detected cancers: results of the prevalence screen of the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *The lancet oncology, 10*(4), 327-340.

Menon, U., McGuire, A. J., Raikou, M., Ryan, A., Davies, S. K., Burnell, M., . . . Amso, N. N. (2017). The cost-effectiveness of screening for ovarian cancer: results from the UK Collaborative Trial of Ovarian Cancer Screening (UKCTOCS). *British Journal of Cancer, 117*(5), 619-627.

Menzel, P., Dolan, P., Richardson, J., & Olsen, J. A. (2002). The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. *Social Science & Medicine, 55*(12), 2149-2158.

Michalovic, E., Hall, S., Duncan, L. R., Bassett-Gunter, R., & Sweet, S. N. (2018). Understanding the effects of message framing on physical activity action planning: the role of risk perception and elaboration. *International journal of behavioral medicine, 25*(6), 626-636.

Miles, A., Taylor, S. A., Evans, R. E., Halligan, S., Beare, S., Bridgewater, J., . . . Oliver, A. (2019). Patient preferences for whole-body MRI or conventional staging pathways in lung and colorectal cancer: a discrete choice experiment. *European radiology, 29*(7), 3889-3900.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine, 151*(4), 264-269.
Mooney, G. (1994). *Key issues in health economics* London: Harvester Wheatsheaf.
Mooney, G., & Lange, M. (1993). Ante-natal screening: What constitutes 'benefit'? *Social Science & Medicine, 37*(7), 873-878.
Moore, P. A., Ziegler, K. M., Lipman, R. D., Aminoshariae, A., Carrasco-Labra, A., & Mariotti, A. (2018). Benefits and harms associated with analgesic medications used in the management of acute dental pain: an overview of systematic reviews. *The Journal of the American Dental Association, 149*(4), 256-265. e253.
Moore, S. F., Price, S. J., Chowienczyk, S., Bostock, J., & Hamilton, W. (2021). The impact of changing risk thresholds on the number of people in England eligible for urgent investigation for possible cancer: an observational cross-sectional study. *Br J Cancer, 125*(11), 1593-1597. doi:10.1038/s41416-021-01541-4
Morris, S., Devlin, N., Parkin, D., & Spencer, A. (2012). *Economic Analysis in Healthcare*: Wiley.
Moser, K., Patnick, J., & Beral, V. (2009). Inequalities in reported use of breast and cervical screening in Great Britain: analysis of cross sectional survey data. *Bmj, 338*, b2025. doi:10.1136/bmj.b2025
Moss, E., Hollingworth, J., & Reynolds, T. (2005). The role of CA125 in clinical practice. *Journal of clinical pathology, 58*(3), 308-312.
Moss, E. L., Moran, A., Reynolds, T. M., & Stokes-Lampard, H. (2013). Views of general practitioners on the role of CA125 in primary care to diagnose ovarian cancer. *BMC women's health, 13*(1), 1-6.
Mühlbacher, A., & Johnson, F. R. (2016). Choice experiments to quantify preferences for health and healthcare: state of the practice. *Applied health economics and health policy, 14*(3), 253-266.
Mühlbacher, A. C., Bridges, J. F., Bethge, S., Dintsios, C., Schwalm, A., Gerber-Grote, A., & Nübling, M. (2017). Preferences for antiviral therapy of chronic hepatitis C: a discrete choice experiment. *The European Journal of Health Economics, 18*(2), 155-165.
Mühlbacher, A. C., Kaczynski, A., Zweifel, P., & Johnson, F. R. (2016). Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. *Health Economics Review, 6*(1), 2. doi:10.1186/s13561-015-0079-x
Mujica-Mota, R. E., Roberts, M., Abel, G., Elliott, M., Lyratzopoulos, G., Roland, M., & Campbell, J. (2015). Common patterns of morbidity and multi-morbidity and their impact on health-related quality of life: evidence from a national survey. *Qual Life Res, 24*(4), 909-918. doi:10.1007/s11136-014-0820-7
Mulley, A. G., Trimble, C., & Elwyn, G. (2012). Stop the silent misdiagnosis: patients' preferences matter. *Bmj, 345*.
Nagle, C. M., Francis, J. E., Nelson, A. E., Zorbas, H., Luxford, K., De Fazio, A., . . . Webb, P. M. (2011). Reducing time to diagnosis does not improve outcomes for women with symptomatic ovarian cancer: a report from the Australian Ovarian Cancer Study Group. *J Clin Oncol, 29*(16), 2253-2258.
Naik, G., Ahmed, H., & Edwards, A. G. (2012). Communicating risk to patients and the public. *British Journal of General Practice, 62*(597), 213-216.
Nash, Z., & Menon, U. (2020). Ovarian cancer screening: Current status and future directions. *Best practice & research Clinical obstetrics & gynaecology, 65*, 32-45.
National Academies of Sciences, E., & Medicine. (2016). *Ovarian cancers: evolving paradigms in research and care*: National Academies Press.
National Screening Committee. (2015). Criteria for appraising the viability, effectiveness and appropriateness of a screening programme. Retrieved from https://www.gov.uk/government/publications/evidence-review-criteria-national-screening-programmes/criteria-for-appraising-the-viability-effectiveness-and-appropriateness-of-a-screening-programme
Nayaradou, M., Berchi, C., Dejardin, O., & Launoy, G. (2010). Eliciting population preferences for mass colorectal cancer screening organization. *Medical decision making : an international journal of the Society for Medical Decision Making, 30*(2), 224-233.
NCRAS. (2019a). Routes to Diagnosis 2006-2016 workbook (a). Retrieved from http://www.ncin.org.uk/publications/routes_to_diagnosis
NCRAS. (2019b). Stage by Routes to Diagnosis, 2015-2016. Retrieved from http://www.ncin.org.uk/publications/routes_to_diagnosis

Neal, D. E., Feit, E. M., & Etzkorn, J. R. (2018). Patient Preferences for the Treatment of Basal Cell Carcinoma: A Mapping Review of Discrete Choice Experiments. *Dermatol Surg, 44*(8), 1041-1049. doi:10.1097/dss.0000000000001564
Neal, R. D., Allgar, V. L., Ali, N., Leese, B., Heywood, P., Proctor, G., & Evans, J. (2007). Stage, survival and delays in lung, colorectal, prostate and ovarian cancer: comparison between diagnostic routes. *British Journal of General Practice, 57*(536), 212-219.
Neal, R. D., Tharmanathan, P., France, B., Din, N. U., Cotton, S., Fallon-Ferguson, J., . . . Emery, J. (2015). Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *Br J Cancer, 112 Suppl 1*(Suppl 1), S92-107. doi:10.1038/bjc.2015.48
Neuman, T., Neuman, E., & Neuman, S. (2010). Explorations of the effect of experience on preferences for a health-care service. *The Journal of Socio-Economics, 39*(3), 407-419.
Nguyen, T. C., Robinson, J., Whitty, J. A., Kaneko, S., & Nguyen, T. C. (2015). Attribute non-attendance in discrete choice experiments: A case study in a developing country. *Economic Analysis and Policy, 47*, 22-33.
NHS (2022). Blood Tests. Retrieved from https://www.nhs.uk/conditions/blood-tests/#:~:text=Some%20test%20results%20will%20be,can%20be%20stressful%20and%20upsetting.
NHS Digital. (2018). *Cervical Screening Programme England 2017-18*. Retrieved from
NHS Digital. (2019). Breast Screening Programme, England 2017-18. Retrieved from https://digital.nhs.uk/data-and-information/publications/statistical/breast-screening-programme/england-2017-18
NHS England. (2016). Achieving World Class Cancer Outcomes: Taking the strategy forward: five year forward. Retrieved from https://www.england.nhs.uk/wp-content/uploads/2016/05/cancer-strategy.pdf
NHS England. (2018). NHS Diagnostic Waiting Times and Activity Data. Retrieved from https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2018/06/DWTA-Report-April-2018.pdf
NHS England. (2022). Cervical screening: quarterly coverage data reports 2021. Retrieved from https://www.gov.uk/government/publications/cervical-screening-quarterly-coverage-data-reports-2021
NHS UK. (2020). Diagnosis: Ovarian Cancer. Retrieved from https://www.nhs.uk/conditions/ovarian-cancer/diagnosis/
NHS UK. (2021a). How to decide if you want breast screening. Retrieved from https://www.nhs.uk/conditions/breast-screening-mammogram/how-to-decide-if-you-want-breast-screening/
NHS UK. (2021b). NHS screening. Retrieved from https://www.nhs.uk/conditions/nhs-screening/
NICE. (2011). Ovarian cancer: recognition and initial management. Retrieved from https://www.nice.org.uk/guidance/CG122
NICE. (2015). Suspected cancer: recognition and referral (NG12). Retrieved from www.nice.org.uk/guidance/ng12
NICE. (2017). Quantitative faecal immunochemical tests to guide referral for colorectal cancer in primary care [DG30]. Retrieved from https://www.nice.org.uk/guidance/DG30
NICE. (2019). Lung cancer: diagnosis and management [DG122]. Retrieved from https://www.nice.org.uk/guidance/ng122
NICE. (2021). Prostate cancer: diagnosis and management [DG131]. Retrieved from https://www.nice.org.uk/guidance/ng131
Nisbett, R. E., & Ross, L. (1980). Human inference: Strategies and shortcomings of social judgment.
Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies* (Vol. 11): sage.
Obadha, M., Barasa, E., Kazungu, J., Abiiro, G. A., & Chuma, J. (2019). Attribute development and level selection for a discrete choice experiment to elicit the preferences of health care providers for capitation payment mechanism in Kenya. *Health economics review, 9*(1), 1-19.
Oberlin, A. M., Pasipamire, T., & Chibwesha, C. J. (2019). Exploring women's preferences for HPV-based cervical cancer screening in South Africa. *International Journal of Gynecology & Obstetrics*.
Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. (2015). Improving risk understanding across ability levels: Encouraging active processing with dynamic icon arrays. *Journal of Experimental Psychology: Applied, 21*(2), 178.

Okeji, M. C., Agwuna, K. K., Ihudiebube-Splendor, C. N., Izge, I. Y., Ekuma, K. K., & Emeter, J. O. (2017). Transvaginal Sonography: perception and attitude of Nigerian women. *BMC women's health, 17*(1), 1-4.
ONS. (2019a). Cancer registration statistics, England: 2017. Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/2017
ONS. (2019b). Cancer survival in England - adults diagnosed. Retrieved from https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancersurvivalratescancersurvivalinenglandadultsdiagnosed
ONS. (2021). *Population estimates by ethnic group, England and Wale*. Retrieved from: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/datasets/populationestimatesbyethnicgroupenglandandwales
Ortiz-Muñoz, B., Aznar-Oroval, E., García, A. G., Peris, A. C., Ballestero, P. P., Yepes, M. S., . . . Garcia, E. G. (2014). HE4, Ca125 and ROMA algorithm for differential diagnosis between benign gynaecological diseases and ovarian cancer. *Tumor Biology, 35*(7), 7249-7258.
Osborne, J. M., Flight, I., Wilson, C. J., Chen, G., Ratcliffe, J., & Young, G. P. (2018). The impact of sample type and procedural attributes on relative acceptability of different colorectal cancer screening regimens. *Patient Preference and Adherence, 12*, 1825-1836. doi:http://dx.doi.org/10.2147/PPA.S172143
Ostermann, J., Brown, D. S., de Bekker-Grob, E. W., Mühlbacher, A. C., & Reed, S. D. (2017). Preferences for health interventions: improving uptake, adherence, and efficiency. *The Patient-Patient-Centered Outcomes Research, 10*(4), 511-514.
Palan, S., & Schitter, C. (2018). Prolific. ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22-27.
Pan, X., & Zuo, Z. (2020). Exploring the role of utility-difference threshold in choice behavior: An empirical case study of bus service choice. *International Journal of Transportation Science and Technology, 9*(2), 128-136.
Papin-Lefebvre, F., Guillaume, E., Moutel, G., Launoy, G., & Berchi, C. (2017). General practitioners' preferences with regard to colorectal cancer screening organisation Colon cancer screening medico-legal aspects. *Health policy, 121*(10), 1079-1084. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1016/j.healthpol.2017.08.013
Park, S. H., Lee, B., Lee, S., Choi, E., Choi, E.-B., Yoo, J., . . . Choi, K. S. (2015). A qualitative study of women's views on overdiagnosis and screening for thyroid cancer in Korea. *BMC cancer, 15*(1), 1-8.
Pavlik, E. J., van Nagell Jr, J. R., DePriest, P. D., Wheeler, L., Tatman, J. M., Boone, M., . . . Kryscio, R. K. (1995). Participation in transvaginal ovarian cancer screening: compliance, correlation factors, and costs. *Gynecologic oncology, 57*(3), 395-400.
Pearce, A. M., Mulhern, B. J., Watson, V., & Viney, R. C. (2020). How are debriefing questions used in health discrete choice experiments? An online survey. *Value in health, 23*(3), 289-293.
Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153-163.
Peitzmeier, S. M., Reisner, S. L., Harigopal, P., & Potter, J. (2014). Female-to-male patients have high prevalence of unsatisfactory Paps compared to non-transgender females: implications for cervical cancer screening. *Journal of General Internal Medicine, 29*(5), 778-784.
Pellat, A., Deyra, J., Coriat, R., & Chaussade, S. (2018). Results of the national organised colorectal cancer screening program with FIT in Paris. *Scientific reports, 8*(1), 1-4.
Peters, Y., & Siersema, P. D. (2020). Public Preferences and Predicted Uptake for Esophageal Cancer Screening Strategies: A Labeled Discrete Choice Experiment. *Clinical and translational gastroenterology, 11*(11).
Peterson, E. B., Ostroff, J. S., DuHamel, K. N., D'Agostino, T. A., Hernández, M., Canzona, M. R., & Bylund, C. L. (2016). Impact of provider-patient communication on cancer screening adherence: a systematic review. *Preventive Medicine, 93*, 96-105.
Philips, Z., Ginnelly, L., Sculpher, M., Claxton, K., Golder, S., Riemsma, R., . . . Glanville, J. (2004). Review of guidelines for good practice in decision-analytic modelling in health technology assessment. In *NIHR Health Technology Assessment programme: Executive Summaries*: NIHR Journals Library.
Phillips, K., Hersch, J., Turner, R., Jansen, J., & McCaffery, K. (2016). The influence of the 'cancer effect'on young women's responses to overdiagnosis in cervical screening. *Patient Education and Counseling, 99*(10), 1568-1575.

