Attributes used for cancer screening discrete choice experiments: A systematic review

Rebekah Hall[a], Antonieta Medina-Lara[a], Willie Hamilton[a], Anne E Spencer[a]

[a] College of Medicine and Health, University of Exeter, Exeter, UK

Corresponding Author:

Rebekah Hall

Address: South Cloisters, University of Exeter, St Luke's Campus, Heavitree, Exeter, EX1 2LU, UK

Phone: +44 (0)1392 725565

Email: rh591@exeter.ac.uk

ORCID ID:

Rebekah Hall: 0000-0003-1804-993X

Antonieta Medina-Lara: 0000-0001-7325-8246

Willie Hamilton: 0000-0003-1611-1373

Anne Spencer: 0000-0002-8163-3103

Abstract

Background: Evidence from discrete choice experiments (DCEs) can be used to enrich understanding of preferences, inform the (re)design of screening programs and/or improve communication within public campaigns about the benefits and harms of screening. However, reviews of screening DCEs 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 DCE. Misspecification or misinterpretation of attributes may lead to non-compensatory behaviours, attribute non-attendance and responses that lack external validity.

Objectives: 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 DCEs estimating preferences towards cancer screening, dated between 1990 and December 2020. Data were synthesised narratively. In-depth analysis of attributes lead 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 (ISPOR) 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 DCEs 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 DCEs.

Declarations

Funding: This research arises from the CanTest Collaborative, which is funded by Cancer Research UK (reference number: C8640/A23385), of which Rebekah Hall is a funded PhD student, Willie Hamilton is director and Anne E Spencer is a senior faculty member.

Conflicts of interest: No conflicts of interest

Author contributions: All authors contributed to the study conception and design. Searches, screening and data abstraction were performed by Rebekah Hall and Antonieta Medina-Lara. The first draft of the manuscript was written by Rebekah Hall and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank Jenny Lowe, University of Exeter for helping to run the database searches that formed part of this review. We would also like to thank Nia Morrish for assisting in the initial screening of search results.

Key points for decision makers

- Discrete choice experiments assessing preferences towards cancer screening are increasingly popular, however, discrepancies between stated choices and observed screening behaviour exist.
- Test-specific and service delivery attributes were included most often across studies and test sensitivity, mortality reduction and time since last test were most likely to be most important to respondents.

- Methodological choices during the design stage of preference-based studies should be made carefully and reported rigorously since decisions relating to attribute development can influence study results

1. Introduction

Screening for breast, cervical and colorectal cancer increases the likelihood of early detection, leading to improved treatment options and a better prognosis: national screening programmes are an effective way to tackle these cancers [1, 2]. However, the effectiveness of universal screening programmes in maximising health gains and reducing mortality relies on the willingness of individuals to participate. Screening rates remain internationally low, inconsistently meeting participation targets [3, 4]. They are even declining in some instances [5]. Uptake is affected by sociodemographic factors such as age, gender, ethnicity and income [6, 7] but it is also possible that low levels of engagement are influenced by disparities between current screening policies and the preferences of the target population.

The role of preferences in healthcare decision-making has gained attention in recent years, with shared-decision making central to the NHS in the UK [8]. Preferences are particularly important where there are multiple effective treatment options available, each with different and competing risks and benefits [9]. This is relevant to many cancer screening services, particularly colorectal cancer, where at least five different modalities have been shown to be cost-effective [10]. Discrete choice experiments (DCEs) have become an increasingly popular method to measure preferences that are otherwise unobservable in everyday life [11]. DCEs are particularly useful with their ability to assess trade-offs between health outcomes (e.g. quality of life, mortality reduction) and other factors associated with interventions such as test-specific characteristics (e.g. specificity, modality) and broader aspects of service delivery (e.g. gender of the health care provider, results delivery method) [12]. Consideration of attributes beyond immediate health outcomes is particularly relevant to cancer screening, which relies

on active engagement from asymptomatic individuals that may offer no immediate benefit (i.e. for many, tests will be normal) but have associated financial, physical or time constraints.

DCEs require respondents to state their choice between a series of hypothetical alternatives consisting of different combinations of attributes which describe key characteristics of, in this example, cancer screening programmes. Respondents' choices are used to infer the value, expressed as utility scores, associated with each attribute, and the willingness to make trade-offs between the attributes incorporated in the experiment [13]. The identification and selection of attributes and associated levels is a fundamental component of designing a DCE; inadequate selection can lead to omitted variable bias, whereby stated choices do not reflect real life choices. Current guidance emphasises the importance of the attribute selection process and the need to strike a balance between the interests of the target population and relevance to the research or policy question of interest [14, 15]. However, within the literature, performing and reporting of attribute selection is often relatively neglected [16].