Phillips, K., Van, S. B., Marshall, D., Walsh, J., & Thabane, L. (2006). A review of studies examining stated preferences for cancer screening. In: Centers for Disease Control and Prevention.
Phillips, K. A., Johnson, F. R., & Maddala, T. (2002). Measuring what people value: a comparison of "attitude" and "preference" surveys. *Health services research, 37*(6), 1659-1679.
Pignone, M. P., Brenner, A. T., Hawley, S., Sheridan, S. L., Lewis, C. L., Jonas, D. E., & Howard, K. (2012). Conjoint analysis versus rating and ranking for values elicitation and clarification in colorectal cancer screening. *Journal of General Internal Medicine, 27*(1), 45-50. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1007/s11606-011-1837-z
Pignone, M. P., Crutchfield, T. M., Brown, P. M., Hawley, S. T., Laping, J. L., Lewis, C. L., . . . Wheeler, S. B. (2014). Using a discrete choice experiment to inform the design of programs to promote colon cancer screening for vulnerable populations in North Carolina. *BMC health services research, 14*, 611. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1186/s12913-014-0611-4
Pignone, M. P., Howard, K., Brenner, A. T., Crutchfield, T. M., Hawley, S. T., Lewis, C. L., & Sheridan, S. L. (2013). Comparing 3 techniques for eliciting patient values for decision making about prostate-specific antigen screening: A randomized controlled trial. *JAMA Internal Medicine, 173*(5), 362-368. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1001/jamainternmed.2013.2651
Pinsky, P. F., Yu, K., Kramer, B. S., Black, A., Buys, S. S., Partridge, E., . . . Prorok, P. C. (2016). Extended mortality results for ovarian cancer screening in the PLCO trial with median 15 years follow-up. *Gynecologic oncology, 143*(2), 270-275.
Pinto-Prades, J. L., McHugh, N., Donaldson, C., & Manoukian, S. (2019). Sequence effects in time trade-off valuation of hypothetical health states. *Health economics, 28*(11), 1308-1319.
Plott, C. R. (1993). Rational individual behavior in markets and social choice processes.
Plott, C. R., & Zeiler, K. (2005). The willingness to pay-willingness to accept gap, the" endowment effect," subject misconceptions, and experimental procedures for eliciting valuations. *American Economic Review, 95*(3), 530-545.
Plumb, A. A., Boone, D., Fitzke, H., Helbren, E., Mallett, S., Zhu, S., . . . Halligan, S. (2014). Detection of extracolonic pathologic findings with CT colonography: a discrete choice experiment of perceived benefits versus harms. *Radiology, 273*(1), 144-152. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1148/radiol.14131678
Polaris, J. J. Z., & Katz, J. N. (2014). "Appropriate" diagnostic testing: supporting diagnostics with evidence-based medicine and shared decision making. *BMC Research Notes, 7*(1), 922. doi:10.1186/1756-0500-7-922
Porta, M., Gallen, M., Belloc, J., & Malats, N. (1996). Predictors of the interval between onset of symptoms and first medical visit in patients with digestive tract cancer. *Int J Oncol, 8*(5), 941-949. doi:10.3892/ijo.8.5.941
Potoglou, D., Burge, P., Flynn, T., Netten, A., Malley, J., Forder, J., & Brazier, J. E. (2011). Best–worst scaling vs. discrete choice experiments: an empirical comparison using social care data. *Social science & medicine, 72*(10), 1717-1727.
Price, S., McManus, J., & Barrett, J. (2019). The transgender population: improving awareness for gynaecologists and their role in the provision of care. *The Obstetrician & Gynaecologist, 21*(1), 11-20.
Puliti, D., Duffy, S. W., Miccinesi, G., De Koning, H., Lynge, E., Zappa, M., & Paci, E. (2012). Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. *Journal of Medical Screening, 19*(1_suppl), 42-56.
Raginel, T., Grandazzi, G., Launoy, G., Trocmé, M., Christophe, V., Berchi, C., & Guittet, L. (2020). Social inequalities in cervical cancer screening: a discrete choice experiment among French general practitioners and gynaecologists. *BMC health services research, 20*(1), 1-10.
Ragland, M., Trivers, K. F., Andrilla, C. H. A., Matthews, B., Miller, J., Lishner, D., . . . Baldwin, L.-M. (2018). Physician nonprofessional cancer experience and ovarian cancer screening practices: results from a national survey of primary care physicians. *Journal of Women's Health, 27*(11), 1335-1341.
Ramezani Doroh, V., Delavari, A., Yaseri, M., Sefiddashti, S. E., & Akbarisari, A. (2019). Preferences of Iranian average risk population for colorectal cancer screening tests. *International journal of health care quality assurance*.
Ran, T., Cheng, C.-Y., Misselwitz, B., Brenner, H., Ubels, J., & Schlander, M. (2019). Cost-effectiveness of colorectal cancer screening strategies—a systematic review. *Clinical Gastroenterology and Hepatology, 17*(10), 1969-1981. e1915.

Robinson, A., Spencer, A., & Moffatt, P. (2015). A framework for estimating health state utility values within a discrete choice experiment: modeling risky choices. *Medical Decision Making, 35*(3), 341-350.
Robinson, K. M., Christensen, K. B., Ottesen, B., & Krasnik, A. (2012). Diagnostic delay, quality of life and patient satisfaction among women diagnosed with endometrial or ovarian cancer: a nationwide Danish study. *Quality of Life Research, 21*(9), 1519-1525.
Rolfe, J., Windle, J., & Bennett, J. (2015). Benefit transfer: insights from choice experiments. In *Benefit Transfer of Environmental and Resource Values* (pp. 191-208): Springer.
Rose, J., & Hensher, D. A. (2004). Modelling agent interdependency in group decision making. *Transportation Research Part E: Logistics and Transportation Review, 40*(1), 63-79.
Rose, J. M., & Bliemer, M. C. (2009). Constructing efficient stated choice experimental designs. *Transport Reviews, 29*(5), 587-617.
Rose, J. M., & Bliemer, M. C. (2013). Sample size requirements for stated choice experiments. *Transportation, 40*(5), 1021-1041.
Rosenberger, R. S., Peterson, G. L., Clarke, A., & Brown, T. C. (2003). Measuring dispositions for lexicographic preferences of environmental goods: integrating economics, psychology and ethics. *Ecological Economics, 44*(1), 63-76.
Round, T., Gildea, C., Ashworth, M., & Møller, H. (2020). Association between use of urgent suspected cancer referral and mortality and stage at diagnosis: a 5-year national cohort study. *British Journal of General Practice, 70*(695), e389. doi:10.3399/bjgp20X709433
Rozbroj, T., Haas, R., O'Connor, D., Carter, S. M., McCaffery, K., Thomas, R., . . . Buchbinder, R. (2021). How do people understand overtesting and overdiagnosis? Systematic review and meta-synthesis of qualitative research. *Social science & medicine, 285*, 114255.
Rupert, D. J., Poehlman, J. A., Hayes, J. J., Ray, S. E., & Moultrie, R. R. (2017). Virtual versus in-person focus groups: Comparison of costs, recruitment, and participant logistics. *Journal of Medical Internet Research, 19*(3), e6980.
Russell, M. (2005). Does Patient ethnicity or Sonographer gender have any bearing on Patient acceptability of transvaginal ultrasound? *Ultrasound, 13*(3), 170-172.
Ryan, M. (2004). Discrete choice experiments in health care. In: British Medical Journal Publishing Group.
Ryan, M., & Bate, A. (2001). Testing the assumptions of rationality, continuity and symmetry when applying discrete choice experiments in health care. *Applied Economics Letters, 8*(1), 59-63.
Ryan, M., Bate, A., Eastmond, C. J., & Ludbrook, A. (2001). Use of discrete choice experiments to elicit preferences. *BMJ Quality & Safety, 10*(suppl 1), i55-i60.
Ryan, M., Gerard, K., & Amaya-Amaya, M. (2008). Discrete choice experiments in a nutshell. In *Using discrete choice experiments to value health and health care* (pp. 13-46): Springer.
Ryan, M., & Skåtun, D. (2004). Modelling non-demanders in choice experiments. *Health economics, 13*(4), 397-402.
Ryan, M., Watson, V., & Entwistle, V. (2009). Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. *Health economics, 18*(3), 321-336.
Ryan, M., & Wordsworth, S. (2000). Sensitivity of willingness to pay estimates to the level of attributes in discrete choice experiments. *Scottish Journal of Political Economy, 47*(5), 504-524.
Salika, T., Lyratzopoulos, G., Whitaker, K. L., Waller, J., & Renzi, C. (2017). Do comorbidities influence help-seeking for cancer alarm symptoms? A population-based survey in England. *Journal of Public Health, 40*(2), 340-349. doi:10.1093/pubmed/fdx072
Salkeld, G., Ryan, M., & Short, L. (2000). The Veil of Experience: Do Consumers Prefer What They Know Best? *Health economics, 9*(3), 267-270. doi:http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291099-1050/issues
Salkeld, G., Solomon, M., Short, L., Ryan, M., & Ward, J. E. (2003). Evidence-based consumer choice: A case study in colorectal cancer screening. *Australian and New Zealand Journal of Public Health, 27*(4), 449-455.
Salsman, J. M., Pavlik, E., Boerner, L. M., & Andrykowski, M. A. (2004). Clinical, demographic, and psychological characteristics of new, asymptomatic participants in a transvaginal ultrasound screening program for ovarian cancer. *Preventive Medicine, 39*(2), 315-322.
Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica, 5*(17), 61-71.
Sasieni, P., Castanon, A., & Cuzick, J. (2009). Effectiveness of cervical screening with age: population based case-control study of prospectively recorded data. *Bmj, 339*, b2968. doi:10.1136/bmj.b2968

Savage, S. J., & Waldman, D. M. (2008). Learning and fatigue during choice experiments: a comparison of online and mail survey modes. *Journal of Applied Econometrics, 23*(3), 351-371.
Scalia, P., Schubbe, D. C., Lu, E. S., Durand, M.-A., Frascara, J., Noel, G., . . . Elwyn, G. (2021). Comparing the impact of an icon array versus a bar graph on preference and understanding of risk information: Results from an online, randomized study. *PLoS One, 16*(7), e0253644.
Scarpa, R., Zanoli, R., Bruschi, V., & Naspetti, S. (2013). Inferred and stated attribute non-attendance in food choice experiments. *American Journal of Agricultural Economics, 95*(1), 165-180.
Schapira, M. M., Walker, C. M., Cappaert, K. J., Ganschow, P. S., Fletcher, K. E., McGinley, E. L., . . . Jacobs, E. A. (2012). The numeracy understanding in medicine instrument: a measure of health numeracy developed using item response theory. *Medical Decision Making, 32*(6), 851-865.
Schwartz, L. M., Woloshin, S., Sox, H. C., Fischhoff, B., & Welch, H. G. (2000). US women's attitudes to false positive mammography results and detection of ductal carcinoma in situ: cross sectional survey. *Bmj, 320*(7250), 1635-1640.
Schwartz, L. M., Woloshin, S., & Welch, H. G. (2005). Can patients interpret health information? An assessment of the medical data interpretation test. *Medical Decision Making, 25*(3), 290-300.
Schwartz, M., Lerman, C., Daly, M., Audrain, J., Masny, A., & Griffith, K. (1995). Utilization of ovarian cancer screening by women at increased risk. *Cancer Epidemiol Biomarkers Prev, 4*(3), 269-273.
Seay, J., Ranck, A., Weiss, R., Salgado, C., Fein, L., & Kobetz, E. (2017). Understanding transgender men's experiences with and preferences for cervical cancer screening: a rapid assessment survey. *LGBT health, 4*(4), 304-309.
Sen, A. (1993). Capability and well-being73. *The quality of life, 30*.
Shahidi, N., & Cheung, W. Y. (2016). Colorectal cancer screening: Opportunities to improve uptake, outcomes, and disparities. *World journal of gastrointestinal endoscopy, 8*(20), 733.
Sharma, A., Beveridge, H., Fallowfield, L., Jacobs, I., & Menon, U. (2006). Postmenopausal women undergoing transvaginal ultrasound screening prefer not to have chaperones. *BJOG: An International Journal of Obstetrics & Gynaecology, 113*(8), 954-957.
Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., . . . Kristjansson, E. (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *Bmj, 358*, j4008.
Shen, J. (2009). Latent class model or mixed logit model? A comparison by transport mode choice data. *Applied Economics, 41*(22), 2915-2924.
Shirley, M. H., Barnes, I., Sayeed, S., Finlayson, A., & Ali, R. (2014). Incidence of breast and gynaecological cancers by ethnic group in England, 2001–2007: a descriptive study. *BMC cancer, 14*(1), 1-10.
Sicsic, J., Krucien, N., & Franc, C. (2016). What are GPs' preferences for financial and non-financial incentives in cancer screening? Evidence for breast, cervical, and colorectal cancers. *Social Science and Medicine, 167*, 116-127. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1016/j.socscimed.2016.08.050
Sicsic, J., Pelletier-Fleury, N., & Moumjid, N. (2018). Women's Benefits and Harms Trade-Offs in Breast Cancer Screening: Results from a Discrete-Choice Experiment. *Value in health, 21*(1), 78-88. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1016/j.jval.2017.07.003
SIGN. (2018). SIGN 135. Management of epithelial ovarian cancer. A national clinical guideline. Retrieved from https://www.sign.ac.uk/media/1073/sign135_oct2018.pdf
Smith-Bindman, R. (2018). Use of advanced imaging tests and the not-so-incidental harms of incidental findings. *JAMA internal medicine, 178*(2), 227-228.
Smith, E. M., & Anderson, B. (1985). The effects of symptoms and delay in seeking diagnosis on stage of disease at diagnosis among women with cancers of the ovary. *Cancer, 56*(11), 2727-2732.
Smith, I. P., DiSantostefano, R. L., de Bekker-Grob, E. W., Levitan, B., Berlin, C., Veldwijk, J., & de Wit, G. A. (2021). Methodological Priorities for Patient Preferences Research: Stakeholder Input to the PREFER Public–Private Project. In (Vol. 14, pp. 449-453): Springer.
Smith, S. M., Campbell, N. C., MacLeod, U., Lee, A. J., Raja, A., Wyke, S., . . . Nicolson, M. C. (2009). Factors contributing to the time taken to consult with symptoms of lung cancer: a cross-sectional study. *Thorax, 64*(6), 523-531. doi:10.1136/thx.2008.096560
Snoswell, C. L., Whitty, J. A., Caffery, L. J., Loescher, L. J., Gillespie, N., & Janda, M. (2018). Direct-to-consumer mobile teledermoscopy for skin cancer screening: Preliminary results