Existing systematic reviews largely focus on the clinical implications of DCE findings and suggestions on how to implement findings in policy decisions [17-21]. In contrast, methodological issues of attribute identification, selection, construction (e.g. assignment of levels, type and complexity of attributes) and how these impact on cognitive burden, have not been explored in detail. Previous reviews also differ considerably in terms of the scope of their search strategies and the application of quality assessment, due to narrow review aims, for instance in terms of search dates or cancer type, with three reviews focusing exclusively on studies relating to colorectal cancer [17, 18, 20]. Existing reviews synthesise studies published prior to 2015; given the growth in DCE publications in recent years [11], this review provides an update using a rigorous search strategy covering seven databases using the expertise of an information specialist. In particular, this review aims to provide a summary of the current methods used to identify and select attributes relating to cancer screening, alongside an in-depth review of included attributes and methodological challenges, serving as

a resource for researchers undertaking future studies on cancer screening to design future preference-based studies on cancer testing.

2. Methods

A systematic review was performed in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) Statement [22]. The protocol was registered with PROSPERO before the data extraction stage (CRD42019153834).

2.1 Inclusion and Exclusion criteria

Studies were included if they reported in English a DCE regarding the preferences for screening of any type of cancer within an asymptomatic population. Studies of symptomatic individuals were classified as diagnostic testing and were excluded. Also excluded were studies using alternative preference-elicitation methods such as ranking/rating or time trade-off, as were review papers, commentaries, letters, abstracts or papers that performed secondary analyses on previously reported, and already identified, DCE studies.

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 used seven relevant databases; MEDLINE, Embase, PsycINFO, HMIC, Web of Science, EconLit and NHS EED using previously validated search terms relating to 'discrete choice experiments' [13]. 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'. 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 the Electronic Supplementary Material 1 (ESM). To avoid excluding relevant studies, searches were not further narrowed using terms related to 'screening'; instead this was done manually during the screening stage. The references and forward citations for studies included at full-text stage were also searched for additional studies.

2.3 Study identification

Study selection was performed by two independent reviewers (RH, NM/AML) 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 a third researcher when necessary.

2.4 Data Extraction

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

2.4.1 In-depth attribute analysis

First, data relating to the number of attributes and levels and selection methods were extracted. Next, attributes from each study were extracted and classified. These can be classified in a number of ways; firstly, under the extra-welfarist approach [23, 24] attributes may be categorised as relating to ‘outcomes’ or ‘process’, an approach taken in previous reviews [25]. Alternatively, attributes may be seen as relating to ‘health outcomes’ and ‘non-health outcomes’ [26]. Given the multifaceted nature of cancer screening, and the likely importance and frequency of attributes beyond outcomes, neither approach perfectly met the purposes of our analysis. Instead, we created our own classification system consisting of four categories: (1) outcome attributes (e.g. mortality reduction and adverse events); (2) test-specific attributes (e.g. test procedure, sensitivity and preparation); (3) service-delivery attributes (e.g. screening interval, location and staffing); (4) monetary attributes (e.g. costs and remuneration). Information relating to the complexity (in particular, the inclusion of a risk and/or time elements), significance and influence of attributes was also extracted.

Finally, the most important attribute(s) from each study was extracted. Where authors did not explicitly state the most important attribute, the 'range method' recommended by the ISPOR Conjoint Analysis Good Research Practices Task Force was used to infer the relative importance of attributes [27].

2.5 Quality assessment

Methods for assessing the quality of DCEs are an ongoing area of methodological development. As such, we followed previous reviews [17, 28] and adapted the ISPOR [15] checklist in order to critically appraise the studies identified within this review. The checklist was initially designed as a guide to good research practices and reporting for the design and implementation of DCEs and consists of 10 domains essential for conducting high quality DCEs. Each domain contains 3 sub-questions and studies were assigned a score 1 for each sub-question where the criteria was met and 0 if unfilled or unclear, giving a maximum score of 30. Since the checklist has been adapted for our purposes, 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.

3. Results

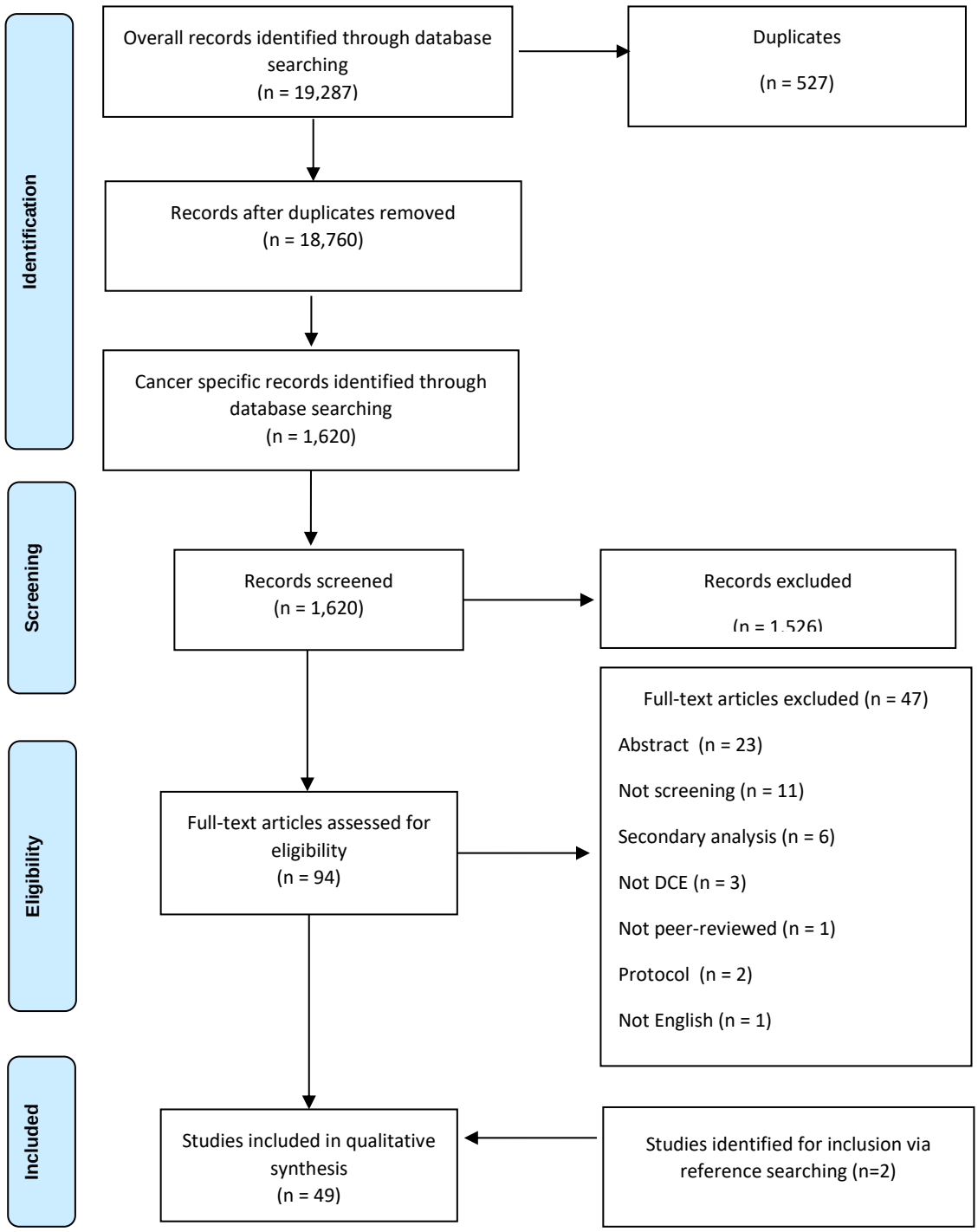
Study selection is shown in the PRISMA diagram in Fig. 1 Once duplicates were removed, there were 1,620 studies. After screening titles and abstracts, 1,526 were excluded, leaving 94 studies to be screened at full-text. During the full-text screening stage, 47 studies were excluded. Two additional studies were found through manual searches of reference lists and forward citations, meaning 49 studies were included in this review.

3.1 Quality of reporting assessment

Scores ranged from 12 to 28 (mean=23) out of 30 (see ESM 2). Common issues related to the reporting and justification of sample size calculations (27/49; 55%), experimental design (30/49; 61%) and model estimation methods (27/49; 55%). In many instances, it was hard to differentiate between poor practice and a lack of detail in reporting. Questions relating to

evidence-based attribute and level selection were generally well-fulfilled (Question 2) indicating methods were clearly reported and justified. Assessing the appropriateness of attribute presentation was more difficult since access to survey instruments was often limited and the wording and definitions used in manuscripts often differed from what was shown to respondents.

Fig 1: PRISMA diagram for systematic review adapted from Moher, et al.