demonstrating willingness-to-pay in Australia. *Journal of telemedicine and telecare, 24*(10), 683-689. doi:http://dx.doi.org/10.1177/1357633X18799582
Soekhai, V., de Bekker-Grob, E. W., Ellis, A. R., & Vass, C. M. (2019). Discrete choice experiments in health economics: past, present and future. *Pharmacoeconomics, 37*(2), 201-226.
Spinks, J., Janda, M., Soyer, H. P., & Whitty, J. A. (2016). Consumer preferences for teledermoscopy screening to detect melanoma early. *Journal of telemedicine and telecare, 22*(1), 39-46. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1177/1357633X15586701
Stevenson, F. A., Cox, K., Britten, N., & Dundar, Y. (2004). A systematic review of the research on communication between patients and health care professionals about medicines: the consequences for concordance. *Health Expect, 7*(3), 235-245. doi:10.1111/j.1369-7625.2004.00281.x
Stewart, T., Lee, Y. A., & Damiano, E. A. (2020). Do transgender and gender diverse individuals receive adequate gynecologic care? An analysis of a rural academic center. *Transgender health, 5*(1), 50-58.
Stone, D. N., & Schkade, D. A. (1991). Numeric and linguistic information representation in multiattribute choice. *Organizational Behavior and Human Decision Processes, 49*(1), 42-59.
Sud, A., Torr, B., Jones, M. E., Broggio, J., Scott, S., Loveday, C., . . . Jhanji, S. (2020). Effect of delays in the 2-week-wait cancer referral pathway during the COVID-19 pandemic on cancer survival in the UK: a modelling study. *The Lancet Oncology, 21*(8), 1035-1044.
Sundar, S., Rick, C., Dowling, F., Au, P., Snell, K., Rai, N., . . . Davenport, C. (2016). Refining Ovarian Cancer Test accuracy Scores (ROCkeTS): protocol for a prospective longitudinal test accuracy study to validate new risk scores in women with symptoms of suspected ovarian cancer. *BMJ open, 6*(8), e010333.
Tahmasebi, F., Nath, R., Sokolovsky, N., Scaffidi, J., & Boley, J. (2018). Incidental finding of raised CA125: a cause for concern. *Crit Care Obst Gyne, 5*(01), 1-5.
Tamirisa, N. P., Goodwin, J. S., Kandalam, A., Linder, S. K., Weller, S., Turrubiate, S., . . . Riall, T. S. (2017). Patient and physician views of shared decision making in cancer. *Health Expectations, 20*(6), 1248-1253. doi:https://doi.org/10.1111/hex.12564
Tang, Y.-T., & Chooi, W.-T. (2021). A systematic review of the effects of positive versus negative framing on cancer treatment decision making. *Psychology & health*, 1-26.
Target Ovarian Cancer. (2016). *Pathfinder 2016. Transforming futures for women with ovarian cancer*. Retrieved from https://targetovariancancer.org.uk/sites/default/files/2020-07/Pathfinder%202016.pdf
Target Ovarian Cancer. (2019). Smear tests won't detect ovarian cancer. Retrieved from https://targetovariancancer.org.uk/news/smear-tests-wont-detect-ovarian-cancer
Target Ovarian Cancer. (2022a). Campaigns on awareness. Retrieved from https://targetovariancancer.org.uk/get-involved/campaign/current-campaigns/campaign-awareness
Target Ovarian Cancer. (2022b). Campaigns on diagnosis. Retrieved from https://targetovariancancer.org.uk/get-involved/campaign/current-campaigns/campaign-awareness
Tariq, S., & Woodman, J. (2013). Using mixed methods in health research. *JRSM short reports, 4*(6), 2042533313479197.
Tervonen, T., Schmidt-Ott, T., Marsh, K., Bridges, J. F., Quaife, M., & Janssen, E. (2018). Assessing rationality in discrete choice experiments in health: an investigation into the use of dominance tests. *Value in health, 21*(10), 1192-1197.
Teti, M., Kerr, S., Bauerband, L., Koegler, E., & Graves, R. (2021). A Qualitative Scoping Review of Transgender and Gender Non-conforming People's Physical Healthcare Experiences and Needs. *Frontiers in Public Health, 9*.
Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review, 34*(4), 273.
Timmis, L. (2020). *Stakeholder perspectives on optimal follow-up care: An interdisciplinary economic and psychological investigation into gynaecological cancer.* (Doctor of Philosophy), Bangor University,
Tokuda, Y., Chinen, K., Obara, H., & Joishy, S. K. (2009). Intervals between symptom onset and clinical presentation in cancer patients. *Intern Med, 48*(11), 899-905. doi:10.2169/internalmedicine.48.1720
Torrance, G. W. (1976). Social preferences for health states: an empirical evaluation of three measurement techniques. *Socio-economic planning sciences, 10*(3), 129-136.

Torrance, G. W., Feeny, D., & Furlong, W. (2001). Visual analog scales: do they have a role in the measurement of preferences for health states? In: Sage Publications Sage CA: Thousand Oaks, CA.
Train, K. E. (2009). *Discrete choice methods with simulation*: Cambridge university press.
Trevena, L. J., Bonner, C., Okan, Y., Peters, E., Gaissmaier, W., Han, P. K., . . . Zikmund-Fisher, B. J. (2021). Current challenges when using numbers in patient decision aids: advanced concepts. *Medical Decision Making, 41*(7), 834-847.
Trevena, L. J., BPsych, H. M. D., Barratt, A., Butow, P., & Caldwell, P. (2006). A systematic review on communicating with patients about evidence. *Journal of evaluation in clinical practice, 12*(1), 13-23.
Tversky, A., & Shafir, E. (1992). Choice under conflict: The dynamics of deferred decision. *Psychological science, 3*(6), 358-361.
Ubel, P. A., Loewenstein, G., & Jepson, C. (2003). Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. *Quality of life Research, 12*(6), 599-607.
UK National Screening Committee. (2015). Criteria for appraising the viability, effectiveness and appropriateness of a screening programme. Retrieved from https://www.gov.uk/government/publications/evidence-review-criteria-national-screening-programmes/criteria-for-appraising-the-viability-effectiveness-and-appropriateness-of-a-screening-programme
van Dam, L., Hol, L., Bekker-Grob, E. W. d., Steyerberg, E. W., Kuipers, E. J., Habbema, J. D. F., . . . Leerdam, M. E. v. (2010). What determines individuals' preferences for colorectal cancer screening programmes? A discrete choice experiment. *European Journal of Cancer, 46*(1), 150-159. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1016/j.ejca.2009.07.014
van Nagell Jr, J. R., DePriest, P. D., Ueland, F. R., DeSimone, C. P., Cooper, A. L., McDonald, J. M., . . . Kryscio, R. J. (2007). Ovarian cancer screening with annual transvaginal sonography: findings of 25,000 women screened. *Cancer, 109*(9), 1887-1896.
van Nagell Jr, J. R., Miller, R. W., DeSimone, C. P., Ueland, F. R., Podzielinski, I., Goodrich, S. T., . . . Pavlik, E. J. (2011). Long-term survival of women with epithelial ovarian cancer detected by ultrasonographic screening. *Obstetrics & Gynecology, 118*(6), 1212-1221.
Vass, C., Rigby, D., Tate, K., Stewart, A., & Payne, K. (2018a). An exploratory application of eye-tracking methods in a discrete choice experiment. *Medical Decision Making, 38*(6), 658-672.
Vass, C. M., Davison, N. J., Vander Stichele, G., & Payne, K. (2020). A picture is worth a thousand words: the role of survey training materials in stated-preference studies. *The Patient-Patient-Centered Outcomes Research, 13*(2), 163-173.
Vass, C. M., & Payne, K. (2017). Using discrete choice experiments to inform the benefit-risk assessment of medicines: are we ready yet? *Pharmacoeconomics, 35*(9), 859-866.
Vass, C. M., Wright, S., Burton, M., & Payne, K. (2018b). Scale heterogeneity in healthcare discrete choice experiments: a primer. *The Patient-Patient-Centered Outcomes Research, 11*(2), 167-173.
Veldwijk, J., Lambooij, M. S., de Bekker-Grob, E. W., Smit, H. A., & de Wit, G. A. (2014). The effect of including an opt-out option in discrete choice experiments. *PLoS ONE, 9*(11), e111805.
Viswanathan, M., & Childers, T. L. (1996). Processing of numerical and verbal product information. *Journal of Consumer Psychology, 5*(4), 359-385.
Waddell, A., Lennox, A., Spassova, G., & Bragge, P. (2021). Barriers and facilitators to shared decision-making in hospitals from policy to practice: a systematic review. *Implementation Science, 16*(1), 74. doi:10.1186/s13012-021-01142-y
Waller, J., Douglas, E., Whitaker, K. L., & Wardle, J. (2013). Women's responses to information about overdiagnosis in the UK breast cancer screening programme: a qualitative study. *BMJ open, 3*(4), e002703.
Wang, J., Andrae, B., Sundström, K., Ploner, A., Ström, P., Elfström, K. M., . . . Sparén, P. (2017). Effectiveness of cervical screening after age 60 years according to screening history: Nationwide cohort study in Sweden. *PLoS medicine, 14*(10), e1002414. doi:10.1371/journal.pmed.1002414
Wang, X., Piantadosi, S., Le-Rademacher, J., & Mandrekar, S. J. (2021). Statistical considerations for subgroup analyses. *Journal of Thoracic Oncology, 16*(3), 375-380.
Webb, E. J., Meads, D., Lynch, Y., Judge, S., Randall, N., Goldbart, J., . . . Murray, J. (2021). Attribute selection for a discrete choice experiment incorporating a best-worst scaling survey. *Value in health, 24*(4), 575-584.

Weinstein, M. C., Siegel, J. E., Gold, M. R., Kamlet, M. S., & Russell, L. B. (1996). Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *Jama, 276*(15), 1253-1258.
Westwood, M., Ramaekers, B., Lang, S., Grimm, S., Deshpande, S., de Kock, S., . . . Kleijnen, J. (2018). Risk scores to guide referral decisions for people with suspected ovarian cancer in secondary care: a systematic review and cost-effectiveness analysis.
Whitaker, K. L., Ghanouni, A., Zhou, Y., Lyratzopoulos, G., & Morris, S. (2017). Patients' preferences for GP consultation for perceived cancer risk in primary care: a discrete choice experiment. *British Journal of General Practice, 67*(659), e388-e395.
Whitty, J. A., Walker, R., Golenko, X., & Ratcliffe, J. (2014). A think aloud study comparing the validity and acceptability of discrete choice and best worst scaling methods. *PLoS One, 9*(4), e90635. doi:10.1371/journal.pone.0090635
Willis, G. B. (2004). *Cognitive interviewing: A tool for improving questionnaire design*: sage publications.
Wilson, J. M. G., & Jungner, G. (1968). Principles and practice of screening for disease.
Woller-Carter, M. M., Okan, Y., Cokely, E. T., & Garcia-Retamero, R. (2012). *Communicating and distorting risks with graphs: An eye-tracking study.* Paper presented at the Proceedings of the human factors and ergonomics society annual meeting.
Woloshin, S., & Schwartz, L. M. (2011). Communicating data about the benefits and harms of treatment: a randomized trial. *Annals of internal medicine, 155*(2), 87-96.
Woof, V. G., Ruane, H., French, D. P., Ulph, F., Qureshi, N., Khan, N., . . . Donnelly, L. S. (2020). The introduction of risk stratified screening into the NHS breast screening Programme: views from British-Pakistani women. *BMC cancer, 20*(1), 452. doi:10.1186/s12885-020-06959-2
World Ovarian Cancer Coalition. (2018). *The Every Women Study. Summary Report*. Retrieved from https://worldovariancancercoalition.org/wp-content/uploads/2019/11/WOCC-Every-Woman-Study-Summary-Report-Nov-08.pdf
Wortley, S., Wong, G., Kieu, A., & Howard, K. (2014). Assessing stated preferences for colorectal cancer screening: A critical systematic review of discrete choice experiments. *Patient, 7*(3), 271-282. doi:http://0-dx.doi.org.lib.exeter.ac.uk/10.1007/s40271-014-0054-3
Xiao, J., Wang, F., Wang, M., & Ma, Z. (2022). Attribute nonattendance in COVID-19 vaccine choice: A discrete choice experiment based on Chinese public preference. *Health Expectations*.
Yegoryan, N., Guhl, D., & Klapper, D. (2020). Inferring attribute non-attendance using eye tracking in choice-based conjoint analysis. *Journal of Business Research, 111*, 290-304.
Yepes-Nuñez, J. J., Zhang, Y., Xie, F., Alonso-Coello, P., Selva, A., Schünemann, H., & Guyatt, G. (2017). Forty-two systematic reviews generated 23 items for assessing the risk of bias in values and preferences' studies. *Journal of clinical epidemiology, 85*, 21-31.
Zhao, Z., Du, L., Wang, L., Wang, Y., Yang, Y., & Dong, H. (2021). Preferred lung cancer screening modalities in China: a discrete choice experiment. *Cancers, 13*(23), 6110.
Zipkin, D. A., Umscheid, C. A., Keating, N. L., Allen, E., Aung, K., Beyth, R., . . . Korenstein, D. (2014). Evidence-based risk communication: a systematic review. *Annals of internal medicine, 161*(4), 270-280.

Appendix 9.2: Best-worst scaling questionnaire

Understanding priorities towards ovarian cancer screening (BWS)

INFORMATION FOR PARTICIPANTS

You are being invited to take part in a research study that aims to understand attitudes around screening for possible ovarian cancer. Before you decide whether to take part in this study it is important for you to understand why the research is being done and what taking part will involve.

Please read the following information carefully and take time to consider whether you wish to take part.

What is the aim of the project?