3.2 Study characteristics

Key study characteristics are summarised in Table 1. Cancer screening DCEs have increased with over 40% (21/49) of studies being published in the last five years. Studies were primarily clustered around five countries: Australia (12/49; 24%) [29-40], USA (9/49; 18%) [35, 36, 41-47], Netherlands (8/49; 16%) [48-55], UK (7/49; 14%) [56-61] and France (7/49; 14%) [62-68]. Studies commonly examined cancer sites with clear existing evidence supporting the efficacy of population-based screening (e.g. cervical 11/49; 22% [32, 34, 56, 60, 64, 68-73], breast 8/49; 16% [30, 47, 61, 64, 66, 73-75]), with colorectal cancer being the most common (26/49; 53%). The remaining studies considered prostate [35, 37, 50, 67], oesophageal [55] and skin cancer screening [38, 40] where the potential benefits and harms of national screening programs are debated [76]. 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 programs where there was an established modality with no close alternatives, such as the use of mammograms for breast cancer screening.

DCEs typically collected preferences of a single population (42/49; 86%), most frequently the general public with mixed prior experience of screening (36/49; 73%), followed by HCPs (10/49; 20%) [41, 51, 52, 60, 61, 66, 67, 70, 74, 77] with patients (9/49; 18%) [29, 30, 33, 34, 38, 40, 70, 72, 77] being considered least often. The target population of each study was reported and classified according to the authors' own definition; therefore the definition of "patients" varied and included those attending screening appointments or purchasing tests [29, 30, 33, 34, 38, 40, 70] and those attending appointments for other reasons [72, 77]. Notably no studies exclusively elicited the preferences of cancer patients.

Sample size calculations were described in less than half of studies (24/49; 49%). Final sample sizes varied from 38-2067 with 45% (22/49) of studies having a sample size between 201-500 participants. For the most part, data were collected using a self-completed survey

administered online (22/49; 45%) and/or by post (19/49; 39%) with just twelve studies (27%) opting to collect data face-to-face [31, 43, 45, 57, 59, 69-75]. Response rates were described inconsistently across studies varying from 5-89% where reported. Self-completed questionnaires encountered lower response rates than interview-administrated studies.

Piloting or pre-testing was reported in 41 studies, with the purpose of testing explicitly reported in the majority of studies. However, the methods used during pre-testing were reported in only 27 studies [30, 33, 35-37, 40, 45-49, 51, 52, 54-61, 65-68, 71, 73-75, 78]. The validity or quality of responses was explicitly verified in 27 (55%) studies using rationality check questions, most commonly by including a choice task with a dominant alternative to check for non-satiation (18/49; 37%) [31, 37, 41-44, 48-50, 55-59, 61, 67, 72, 78]. Further details of each study are found in ESM 3.

Table 1: Study characteristics

Characteristic		n (%)	References
Year of publication	2000-2005	4 (8%)	[29-31, 56]
	2006-2010	9 (18%)	[32, 33, 41, 48, 49, 62, 63, 69, 78]
	2011-2015	15 (31%)	[34-37, 42-44, 50-53, 57-59, 70]
	2016- present	21(43%)	[38-40, 45-47, 54, 55, 60, 61, 64-68, 71-75, 77]
Country*	Australia	12 (24%)	[29-40]
	Canada	2 (4%)	[41, 78]
	France	7(14%)	[62-68]
	Netherlands	8 (16%)	[48-55]
	UK	6 (12%)	[56-61]
	USA	9 (18%)	[35, 36, 41-47]
	Other	8 (16%)	[69-75, 77]
Cancer site and intervention(s)*	**Breast**	**8 (16%)**	
	Mammogram	7 (14%)	[30, 47, 61, 64, 66, 73, 75]
	Other	2 (4%)	[74, 75]
	Cervix	**11 (22%)**	
	Standard care/Pap-smear	9 (18%)	[32, 34, 56, 60, 64, 68, 69, 71, 73]
	Liquid-based cytology	3 (6%)	[32, 34, 56]
	Human papillomavirus (HPV) testing	2 (4%)	[32, 72]
	Self-sampling	3 (6%)	[60, 68, 70]
	Other	2 (4%)	[32, 70]
	Colon	**26 (53%)**	
	Faecal occult blood test (FOBT)	22 (45%)	[29, 31, 33, 36, 39, 41-46, 48, 49, 51-53, 62-65, 77, 78]
	Faecal Immunochemical Test (FIT)	5 (10%)	[39, 46, 49, 54, 77]
	Colonoscopy	10 (20%)	[36, 41-46, 48, 77, 78]
	CT colonography	10 (20%)	[36, 41-45, 57-59, 78]
	Sigmoidoscopy	10 (20%)	[36, 42-46, 48, 49, 77, 78]
	Barium enema	3 (6%)	[41, 77, 78]
	Blood test	4 (8%)	[39, 51, 52, 63]
	Other	3 (6%)	[39, 53, 77]
	Oesophagus	1 (2%)	
	Upper endoscopy	1 (2%)	[55]
	Nasal endoscopy	1 (2%)	[55]
	Pill on a string	1 (2%)	[55]
	Breath test	1 (2%)	[55]
	Blood test	1 (2%)	[55]
	Prostate	**4 (8%)**	