In particular, we want to learn what characteristics of diagnostic testing are most important. Screening for cancer can help to identify the disease at an earlier stage where it is more curable. Currently the NHS does not offer screening for ovarian cancer but research to develop a suitable test are ongoing. In this study we want to learn what characteristics of a possible test are most important to people who may be eligible for screening if/when a test becomes available.

Why have I been invited to take part?

We are approaching you because we are seeking responses from women and people with ovaries over the age of 40. You must be able to complete the survey in English to take part. **Please do not take part in this survey if you have ever undergone a procedure to remove both of your ovaries.**

What will I be asked to do?

Should you agree to take part, you will be asked to complete a survey lasting about 15 minutes. The survey includes questions about your background, health history and your views on ovarian cancer testing. There are no right or wrong answers- the purpose of the survey is to better understand the opinion of women.

What are the possible disadvantages and risks of taking part?

Participating in the research is not anticipated to cause you any disadvantages or discomfort. Some questions may be considered sensitive, however, you do not have to provide responses to any questions you feel uncomfortable answering. During the survey you will be asked to imagine you might have cancer, for some this may cause anxiety.

Will I be paid?

You will receive a payment of £1.75 for completing the survey. Payments will be paid directly into your Prolific account within 10 working days.

Please note: In order to receive the payment you must complete the full survey and click the completion link at the end of the survey. To ensure responses are high quality, three "attention check" questions have been included throughout the survey. You must correctly complete all three questions to receive payment.

Can I change my mind and withdraw from the project?

If you decide you no longer wish to take part during the survey, simply exit the webpage to withdraw. Your incomplete responses will be permanently deleted. If you decide to withdraw after submitting your responses, please contact us via your Prolific account or directly by email. You can withdraw from the study for up to 14 days after completion and do no not have to give a reason. After 14 days it may no longer be possible to withdraw your submission because anonymization will mean we cannot link responses to individual participants.

Is the survey confidential?

All your answers to the survey are completely confidential and anonymous. You will not be asked your name or any other identifying information. Your responses will be securely stored on an encrypted password protected computer and managed according to a law called the Data Protection Act (2018). Your anonymised data will be stored for a period of five years.

The results of the study may be published in academic journals or conferences but any included data will not be individually identifiable.

The University of Exeter processes personal data for the purposes of carrying out research in the public interest. The University will endeavour to be transparent about its processing of your personal data and this information sheet should provide a clear explanation of this. If you do have any queries about the University's processing of your personal data that cannot be resolved by the research team, further information may be obtained from the University's Data Protection Officer by emailing dataprotection@exeter.ac.uk or at www.exeter.ac.uk/dataprotection

What if I have any questions?

If you have any questions about our project, either now or in the future, please feel free to contact Rebekah Hall by emailing rh591@exeter.ac.uk or Anne Spencer a.e.spencer@exeter.ac.uk

---

This project has been reviewed and approved by the University of Exeter UEBS ethics comittee application number (eUEBS003725v4.2)

Consent Form

- I understand that my participation is voluntary and that I am free to withdraw for up to 14 days without giving any reason and without my legal rights being affected.
- I understand that my data from the study will be fully anonymised and will be looked at by members of the research team and may potentially be shared with other researchers in future if appropriate.
- I understand that relevant sections of the data collected during the study may be looked at by individuals from the University of Exeter, Cancer Research UK or regulatory authorities for audit purposes
- I understand that the results of the study may be published in academic journals but my anonymity will be preserved
- I understand that my anonymised data will be securely stored on an encrypted password protected computed for a period of five years.
- I understand that in order to receive payment for this survey I must complete the full survey and click the link at the end of the survey. I must also correctly complete three attention check questions.

I confirm that I have read the information above and agree to take part in the study:

- ○ Yes
- ○ No

Previous Next

Please enter your ProlificID

Previous Next

Have you ever had a medical procedure that involved the removal of both of your ovaries?

- ○ Yes
- ○ No

Previous Next

In this survey we would like to understand what characteristics a screening test for ovarian cancer are most important to individuals who may be eligible for screening in the future if a test does become available.

What is ovarian cancer?

- Ovarian cancer occurs when the cells in and around the ovaries and fallopian tubes become abnormal, grow out of control and form a lump called a "tumour".
- Ovarian cancer is the 6th most common cancer for women in the UK. Over 7,500 women are diagnosed annually. Most of these cases occur in women over the age of 40.
- Each year approximately 4,200 people will die from ovarian cancer.
- Diagnosing ovarian cancer early means more treatment options are available and increases the chance of being cured or living longer.
- Screening people for ovarian cancer may be one way to diagnose people earlier and improve the chance of survival.

What is cancer screening?

- Cancer screening is a test that looks for early signs of cancer in people without symptoms.
- It can help spot cancers at an early stage, when treatment is more likely to be successful.
- In some cases screening may also prevent cancer from developing (for example cervical screening can identify pre-cancerous cells)
- In the UK there are screening tests for three types of cancer; breast cancer, cervical cancer and colorectal cancer
- There is currently no screening test for ovarian cancer, however, developing a suitable test is an area of ongoing research.

Previous

Next

Section 1: Background questions

In this section, we are interested in getting to know you a little better by asking a few background questions.

*What is your age?

*Which of the following best describes your current relationship status?

- ○ Married
- ○ Widowed
- ○ Divorced
- ○ Separated
- ○ In a domestic partnership or civil union
- ○ Single, but cohabiting with a significant other
- ○ Single, never married
- ○ I prefer not to say

What is your ethnicity?

- ○ White/Caucasian
- ○ Mixed-White and Black African
- ○ Mixed- White and Black Caribbean
- ○ Mixed- White and Asian
- ○ Asian- Indian
- ○ Asian- Pakistani
- ○ Asian- Bangladeshi
- ○ Asian-Chinese
- ○ Black-African
- ○ Black- Caribbean
- ○ Arab
- ○ I prefer not to say
- ○ Other:

Which of the following categories best describes your employment status?

- ○ Employed, working full-time
- ○ Employed, working part-time
- ○ Self-employed or freelance
- ○ Not employed, looking for work
- ○ Not employed, NOT looking for work
- ○ Not employed, NOT looking for work
- ○ Retired
- ○ Long-term sick or disabled, not able to work
- ○ Student
- ○ Volunteering
- ○ Looking after home or relative
- ○ Prefer not to say

***What is your highest level of education?**

- ○ 1-4 O levels/CSEs/GCSEs (any grades), Entry level, Foundation Diploma
- ○ NVQ Level 1, Foundation GNVQ, Basic Skills
- ○ 5+ O levels (passes)/ CSEs (grade 1)/ GCSEs (grades A*-C), School Certificate, 1 A level/ 2-3 AS levels/VCEs, Higher Diploma
- ○ NVQ Level 2, Intermediate GNVQ, City and Guilds Craft, BTEC First/General Diploma, RSA Diploma
- ○ Apprenticeship
- ○ 2+ A levels/VCEs, 4+ AS levels, Higher School Certificate, Progression/Advanced Diploma
- ○ NVQ Level 3, Advanced GNVQ, City and Guilds Advanced Craft, ONC, OND, BTEC National, RSA Advanced Diploma
- ○ Undergraduate degree (BA, BSc)
- ○ Postgraduate degree (e.g. MSc, MA, PhD)
- ○ NVQ Level 4-5, HNC, HND, RSA Higher Diploma, BTEC Higher Level
- ○ Professional qualifications (e.g. teaching, nursing, accounting)
- ○ Other vocational/ work-related qualifications
- ○ Foreign qualifications
- ○ No qualifications
- ○ I prefer not to say

***What is the total annual income of your household (before tax)?**

- ○ Prefer not to say
- ○ £0-£9,999
- ○ £10,000-£19,999
- ○ £20,000-£29,999
- ○ £30,000-£39,999
- ● £40,000-£49,999
- ○ £50,000-£59,999
- ○ £60,000-£69,999
- ○ £70,000+

Previous

Next

Section 2- Your views on ovarian cancer screening

In the next set of questions, we want to learn more about what you think about ovarian cancer screening tests.

In each question you will be shown a list of features based on the types of tests that might be offered to women in the future if ovarian cancer screening becomes available.

For each question, we would like you to select which feature of a medical test you think is **most important** and which feature is **least important** if you were given a choice about what test to take. The features will change between questions so please only consider the features listed in each case. Please only choose **one** feature as most important and **one** feature as the least important.

During this section we are trying to measure your preferences for 16 different features of screening tests. To do this we need to ask quite a lot of very similar questions. The task may start to feel repetitive but each question is slightly different and asks about different combinations of the features we are interested in hearing your opinions on. Please take your time and try and pay close attention.

There are no right or wrong answers. We are simply interested in your opinion.

To help you get used to this style of question, please complete the example question about choosing a restaurant to eat in below:

*Considering only the features below, which is the **most important** and which is the **least important** when choosing a restaurant to eat at?

Most important		Least important
○	Distance from home	○
○	Taste of the food	○
○	Price of the food	○
○	Customer service	○
○	Variety of food on the menu	○

Previous

Next

Section 2- Your views on ovarian cancer testing

For each question, we would like you to select which feature of a medical test you think is **most important** and which feature is **least important** if you were given a choice about what test to take. The features will change between questions so please only consider the features listed in each case. Please only choose **one** feature as most important and **one** feature as the least important.

You can hover over each feature for a more detailed description

Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Type of test	○
○	Accuracy: chance of *false-positive* result	○
○	Chance of cancer diagnosis	○
○	Waiting time for the test	○

Hover over each feature for a more detailed description

*Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Action required by you to arrange the result	○
○	Test location	○
○	Type of test	○
○	Chance of needing a follow up test	○

Hover over each feature for a more detailed description

*Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Waiting time for the test	○
○	Chance of being unnecessarily diagnosed and treated for a cancer that would never have caused symptoms or death	○
○	Screening interval	○
○	Action required by you to arrange the test	○

Hover over each feature for a more detailed description

*Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Test location	○
○	Screening interval	○
○	Accuracy: chance of *false-negative* result	○
○	Accuracy: chance of a *false-positive* result	○

Hover over each feature for a more detailed description

*Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Chance of a follow up test	○
○	Waiting time for the test/span>	○
○	Chance of dying from ovarian cancer	○
○	Accuracy: chance of *false-negative* result	○

Hover over each feature for a more detailed description

*Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Waiting time for the result	○
○	Who performs the test	○
○	Waiting time for the test	○
○	Test location	○

Hover over each feature for a more detailed description

*Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Screening interval	○
○	Chance of cancer diagnosis	○
○	Chance of needing a follow up test	○
○	Waiting time for the result	○

Hover over each feature for a more detailed description

*Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Accuracy: chance of *false-negative* result	○
○	Waiting time for the result	○
○	Chance of being unnecessarily diagnosed and treated for a cancer that would never have caused symptoms or death	○
○	Type of test	○

Hover over each feature for a more detailed description

*Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Chance of dying from ovarian cancer	○
○	Type of test	○
○	Who performs the test	○
○	Screening interval	○

Hover over each feature for a more detailed description

*Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Accuracy: chance of a *false-positive* result	○
○	Action required by you to arrange the result	○
○	Waiting time for the result	○
○	Chance of dying from ovarian cancer	○

Hover over each feature for a more detailed description

*Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Who performs the test	○
○	Chance of needing a follow up test	○
○	Accuracy: chance of a *false-positive* result	○
○	Chance of being unnecessarily diagnosed and treated for a cancer that would never have caused symptoms or death	○

Hover over each feature for a more detailed description

*Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Chance of being unnecessarily diagnosed and treated for a cancer that would never have caused symptoms or death	○
○	Chance of dying from ovarian cancer	○
○	Test location	○
○	Chance of cancer diagnosis	○

Hover over each feature for a more detailed description

*Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Type of test	○
○	Test location	○
○	Attention check- please select as most important	○
○	Chance of cancer diagnosis	○

Hover over each feature for a more detailed description

*Considering only the features below, which is the **most important** and which is the **least important** when choosing a test for possible ovarian cancer?

Most important		Least important
○	Chance of cancer diagnosis	○
○	Accuracy: chance of *false-negative* result	●
●	Action required by you to arrange the test	○
○	Who performs the test	○

Hover over each feature for a more detailed description

Previous Next

How easy or difficult did you find making choices between the most and least important characteristics in the last section?

- ○ Very easy
- ○ Easy
- ○ Neither easy nor difficult
- ○ Difficult
- ○ Very difficult

During the last section you were shown different combinations of 16 features related to screening for possible ovarian cancer. You have already shown that some features are more important than others to you but were there any features that you would consider completely unimportant?

Please select any features that would never be an important issue to you if you were making a decision about whether to undergo a screening for possible ovarian cancer.

- [ ] Type of test
- [ ] Accuracy: chance of false-negative result
- [ ] Accuracy: chance of a false-positive result
- [ ] Chance of needing a follow up test
- [ ] Chance of dying from ovarian cancer
- [ ] Chance of cancer diagnosis
- [ ] Chance of being unnecessarily diagnosed and treated for a cancer that would never have caused symptoms or death
- [ ] Screening interval
- [ ] Test location
- [ ] Action required by you to arrange the test
- [ ] Who performs the test
- [ ] Waiting time for the test
- [ ] Waiting time for the result

During the last section you were shown different combinations of 16 features related to screening for possible ovarian cancer. You have already shown that some features are more important than others to you but were there any features that you would consider completely unimportant?

Please select any features that would always be an important issue to you if you were making a decision about whether to undergo screening for possible ovarian cancer.

- [ ] Type of test
- [ ] Accuracy: chance of false-negative result
- [ ] Accuracy: chance of a false-positive result
- [ ] Chance of needing a follow up test
- [ ] Chance of dying from ovarian cancer
- [ ] Chance of cancer diagnosis
- [ ] Chance of being unnecessarily diagnosed and treated for a cancer that would never have caused symptoms or death
- [ ] Screening interval
- [ ] Test location
- [ ] Action required by you to arrange the test
- [ ] Who performs the test
- [ ] Waiting time for the test
- [ ] Waiting time for the result

These are all the features you were asked about in the last section:

- Type of test
- Accuracy: chance of **false-negative** result
- Accuracy: chance of a **false-positive** result
- Chance of needing a follow up test
- Chance of dying from ovarian cancer
- Chance of cancer diagnosis
- Chance of being unnecessarily diagnosed and treated for a cancer that would never have caused symptoms or death
- Screening interval
- Test location
- Action required by you to arrange the test
- Who reviews the result
- Waiting time for the test
- Waiting time for the result

Are there any other characteristics that would be important for you to know before deciding whether or not to take a test to screen for ovarian cancer?

Please list any thoughts below:

Section 4: A little more about you

In this section we will ask a little more about you and your views around health care.

In general, how would you rate your overall health?

- ○ Very poor
- ○ Poor
- ○ Fair
- ○ Good
- ○ Very good

To what extent do you worry about your risk of cancer?

- ○ Not at all
- ○ A little
- ○ A moderate amount
- ○ A lot
- ○ A great deal

Have you ever been diagnosed with any type of cancer?

- ○ Yes
- ○ No
- ○ I prefer not to say

To your knowledge, have any of your family or friends ever been diagnosed with ovarian cancer?

- ○ Yes
- ○ No
- ○ I prefer not to say

Was this person your:

- ○ Blood relative
- ○ Non-blood relative (e.g. step-relatives, in-law, adoptive relatives)Non-blood relative
- ○ Friend
- ○ Acquaintance/ work colleague
- ○ Other:

Based on the information below, how often do you visit the GP on average each year?
Please answer "0" to show you are paying attention.

Have you every undergone testing for possible ovarian cancer?

- ○ Yes
- ○ No
- ○ I don't know
- ○ I prefer not to say

Have you ever had a transvaginal ultrasound?

- ○ Yes
- ○ No
- ○ I don't know
- ○ I prefer not to say

A transvaginal ultrasound is a type of pelvic ultrasound used to examine female reproductive organs. This is an internal examination.

Do you regularly attend cervical screening (smear test) appointments?

- ○ I attend every time I receive an invitation
- ○ I attend cervical screening sometimes
- ○ I used to attend regularly but have stopped
- ○ I have never attended

Regularly means whenever you receive an invitation letter in the post.
For a woman your age with normal past screening results, this will be every 3 years.

How much confidence and trust do you have in general practitioners (GPs)?

- ○ A great deal
- ○ A lot
- ○ A moderate amount
- ○ A little
- ○ None at all

When seeking help for medical issues, how much do you feel able to be involved in decisions about the treatment process?

- ○ A great deal
- ○ A lot
- ○ A moderate amount
- ○ A little
- ○ Not at all

When seeking help for medical issues, how much do you wish to be involved in decisions about the treatment process?

- ○ A great deal
- ○ A lot
- ○ A moderate amount
- ○ A little
- ○ Not at all

In general, how willing are you to take risks?

- 1- unwilling to take risks
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10- fully prepared to take risks

Thank you for completing this survey.

Important: Please click here to return to prolific and confirm your submission

If you have any questions or concerns about the survey please contact Rebekah Hall at rh591@exeter.ac.uk or Prof Anne Spencer at A.E.Spencer@exeter.ac.uk

Postal address: University of Exeter Medical School, Room 1.15, South Cloisters, St Luke's Campus, Magdalen Road, City, Exeter, EX1 2LU

More information about ovarian cancer and the tests available can be found at :

Cancer Research UK: https://www.cancerresearchuk.org/about-cancer/ovarian-cancer

Target ovarian cancer: https://www.targetovariancancer.org.uk

The Eve Appeal: https://eveappeal.org.uk/gynaecological-cancers/ovarian-cancer/

Please speak to your GP if you are concerned about your risk of ovarian cancer.

Appendix 9.3: Design specifications when specifying the communication format for the screening DCE based on risky attributes

vii. Absolute risks not relative risks

The use of absolute risks (e.g. 4 in 100 will die) rather than relative risks (e.g. 50% increase in deaths) is consistently recommended across the risk communication literature in order to maintain a neutral communication style and maximise understanding. Relative risks have been shown to manipulate or persuade audiences by magnifying risk perceptions and decreasing understanding (Akl *et al.*, 2011; Garcia-Retamero & Cokely, 2017; Zipkin *et al.*, 2014). Relative risks typically have a greater impact on decision-making, therefore mixing attributes expressed in terms of absolute and relative risks within the same experiment would introduce potential bias (Zipkin *et al.*, 2014).

viii. Frequencies not percentages

There is mixed evidence about the acceptability and effectiveness of percentages versus simple frequencies. Studies comparing the two approaches demonstrate conflicting results (Hoffrage *et al.*, 2002; Trevena *et al.*, 2006; Woloshin & Schwartz, 2011). However, research suggests rare events (i.e. those occurring less than 1%) may be less well understood when represented as percentages due to the use of decimal points meaning simple frequencies are preferable (Trevena *et al.*, 2021).

ix. Consistent denominator for frequencies

To aid understanding and comparisons between alternatives uniformity in the size of the denominator is recommended both across and within attributes (Garcia-Retamero & Cokely, 2017; Garcia-Retamero & Galesic, 2009; Garcia-Retamero *et al.*, 2010). Consistency in the denominator also requires the population of interest to be the same across all attributes (i.e. all those at risk). In practice, this means attributes such as false negatives should be expressed in terms of the number of people screened taking into account disease prevalence rather than in terms of only those with the disease (i.e. approach used in the diagnostic study) (Garcia-Retamero & Cokely, 2017).

x. Consistent framing of attributes

How risks are framed has a proven impact on the decision-making. Positively framed risks (e.g. chance of surviving) typically increase the acceptability of interventions compared to negatively framed risks

(e.g. chance of dying) (Akl *et al.*, 2011; Michalovic *et al.*, 2018; Zipkin *et al.*, 2014). To avoid biases in the willingness to trade between attributes, a consistent framing perspective should be adopted. However, the decision between positive or negative phrasing may introduce a potential bias regarding the selection of the opt-out "no screening" option (Lee & Kang, 2018; Tang & Chooi, 2021).

xi. Visual aids

Evidence on the effectiveness of visual aids for the communication of risk is disputed. Several studies have demonstrated the use of graphics improves understanding and can help to avoid biases such as framing effects and denominator neglect, particularly amongst those with lower numeracy skills (Garcia-Retamero & Cokely, 2017; Garcia-Retamero *et al.*, 2010). However, many studies have also shown visual aids do not influence the understanding or interpretation of risk information. Specifically in the context of DCEs, (Vass *et al.*, 2018a) showed no difference in preferences between risks expressed visually or as percentages.

The best choice of visual aid is also disputed and varies according to the context with studies providing conflicting evidence showing icon arrays, bar charts, decision trees and line graphs are all most preferred by audiences and/or most effective in varying contexts (Corso *et al.*, 2001; Scalia *et al.*, 2021; Zipkin *et al.*, 2014).

Importantly, studies demonstrating that visual aids can have a detrimental effect on risk comprehension where graphics are overly complex or aimed at audiences with poor graph literacy. Regardless of the choice of visual display, transparency and simplicity are fundamental. To maximise the effectiveness of visual aids there are some universal guidelines that aid the interpretability of graphics. Satisfying these criteria has been found to be more important that the choice of specific graphic (Ancker *et al.*, 2006).

Key criteria to improve the effectiveness of visual aids:

- visuals should be supplemented by numerical risks (Garcia-Retamero & Cokely, 2017; Okan *et al.*, 2015; Trevena *et al.*, 2021)
- ensure spatial features of visuals (e.g. height of bars, axis scales) are conventional and representative (e.g. avoid truncated scales) (Trevena *et al.*, 2021)
- Use incremental risk format showing the (risk with and without intervention displayed in same array
- Depict both positive and negative outcomes within the same visual (e.g. stacked bar charts or icon arrays showing outcomes for the entire population)
- assess the graph literacy of the target audience (Garcia-Retamero & Cokely, 2017)
- validate visual aids by conducting usability studies with the target audience before implementation (Okan *et al.*, 2015; Woller-Carter *et al.*, 2012).

One review, based on 19 studies, concluded that the choice of a specific graphic is not as important as whether the graphic frames the frequency of an event with a visual representation of the total population in which it occurs (Ancker *et al.*, 2006).

Appendix 9.4: Interview schedule for the think-aloud pilot interviews

Introduction to purpose of the study and verbal consent (8 mins)

Introduction of interviewers

Verbal consent
(Written consent received prior to interview)
Reminder of the information sheet. Recap of consent form (free to withdraw, interview is recorded for transcription at which point it will be deleted etc.) Ask if they are happy to go ahead and start interview.

Background to study
We are interested in testing the usability of the questionnaire we have developed, it not a test of the user themselves, questions are based on your own opinion and there are no right or wrong answers

(Recording starts here)

Introduction to thinking-aloud (5 mins)

Explain that we're interested in testing the system that we're about to present and that we're not testing the user specifically

Ask the participant to tell us everything they're thinking about from the moment they read the task and when they complete it.

Tell them that they don't need to plan/think out what they want to say. Just act as if they're by themselves Emphasize that the important thing is to keep talking

Explain that if you notice long periods of silence you may interrupt and remind them to keep talking.

You will not be able to respond to any questions during the observation but will happily answer any questions at the end of the task

Warm up exercise (5 mins)

To help the user get a feel for actually performing a Think-Aloud, perform an example of think-aloud e.g. "Please think aloud as you name how many windows are in your house?"

Think-aloud tasks (20 mins)

Introduction to the task: discrete choice experiments and attributes and levels included

Remind participant: most important thing is to keep talking, do not worry about making sense or explaining what you are doing to me. Act as if you are alone and are thinking out loud as you complete tasks.

Does the user have any questions before tasks begin?

During tasks: interviewers will remain silent unless participant stops talking then probes will be used to encourage talking. If the user becomes frustrated, take a quick break

Probes:

- What are you thinking now?
- Why did you choose that option?
- What did you consider when making that choice?
-

Follow up questions (5 mins)

How difficult did you find completing the questions?

Which attributes did you base your decisions on?
Was there anything you did not understand as you were completing the task?
Is there anything you would like to change about the way the questions were presented?
Would you have answered differently if you were alone?

Wrap up discussion (5 mins)

Any further feedback or thoughts?
Any questions?
Thank you for completing the interview

Appendix 9.5: Results of think-aloud interviews used to pilot the screening DCE

Findings from interviews centred around five key themes.

i. Understanding and acceptability of the task

Some participants initially struggled with format of the question and were distracted by the format and spotting what changed between the questions:

" I've got the same question up again, it looks like any way" p3

"They've swapped back the other way… No it's definitely swapped" p2

However, after an initial teething period, participants appeared to grow accustomed to the format of questions and credited the visual layout for aiding with decisions:

"I like the way you've got them in columns so that you could flip from one to the other and easily compare the figures" p3

"it seemed quite clear. I think it's quite a good way of putting it across" p2

And the intention to trade-off between the risks and benefits appeared to be understood and well-considered by all participants:

"False negatives, meh not great. Weighing it up against the top two, which I think are going to be a much bigger deal, the false positives the biggest deal in my head so yeah i'd live with that for the trade off in the false positives". p1

"I can see where you're going with this up trying to get us to weigh up all the factors" p4

The choice context also appeared to be well-integrated into the decision-making process, with most participants acknowledging the frequency of screening and the size of the population screened within the question:

"first of all I'm looking at what we've got annually, so this is going to be every year for 10 years which I think is quite important for me so I have to kind of consider that in what my options are" p1

"So out of 10,000 women. If everybody is screened, every year, for 10 years." p5

Respondents appeared to use a combination of vertical (by alternatives) and horizontal (by attributes) comparisons when considering choice tasks:

"I'm looking at the various figures for test A and test B and just comparing them for each one of the four categories." p3

"Right this one, 10 deaths. 994 false negatives, 16 people missing the result. Three overdiagnosis but you're under the 10% of false positives, its still too many, but it's better. Test B is four tenths again and 30 deaths, I think that's way too high." p2

The majority of participants focused on their own personal decisions, however, one participant took a more altruistic approach and considered the implications of the particular screening programme for a range of people including her children and even the budgetary consequences for the NHS:

"It's so hard, I'm trying to think what would I want my daughter, when she grows up if she was at risk of ovarian cancer I'd be really worried about the false positive results, so I had to keep thinking about that in terms of where do I stand on the spectrum." p1

"That and the false positives, test A is just not worth it for so many women, especially the younger ones with that one, and the fact that you're only getting a 25% reduction is that worth NHS resources?" p1

Respondents appeared to be engaged and well-motivated by the research question, drawing on personal experiences and considered the consequences of the decisions they made. Perhaps due to this investment, participants struggled at times and commented on the difficulty of their deliberations. However, this appeared to be due to the importance of all attributes rather than any difficulty in understanding:

"Oh, this one's a difficult one, to choose between the higher false positives and false negatives" p3

"*This is a bit more tricky*." p2

"Right… you've really thrown some hard ones in" p1

Despite finding it hard to make a choice in some choice tasks, participants completed all twelve questions without appearing to get distracted or experience loss of focus:

"Oh. That was it? Did I go through them too quickly?" p2

ii. Attribute understand and importance

All attributes appeared to be well understood. One participant asked for further clarification on "overdiagnosis" during the introductory reading but appeared to understand the concept well once completing the tasks.

Overdiagnosis appeared to be the most divisive attribute with some finding the concept highly important, whereas others saw it as less of an issue, with one woman stating she had largely ignored this attribute during deliberations:

"Overdiagnosis is 22, I must admit that does worry me a bit the thought of having to have surgery unnecessarily is worrying." p2

"The overdiagnosed on that is fab, that's great. I mean you can rest assured that you've got hardly any chance of having that happen compared to [test A], the comparison of the pair is huge and again, if you then end up diagnosed your treatments going to be so much more appropriate." p1

"if I had cancer, I think I'd rather have everything thrown at it than nothing." p4

"Test A has more overdiagnosis, that's not so much of a concern really as I think I would prefer to be diagnosed regardless." p3

The magnitude of false positives appeared to be barrier for some participants, particularly at the start of the choice tasks:

"They were all really important but the ones that catch your eye I suppose were the false positives, the second one down. Because of the absolutely high numbers, it tends to catch your eye more doesn't it?" p2

"I'm still really, i'm floored by that number, over 10 years, 10,000 women, at some point that's huge, it's just it's almost its getting on for 50% chance of you having a false positive in 10 years, that is just crazy." p1

However, all participants appeared to exhibit trade-offs between two or more of the attributes in each choice tasks, and when asked, no women said they focused on a single attribute alone:

"Weighing up the odds on there, because again that's 10% [false positives] but then its only a quarter of the people that would die without the tests, I think test B again seems to be the best choice." p2

"There's got to be more to it than just looking at the deaths, we've got to be thinking about what are my chances, if I got diagnosed, of not having lots of horrific unnecessary treatments." p5

Overall, women appeared to have a good grasp on the meaning of all the attributes and considered the underlying consequences of each of the benefits and harms rather than solely focusing on the differences in numbers:

"I was very aware of what the consequences were. I mean with false positive you're likely to have surgery, that is unnecessary or follow up tests that are unnecessary. With the false negatives, you may not have the treatment that you need in time to save your life." p3

"Overdiagnosed, zero brilliant. So the chances of going through treatment when it's not necessary, fab!" p1

iii. Interpretation and presentation of risk information

All interview respondents converted the raw figures used to describe levels into percentages, fractions and/or ratios whilst completing choice tasks and were able to do so with minimal mistakes. When asked, four of the five women acknowledged they were comfortable working with numbers but suggested others may experience problems. One woman explained that due media coverage of the COVID-19 pandemic she now felt more comfortable with these kinds of statistics:

"I think people understand a "1 in 4". You know with all this covid that we've had over the last two years and we've had all these quotes thrown at us and its something I'm becoming familiar with." p1

Participant 3 expressed the most difficultly (despite not making any mistakes) and suggested the inclusion of percentages may help:

"I would have liked to have seen the numbers but say in brackets at the side the percentage, and then perhaps with an Asterix saying per 10,000 or compared to the number with no screening. I think that would explain it more rather than having to work it out yourself." p3

"I couldn't do the percentages for the false positives my brain doesn't work that well." p3

Participants all varied in their approach to risk interpretation and often used several approaches as they compared different attributes. For instance, mortality reduction was often interpretated as a risk reduction compared to the no screening option:

"This one is equal on the deaths of 50% of the no screening" p3

"It reduces the number of ovarian cancer deaths by about 50%" p4

"Test A is definitely better than test B on the ovarian cancer deaths, knocking down to 75% less deaths." p1

However, given the remaining attribute levels for the no screening alternative were set to zero, risks for remaining attributes tended to focus on difference between the two test alternatives.

Interesting, the magnitude of false positives appeared to lead people to focus more on the time scale and/or number of people screened:

"Yeah I think both, I think 2000 and 3000 false positive results is too many out of that many people." p4

"I think it would totally put my mind, a lot more at rest about having that test, knowing that I've got such a lower false positive result chance, which is still not, I mean its still over 10 years, a 1 in 10 chance but it's so much better than test A." p1

Whilst comparisons of attributes with smaller levels were typically compared in the raw format with little reference to the wider population:

"Test A seven false negatives and test B 16. None with no screening." p2

"False negatives, are very similar at 10 and 13" p5

"7 and the 11 on overdiagnosis its still quite similar and quite high" p2

iv. No screening alternative

Attitudes towards the no screening alternative varied across participants. One participant appeared to adopt a neutral attitude towards screening from the outset of the task (quote)

However, remaining participants expressed favourable opinions about screening which impacted their willingness to select the opt-out alternative even where expressing displeasure at both options:

"I was very tempted to say I wouldn't be screened but I'm, like I say I would, I feel that we should be screened for it really." p2

“It makes me question. Is it worth a yearly screening test? is it worth that not just for me, but for the NHS and the cost.” p1

“You’ve just totally thrown me because I’m looking at the false positives and thinking yeah but that's over 10 years, a third of women are going to get told its positive to go through all that and then find out they’re not. That’s too great a figure, that's just not what you want, is it worth doing that?” p1

“There doesn't seem to be a huge advantage in taking a test.” p3

“I think.I would go for test B because I think it would be an important thing to be tested but it's still quite worrying the amount of false positives and overdiagnosed ones.” p2

One of these participant even outright stating they had chosen to ignore that option:

“I’m ignoring the no screening because to me it that is a no brainer.” p3

But later acknowledged this reluctance caused them to behave counterintuitively or inconsistently as tasks progressed (contradictory):

“I’ll choose test B, that may seem to be counter to what I’ve already chosen but that, that feels right.” p3

In total, just two of the women selected the “no screening” alternative during the survey but showed reluctance in doing so:

“That's difficult. I’m on one that believes in screening for different thing if it saves lives but it's hard to decide over those.” p2

“I think I would choose no screening there, I know I said before, I think we should have screening and but that's a very high amount of false positives.” p2

v. Missing contextual information

Throughout the survey, participants highlighted additional information they felt essential to their decision-making. The absence of this information, often meant women were required to make assumptions in order to facilitate decision-making.

One participant struggled to interpret mortality reduction without the underlying knowledge of how many people per 10,000 were diagnosed:

“I'd love to know what the actual figures of what the cancer rate is compared to that [deaths]” p1

Once told this information after completing tasks, they indicated their answers would have differed on the basis of this knowledge:

“Once you say that to me, and you said, there are 70 per 10,000. That over half of them will die (without screening) that is, in my mind significant…. yeah there's definite reasons for me to know that that's my risk, it’s not 40 my risk is 60% of dying if it's not detected and put in that context that means a lot more to me.” p1

Two participants both made assumptions about their personal risk of cancer and expressed a desire to know this information before committing to screening. Both stating their decisions were subject to change based on increased knowledge of risk factors:

“I’d want to know what are my risks, me personally, before I even look at this where are my risks is there some sort of checklist or questionnaire?” p5

“I put a caveat on that of being somebody that’s at really high risk that needs to have be part of the screening program was my caveat.” p1

Finally, another participant became confused between screening and diagnostic testing, referring to those experiencing symptoms and leading the participant to become concerned about the poor performance of the tests described by the hypothetical scenarios.

“if you’re waltzing round and think you've got, I don’t know, either IBS or bad case of constipation or things, I believe you get bloated don’t you? And you're walking around thinking that or even that is menopause or something like that.” p2

This highlighted the need for further clarity about the purpose of screening and the importance of the debriefing information provided at the end of the study.

Appendix 9.6: Key socio-demographic characteristics of pilot study respondents

Characteristic	n (%)
Age	
Mean (SD)	52.0 (9.1)
Range	40-74
Ethnicity, n (%)	
White	36 (90)
Other	4 (10)
Prefer not to say	-
Previously tested for ovarian cancer, n (%)	
Yes	5 (13)
No	35 (88)
Don't know	-
Self-reported health, n (%)	
Very good	5 (13)
Good	21 (53)
Fair	10 (25)
Poor	4 (10)
Very poor	-
Prefer not to say	-

Appendix 9.7: Ngene syntax used to generate the Bayesian Efficient experimental design used in the final DCE study

```
design
;alts = alt1*, alt2*, optout
;rows = 12
;eff = (mnl, d, median)
;bdraws = sobol(100)
;model :
U(alt1) = bmortality[(n,-0.03,0.04)] * mortality[10,20,30] + bFP[(n,-
0.001,0.0003)]* FP[1000, 2000, 3000, 4000] + bFN[(n,-0.12,0.06)]*
FN[3,7,10,13,16,20] + bOD[(n,-0.060,0.037)]* OD[0,3,7,10,13,16]/
U(alt2)= bmortality          * mortality                  + bFP        * FP
+ bFN        * FN                + bOD      * OD/
U(optout)= b3[(n,-3.77,2.43)]
$
```

Appendix 9.8: Final version of the screening DCE survey instrument

Women's preferences towards ovarian cancer screening

INFORMATION SHEET FOR PARTICIPANTS

VERSION NUMBER [4]: DATE [02/12/21]

Thank you for showing an interest in this survey. Please take time to read the following information carefully before deciding whether or not to take part.

What is the aim of the project?
Ovarian cancer is the 6th most common cancer in women in the UK. A screening test for ovarian cancer (similar to cervical screening or 'smear test') could help to identify the disease earlier and improve survival outcomes. Research to develop a suitable screening test is ongoing but it important to make sure any potential test is acceptable to patients and the public. In this study we would like to understand the attitudes and preferences of womena and people with ovaries towards a hypothetical test for ovarian cancer. In particular, we want to learn what characteristics of diagnostic testing are most important.

Why have I been invited to take part?
We are approaching you because we are seeking responses from women and people with ovaries over the age of 40. You do not need to have any prior knowledge of ovarian cancer and you do not have to have been previously tested for ovarian cancer to take part in the study. You must be able to complete the survey in English to take part. **Please do not take part in this survey if you have ever undergone a procedure to remove both of your ovaries.**

What will I be asked to do?
Should you agree to take part, you will be asked to complete a survey lasting around 15 minutes. During the survey you will be shown descriptions of two different medical tests and asked which test you would prefer to have. In total, we will ask you about 6 pairs of tests. At the end of the survey we will ask you some additional questions about yourself. This will help us to better understand how attitudes might vary from person to person.

What are the possible disadvantages and risks of taking part?
Participating in the research is not anticipated to cause you any disadvantages or discomfort. Some questions may be considered sensitive, however, you do not have to provide responses to any questions you feel uncomfortable answering. During the survey you will be asked to imagine you might have cancer, for some this may cause anxiety.

Will I be paid?
You will receive a payment of £2.00 for completing the survey. Payments will be paid directly into your Prolific account within 10 working days.

Please note: In order to receive the payment you must complete the full survey and click the completion link at the end of the survey. To ensure responses are high quality, an “attention

check" question has been included within the survey. You must correctly complete this question to receive payment.

Can I change my mind and withdraw from the project?
If you decide you no longer wish to take part during the survey, simply exit the webpage to withdraw. Your incomplete responses will be permanently deleted. If you decide to withdraw after submitting your responses, please contact us via your Prolific account or directly by email. You can withdraw from the study for up to 14 days after completion and do no not have to give a reason. After 14 days it may no longer be possible to withdraw your submission because anonymization will mean we cannot link responses to individual participants.

Is the survey confidential?
All your answers to the survey are completely confidential and anonymous. You will not be asked your name or any other identifying information. Your responses will be securely stored on an encrypted password protected computer and managed according to a law called the Data Protection Act (2018). Your anonymised data will be stored for a period of five years.

In line with the Cancer Research UK data sharing guidelines, your data may be shared with other researchers in the future at our discretion. Any shared data will be fully anonymised. For more information: https://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/submission-of-a-data-sharing-and-preservation-strategy/data-sharing-guidelines

The results of the study may be published in academic journals or conferences but any included data will not be individually identifiable.

The University of Exeter processes personal data for the purposes of carrying out research in the public interest. The University will endeavour to be transparent about its processing of your personal data and this information sheet should provide a clear explanation of this. If you do have any queries about the University's processing of your personal data that cannot be resolved by the research team, further information may be obtained from the University's Data Protection Officer by emailing dataprotection@exeter.ac.uk or at www.exeter.ac.uk/dataprotection

What if I have any questions?
If you have any questions about our project, either now or in the future, please feel free to contact Rebekah Hall by emailing rh591@exeter.ac.uk

Complaints
If you have any complaints about the way in which this study has been carried out please contact the Chair of the College of Medicine and Health Research Ethics Committee:-

Mark Tarrant, PhD
Chair of the CMH Research Ethics Committee
Email: cmhethics@exeter.ac.uk

This project has been reviewed and approved by the University of Exeter College of Medicine and Health Research Ethics Committee (REF NUMBER: 20/09/261)

Consent Form

- I understand that my participation is voluntary and that I am free to withdraw for up to 14 days without giving any reason and without my legal rights being affected.

- I understand that my data from the study will be fully anonymised and will be looked at by members of the research team and may potentially be shared with other researchers in future if appropriate.

- I understand that relevant sections of the data collected during the study may be looked at by individuals from the University of Exeter, Cancer Research UK or regulatory authorities for audit purposes

- I understand that the results of the study may be published in academic journals but my anonymity will be preserved

- I understand that my anonymised data will be securely stored on an encrypted password protected computed for a period of five years.

- I understand that in order to receive payment for this survey I must complete the full survey and click the link at the end of the survey. I must also correctly complete an attention check question randomly placed within the survey.

I confirm that I have read the information above and agree to take part in the study:

Yes

No

Have you ever had a medical procedure that involved the removal of both of your ovaries?

○ Yes ○ No

Please enter your ProlificID

Next

Thank you for agreeing to take this survey

The survey will present some information about ovarian cancer and describe some tests. We will then ask you some questions about your health and about cancer testing. Later we will ask you to consider different tests for ovarian cancer.

Let's start with some information about ovarian cancer...

Next

Introduction to ovarian cancer screening

Ovarian cancer occurs when the cells in and around the ovaries and fallopian tubes become abnormal, grow out of control and form a lump called a “tumour”.

Ovarian cancer is the 6th most common cancer for women in the UK. Over 7,000 women are diagnosed annually. Most of these cases occur in women over the age of 40.

Screening tests can help to identify certain types of cancer earlier, **before any symptoms arise**. Earlier diagnosis means more treatment options are available and can help to improve the chance of being cured or living longer.

In the UK screening tests are currently available for breast (mammogram), cervical ("smear test") and colorectal cancer.

There is currently no recommended screening test for ovarian cancer, however, research is ongoing and it is hoped a suitable test will be developed and approved in the future.

To be approved any screening test must be proven to save lives, however, all medical tests are also involve some risks (such as incorrect results or side effects). This means it is important to make sure any potential test meets the expectations of people who may be invited to have them and the balance of benefits and potential harms is acceptable to patients and the public.

In this survey we would like to find out the most important aspects of testing are most important to people with ovaries and how people balance the potential benefits (e.g. improved chance of survival) against potential harms.

Next

Ovarian cancer risk factors

Deciding whether to undergo screening is a personal decision and everyone is different.

Some people may find it helpful to have a better understanding of their risk of developing ovarian cancer before making a decision.

Without screening approximately 65 in 10,000 people will develop ovarian cancer over a 10-year period.

There is no way to know for sure who will develop ovarian cancer, however, there are some factors that increase or decrease the personal risk.

Factors that increase the risk of ovarian cancer:

- Family history of ovarian cancer
- Getting older
- Previous cancer diagnosis (especially if you were diagnosed before the age of 40)
- Using hormone replacement therapy (HRT)
- Smoking
- Obesity
- Certain medical conditions (e.g. diabetes, endometriosis)

Factors that may decrease the risk of ovarian cancer:

- Taking the combined contraceptive pill at some point in your life
- Having children and/or breastfeeding
- Having a hysterectomy or sterilisation ("tubes tied")

If you would like more information about risks of ovarian cancer please visit the Cancer Research website: https://www.cancerresearchuk.org/about-cancer/ovarian-cancer/risks-causes

Next

Part 1: Ovarian cancer knowledge

We would like to learn more about your knowledge and experience of ovarian cancer.

Have you ever undergone testing for possible ovarian cancer?

- ○ Yes
- ○ No
- ○ I don't know
- ○ Prefer not to say

Which of the following do you recognise as a symptom of ovarian cancer?

Check all that apply

- [ ] Feeling constantly bloated
- [ ] A swollen tummy
- [ ] Discomfort in your tummy
- [ ] Persistent indigestion or feeling sick
- [ ] Discomfort in your pelvic area
- [ ] A change in bowel habits
- [ ] Back pain
- [ ] Pain during sex
- [ ] Feeling full quickly or loss of appetite
- [ ] Feeling tired all the time
- [ ] Unintentional weight loss
- [ ] Needing to pee more often or more urgently than usual
- [ ] None

How confident are you that you would notice a symptom of ovarian cancer?

When was the last time you visited your GP?

This is an attention check question. Pease enter 'yes' to show that you are paying attention

Next

Part 2: Preferences towards ovarian cancer screening

During this section of the survey you will be asked to choose between screening tests which differ in terms of 4 characteristics:

1. Ovarian cancer deaths

2. False-positive results

3. False-negative results

4. Overdiagnosed cancers

These 4 characteristics are described in more detail on the next pages.

The rate at which these benefits and harms occur is described based on 10,000 people undergoing yearly screening over a period of 10-years. This will hopefully make it easier for you to compare the benefits and harms against each other.

In total, approximately 9 million people in England and Wales would be eligible for this hypothetical screening test (people with ovaries, aged 50-75 years old).

Next

1. Ovarian cancer deaths

This is the number of people who will die from ovarian cancer.

In this study, having no screening will lead to 40 deaths per 10,000 women over 50 years old.

For any screening test to be approved there must be strong evidence that the test reduces the number of deaths from ovarian cancer compared to no screening.

The screening tests you will be shown could reduce ovarian cancer deaths to:

- 30 deaths per 10,000 women over 50 years old
- 20 deaths per 10,000 women over 50 years old
- 10 deaths per 10.000 women over 50 years old

Next

2. False-positive results

These are people who do not have cancer but receive a positive (or abnormal) result.

People who receive an incorrect possible result will undergo unnecessary, often invasive testing.

A small proportion of these people (about 3%) will undergo unnecessary surgery because of the incorrect result.

Choosing not to be screened means there is no risk of false-positive results.

Over a 10-year period, the screening tests you will be shown in this study may result in:

- 1000 false-positive results per 10,000 women screened, leading to 30 unnecessary surgeries
- 2000 false-positive results per 10,000 women screened, leading to 60 unnecessary surgeries
- 3000 false-positive results per 10,000 women screened, leading to 90 unnecessary surgeries
- 4000 false-positive results per 10,000 women screened, leading to 120 unnecessary surgeries

Next

3. False-negative results

These are people who have cancer but receive a negative (or normal) result.

An incorrect negative result leads to false reassurance that they are disease-free and will mean diagnosis and treatment will be delayed.

Choosing not to be screened means there is no risk of false-negative results.

Over a 10-year period, the screening tests you will be shown in this study may result in:

- 3 false-negative results per 10,000 women screened
- 7 false-negative results per 10,000 women screened
- 10 false-negative results per 10,000 women screened
- 13 false-negative results per 10,000 women screened
- 16 false-negative results per 10,000 women screened
- 20 false-negative results per 10,000 women screened

Next

4. Overdiagnosed cancers

These are people who have cancer and are correctly diagnosed using the test. However, the cancer would never lead to death and may even never cause any symptoms.

These people will undergo unnecessary treatment.

Treatments for cancer (e.g. chemotherapy, radiation, surgery) have serious side-effects and often have long-term physical, mental and sometimes financial consequences (due to lost work).

Doctors are unable to tell which patients have a life-threatening disease and who has been over-diagnosed so everyone is offered treatment.

Choosing not to be screened means there is no risk of being over-diagnosed with ovarian cancer.

Over a 10-year period, the screening tests you will be shown in this study may result in:

- 0 cases of over-diagnosed cancer per 10,000 women screened
- 3 cases of over-diagnosed cancer per 10,000 women screened
- 7 cases of over-diagnosed cancer per 10,000 women screened
- 11 cases of over-diagnosed cancer per 10,000 women screened
- 16 cases of over-diagnosed cancer per 10,000 women screened

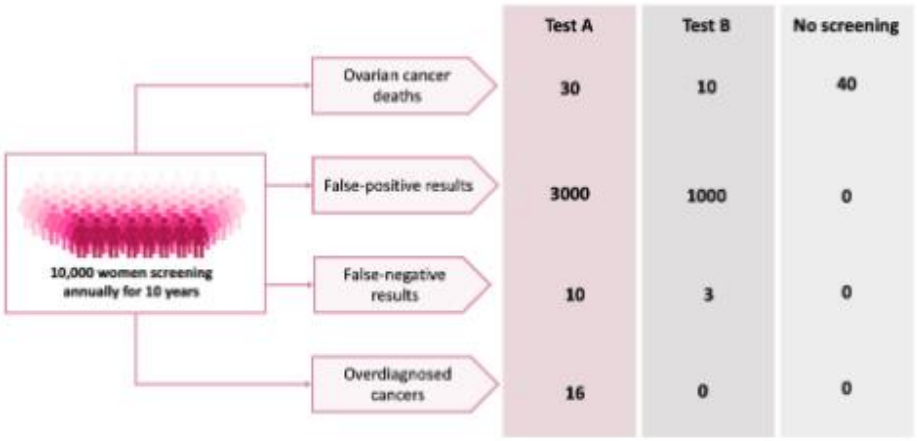
Next

Shortly you will be asked to complete 12 questions comparing test options.

Each question will have 3 options to choose from: Test A, Test B or no screening test.

Each option is described by the 4 characteristics previously described.

The questions will look like this:



	Test A	Test B	No screening
Ovarian cancer deaths	30	10	40
False-positive results	3000	1000	0
False-negative results	10	3	0
Overdiagnosed cancers	16	0	0

For each question, please look at each option and decide which option you would choose if you were asked to make this decision in real life.

These questions are designed to reflect real medical decisions which can be difficult to consider.

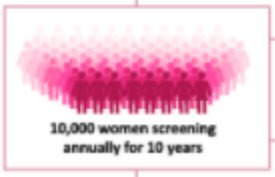
There are no right or wrong answers. We are interested in your opinion.

Next

Warm up question

To help you become familiar with the format of the questions, please answer the warm up question below

If you were given the choice between the options below, which would you choose?



	Test A	Test B	No screening
Ovarian cancer deaths	20	10	40
False-positive results	2000	1000	0
False-negative results	10	3	0
Overdiagnosed cancers	7	0	0

- ○ **Test A**
- ○ **Test B**
- ○ **I would choose not to be screened**

Use your mouse to hover over the characteristics below if you would like a reminder of the characteristic definitions

Ovarian cancer deaths

False-positive results

False-negative results

Overdiagnosed cancers

Next

Great! You will not be asked twelve more questions comparing different test options using the same question style.

Remeber there are no right or wrong answers.

Next

12 DCE QUESTIONS HERE

In this next section you will be asked some follow-up questions about your decision process in the last section. Again, there are no right or wrong answers.

How easy or difficult did you find making your choices?

- ○ Very easy
- ○ Easy
- ○ Neither easy or difficult
- ○ Difficult
- ○ Very difficult

Which characteristics did you consider when making your choices?.

- ☐ Ovarian cancer deaths
- ☐ False-positive results
- ☐ False-negative results
- ☐ Overdiagnosed cancers

What characteristics did you ignore when making your choices?

- ☐ Ovarian cancer deaths
- ☐ False-positive results
- ☐ False-negative results
- ☐ Overdiagnosed cancers
- ☐ I did not ignore any of the characteristics

Why did you ignore certain characteristics?

- [x] There were too many characteristics to look at
- [ ] The other characteristics were unclear
- [ ] The other characteristics were not important to me
- [ ] Other:

You indicated that you would choose not to be screening in any of the scenarios. Please could you briefly explain this decision?

Please rank the four characteristics from most to least important:

Your choices	Your ranking
Overdiagnosed cancers	
Ovarian cancer deaths	
False-negative results	
False-positive results	

Double-click or drag-and-drop items in the left list to move them to the right - your highest ranking item should be on the top right, moving through to your lowest ranking item.

Next

Part 3: Background questions

In the last part of the survey we want to learn a bit more about you and your background. Your answers will be used to understand how preferences towards ovarian cancer testing might vary between different people.

What is your age?

What is your ethnicity?

- ○ Prefer not to say
- ○ White/Caucasian
- ○ Mixed- white and black African
- ○ Mixed- white and black Caribbean
- ○ Mixed- White and Asian
- ○ Asian- Indian
- ○ Asian- Bangladeshi
- ○ Asian- Chinese
- ○ Black-African
- ○ Black- Caribbean
- ○ Arab
- ○ Other:

Which of the following best describes your current relationship status?

- ○ Single
- ○ In a relationship
- ○ Married/In a registered civil partnership
- ○ Separated/divorced
- ○ Widowed
- ○ Prefer not to say

How many children do you have?

Which of the following categories best describes your employment status?

Please choose...

What is the total annual income of your household (before tax)?

- ○ Prefer not to say
- ○ £0-£9,999
- ○ £10,000-£19,999
- ○ £20,000-£29,999
- ○ £30,000-£39,999
- ○ £40,000-£49,999
- ○ £50,000-£59,999
- ○ £60,000-£69,999
- ○ £70,000 or more

What is the highest level of education you have completed?

Please choose...

Next

How is your health in general?

- ○ Very good
- ○ Good
- ○ Fair
- ○ Bad
- ○ Very bad
- ○ Prefer not to say

Compared to the average woman of your age, how would you describe your risk of developing ovarian cancer?

- ○ Very high risk
- ○ High risk
- ○ Average risk
- ○ Low risk
- ○ Very low risk
- ○ Prefer not to say
- ○ Don't know

To what extent do you worry about your risk of ovarian cancer?

- ○ Not at all
- ○ A little
- ○ A moderate amount
- ○ A lot
- ○ A great deal
- ○ Prefer not to say

On average, how often do you visit your GP every year? Please enter 'yes' to show that you are paying attention

Have you ever been diagnosed with any type of cancer?

- ○ Yes
- ○ No
- ○ Prefer not to say

To your knowledge, have any of your family or friends been diagnosed with ovarian cancer?

- ○ Yes
- ○ No
- ○ Prefer not to say

Was this person your:

Check all that apply

- [ ] Blood relative
- [ ] Non-blood relative
- [ ] Friend
- [ ] Acquaintance/work colleague
- [ ] Prefer not to say
- [ ] Other:

Do you regularly attend cervical screening (smear test) appointments?

- ( ) I attend every time I receive an invitation
- ( ) I attend cervical screening sometimes
- ( ) I used to attend regularly but have stopped
- ( ) I have never attended

Regularly means whenever you receive an invitation letter in the post. This is every 3 years if you are under 45 years old or every 5 years if you are under 64 years old.

In general, how willing are you to take risks?

Next

Part 4: Background questions

In this final section you will be asked some questions to understand how comfortable you are with probabilities.

These are not attention check questions. You will still receive payment even if you answer incorrectly.

The PSA (prostate specific antigen) is a blood test that looks for prostate cancer. The test has false alarms so about 30% of men who have an abnormal test turn out not to have prostate cancer. John had an abnormal test. What is the chance that John has prostate cancer?

- ○ 0%
- ○ 30%
- ○ 70%
- ○ 100%

James starts a new blood pressure medicine. The chance of a serious side effect is 0.5%. If 1000 people take this medicine, about how many would be expected to have a serious side effect?

- ○ 1 person
- ○ 5 people
- ○ 50 people
- ○ 500 people

A medical study will randomly assign people so that people are equally likely to get medicine A or medicine B. If there are 300 people in the study, about how many are expected to get medicine A?

- ○ 100 people
- ○ 150 people
- ○ 200 people
- ○ 250 people
- ○ I am not sure

Natasha started a new medicine and was given a handout showing the chance that side effects will occur. Which side effect is Natasha least likely to get?

- a. Dizziness: 1 in 5 people
- b. Nausea: 1 in 10 people
- c. Stomach pain: 1 in 100 people
- d. Allergic reaction: 1 in 200 people
- I am not sure

Amanda is told she has a 1 in 256 chance of dying from cancer and a 1 in 407 chance of dying from a stroke. 6. Which is bigger, Amanda chance of dying from a stroke or cancer?

- Stroke
- Cancer
- Chances are the same
- I am not sure

Next

If you have any additional comments about any of the questions or comments about the survey you have just completed, please leave them below:

Submit

Thank you for completing this survey.

Important: Please click here to return to prolific and confirm your submission

---

Your responses to this survey will add to a body of research which we hope to understand public priorities around a potential future test for ovarian cancer. Currently screening for ovarian cancer is not reccomended since evidence from large trials suggests there is no benefit to screening for ovarian cancer. However, research is ongoing and promising new developments have been made in recent years. This information from this survey will be useful in tailoring ongoing research and policy decisions in the development of screening tests.

We know that cancer is a sensitive subject and being asked to imagine the scenarios we have shown during the survey may have cause some people to feel anxious or concerned. Please speak to your GP if you are concerned about your risk of ovarian cancer.

As with most cancers, early recognition of symptoms will help increase the chances of sucessful treatment. Being of aware of the symptoms will help you to spot them more easily.

Common symptoms of ovarian cancer include:

- feeling constantly bloated
- a swollen tummy
- discomfort in your tummy or pelvic area
- feeling full quickly when eating
- needing to pee more often than usual

Please contact your GP if you have any of these symptoms and do not go away. More information on ovarian cancer can be found on the NHS website: https://www.nhs.uk/be-clear-on-cancer/symptoms/ovarian-cancer

More information about ovarian cancer and the tests available can be found at :

Cancer Research UK: https://www.cancerresearchuk.org/about-cancer/ovarian-cancer

Target ovarian cancer: https://www.targetovariancancer.org.uk

The Eve Appeal: https://eveappeal.org.uk/gynaecological-cancers/ovarian-cancer/

If you are currently living with ovarian cancer and have been affected by any of the issues in this study, please speak to your clinical nurse specialist.

If you have any questions or concerns about the survey please contact Rebekah Hall at rh591@exeter.ac.uk or Prof Anne Spencer at A.E.Spencer@exeter.ac.uk

Postal address: University of Exeter Medical School, Room 1.15, South Cloisters, St Luke's Campus, Magdalen Road, City, Exeter, EX1 2LU

If you would like to be kept informed of the results from this study, please contact Rebekah Hall via your prolific account.

Appendix 10.1: Health-related characteristics of respondents completing the ovarian cancer screening DCE survey

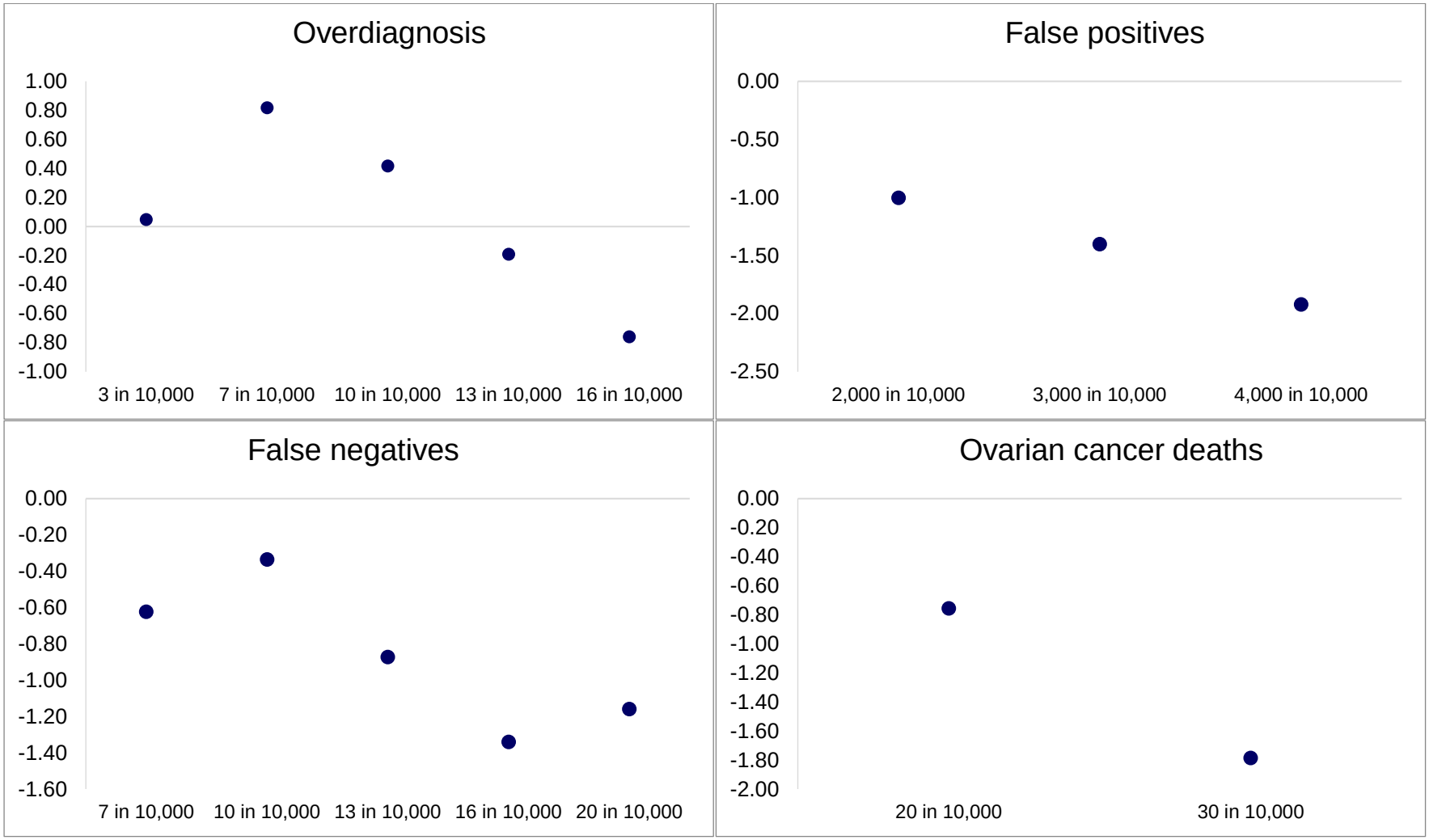
Characteristic	
Perceived risk of ovarian cancer, n (%)	
Very low	12 (5%)
Low	45 (18%)
Average	149 (60%)
High	22 (9%)
Very high	1 (0.4%)
Don't know	21 (8%)
Ovarian cancer-related worry, n (%)	
A great deal	6 (2%)
A lot	13 (5%)
A moderate amount	46 (18%)
A little	108 (44%)
Not at all	76 (30%)
Confidence to recognise OC symptoms, n (%)	
1-Not at all	74 (30%)
2	118 (47%)
3	36 (14%)
4	20 (8%)
5-Extremely confident	2 (1%)
Symptom recognition, n (%)	
Feeling constantly bloated	149 (60%)
Swollen tummy	143 (57%)
Discomfort in your tummy	136 (54%)
Persistent indigestion or feeling sick	55 (22%)
Discomfort in your pelvic area	169 (68%)
A change in bowel habits	77 (31%)
Back pain	105 (42%)
Pain during sex	103 (41%)
Feeling full quick or loss of appetite	76 (30%)
Feeling tired all the time	118 (47%)
Unintentional weight loss	151 (60%)
Needing to urinate more often or more urgently that usual	85 (34%)
None	29 (12%)
Personal history of cancer, n (%)	22 (9%)
Knew someone who was diagnosed with ovarian cancer, n (%)	41 (16%)
Previously tested for ovarian cancer, n (%)	26 (10%)
Cervical cancer screening attendance, n (%)	
Attends every time	157 (63%)
Attends sometimes	35 (14%)
Used to attend but stopped	47 (19%)
Never attended	11 (4%)
Breast cancer screening attendance, n (%)	
Attends every time	52 (21%)
Attends sometimes	3 (1%)
Used to attend but stopped	13 (5%)
Never attended	13 (5%)
Not eligible	169 (68%)
Self-reported overall health, n (%)	
Very good	34 (14%)
Good	124 (50%)
Fair	77 (31%)
Poor	14 (6%)
Very poor	1 (0.4%)

Appendix 10.2: Multinomial logit model used to assess the functional form of attributes

	Coeff.	[95% confidence interval]	
Ovarian cancer deaths			
10 per 10,000	Ref	-	-
20 per 10,000	-0.76***	-1.00	-0.51
30 per 10,000	-1.79***	-2.08	-1.49
False negatives			
3 per 10,000	Ref	-	-
7 per 10,000	-0.62***	-0.93	-0.32
10 per 10,000	-0.34**	-0.65	-0.02
13 per 10,000	-0.87***	-1.16	-0.59
16 per 10,000	-1.34***	-1.72	-0.96
20 per 10,000	-1.16***	-1.64	-0.68
False positives			
1000 per 10,000	Ref	-	-
2000 per 10,000	-1.00***	-1.26	-0.74
3000 per 10,000	-1.40***	-1.69	-1.11
4000 per 10,000	-1.92***	-2.42	-1.42
Overdiagnosed cancers			
0 per 10,000	Ref	-	-
3 per 10,000	0.05	-0.21	0.31
7 per 10,000	0.82***	0.31	1.33
10 per 10,000	0.42	-0.10	0.94
13 per 10,000	-0.19	-0.42	0.03
16 per 10,000	-0.76***	-1.04	-0.47
Alternative-specific constants			
Neither test	-2.54***	-3.02	-2.07
ASC_B	-0.11***	-0.20	-0.01
Model fit statistics			
LL	-2837.09		
Observations	9,000		
N	250		

Key: ***significant at 99% confidence level; **significant at 95% confidence level; *significant at 90% confidence level

Appendix 10.3: Attribute coefficient plots used to visually assess the functional form of attributes



Appendix 10.4: Mixed logit model excluding respondents who failed the rationality check (n=12)

	Coeff.	95% confidence interval	SD
Ovarian cancer deaths	-0.14***	-0.12 – [-0.16]	0.10***
False negatives	-0.06***	-0.04 – [-0.07]	0.05***
False positives	-0.00***	-0.00 – [-0.00]	0.00***
Overdiagnosed cancers	-0.06***	-0.05 – [-0.07]	0.04***
Neither test	-2.33***	-1.57 – [-3.02]	5.31***
Model fit statistics			
LL	-1803.12		
Observations	8,568		
N	238		

Key: ***significant at 99% confidence level; **significant at 95% confidence level; *significant at 90% confidence level

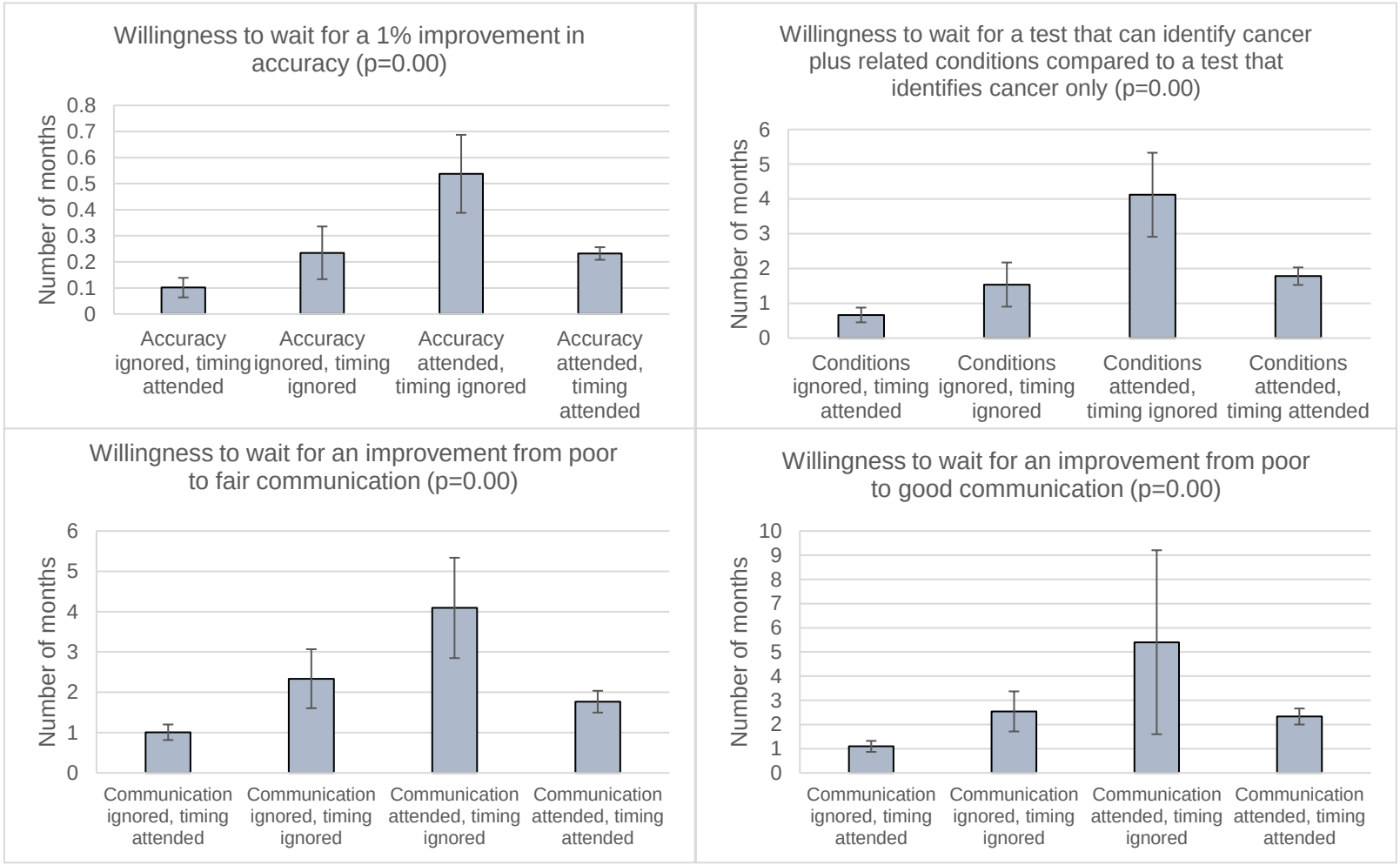
Appendix 10.5: Model fit statistics for alternative latent class models with 2-7 classes.

A 5-class model performed best in terms of AIC and CAIC whereas a 7-class model performed best in terms of log-likelihood and BIC. A five-class model was ultimately chosen based on the class membership percentages and sample size.

Number of classes	AIC	BIC	CAIC	LL
2	4312.24	4358.02	4371.02	-2143.12
3	4081.14	4151.57	4171.57	-2020.57
4	3905.66	4000.74	4027.74	-1925.83
5	3872.25	3991.98	4025.98	-1902.12
6	3857.57	4001.95	4042.95	-1887.78
7	3828.54	2997.57	4045.57	-1866.27

Appendix 11.1: Diagnostic DCE willingness to accept estimates adjusted for different combinations of attribute non-attendance

P-value from ANOVA analysis comparing WTA estimates based on different attendance patterns are shown in brackets.



Appendix 11.2: Screening DCE: WTA estimates based on difference attribute attendance patterns

Shown WTA estimates assume full attendance of the mortality attribute. This represents the majority of participants in all instances. However, all possible attendance patterns were calculated and analysed for differences.

T-tests found no significant differences in WTA estimates for any attribute according to question frame for any attribute attendance pattern.

ANOVA tests were used to identify differences in WTA estimates based on attribute attendance patterns within question frames. Significant differences were found in all instances with exception of estimates of the willingness to accept overdiagnosed cancers in exchange for a reduction in ovarian cancer mortality (p=0.13).

Estimates show the WTA harms in exchange for a 1-person reduction in ovarian cancer mortality per 10,000 people screened.