	Prostate-specific antigen (PSA) test	4 (8%)	[35, 37, 50, 67]
	Skin	**2 (4%)**	
	Teledermoscopy	2 (4%)	[38, 40]
	Self-examination	2 (4%)	[38, 40]
	Examination by professional	2 (4%)	[38, 40]
Population*	General public- mixed experience	36 (73%)	[31-33, 35-37, 39-46, 48-60, 63, 66, 67, 69, 71, 73-75, 78]
	General public- Screening naïve only	1 (2%)	[60]
	Patients/ previously screened individuals only	9 (18%)	[29, 30, 33, 34, 38, 40, 70, 72, 77]
	Healthcare providers	10 (20%)	[32, 41, 47, 57, 59, 62, 64, 65, 68, 69]
Sample size*	< 50	2 (4%)	[38, 45]
	50-100	7 (14%)	[30, 41, 42, 47, 57, 59]
	101-200	8 (16%)	[32, 40, 43, 44, 49, 64, 68]
	201-500	22 (45%)	[29, 31, 32, 35, 36, 48, 50, 54, 60, 62, 64, 65, 69-75, 77]
	501-1000	9 (18%)	[37, 48, 51, 55, 56, 58, 63, 66, 78]
	>1000	8 (16%)	[33, 39, 41, 46, 51, 53, 61, 67]
Response rate (%)*	<20	5 (10%)	[47, 60, 61, 65, 68]
	20-50	13 (27%)	[30, 39, 40, 42, 48-50, 55, 56, 62, 63]
	51-80	16 (33%)	[29, 31, 33, 44, 46, 49, 51-53, 57-59, 64, 66, 78]
	>80	7 (14%)	[35-37, 43, 57, 64, 75]
	Not stated	12 (24%)	[32, 34, 41, 45, 54, 67, 69, 71-74, 77]
Survey administration method*	Self-completed- online	22 (45%)	[35-41, 46, 47, 51-54, 57-61, 64, 66-68]
	Self-completed- postal	18 (37%)	[29, 30, 32, 33, 38, 39, 42, 48-50, 55, 56, 60, 62, 63, 65, 68, 78]
	Interview-led/ in-person	12 (24%)	[31, 43, 45, 57, 59, 69-75]
	Not specified	1 (2%)	[77]
Attribute selection methods*	Literature review/ existing literature	39 (80%)	[30, 32, 33, 35-38, 41-55, 60-68, 70, 71, 73-75, 77, 78]
	Expert opinion	17 (35%)	[37, 41, 43, 45, 48-50, 54, 55, 61-63, 67, 70, 71, 77, 78]
	Interviews with target population	15 (31%)	[37, 44, 47-50, 60, 61, 64, 66, 68, 71, 73-75]
	Focus groups with target population	10 (20%)	[29, 31, 41, 54, 55, 66, 69, 70, 73, 78]
	Previous work by authors	7 (14%)	[34-36, 38, 40-42]
	None or authors assumptions only	4 (8%)	[56-59]
	Other	7 (12%)	[31, 32, 39, 40, 46, 69, 72]
Level specification methods*	Literature review/ existing literature	28 (57%)	[30, 32, 34-36, 39, 40, 44, 48-55, 58, 59, 61, 63, 64, 71, 75, 78]
	Current policy/practice	8 (16%)	[32, 34, 54, 56, 60, 65, 66, 78]
	Qualitative methods	5 (10%)	[42, 44, 50, 73, 74]
	Expert opinion	4 (8%)	[30, 50, 54, 71]
	Trial data	5 (10%)	[31, 62, 63, 67, 69]
	None reported/ authors assumptions	8 (16%)	[29, 41, 43, 47, 68, 70, 72, 77]
	Other	6 (12%)	[35, 36, 38, 40, 56, 57]
Purpose of piloting*	Coverage of attributes and levels was checked	34 (69%)	[30, 32-37, 40, 43, 46, 48-55, 58, 60-67, 69-71, 73-75, 78]
	Understanding and complexity was checked	36 (73%)	[30, 32-37, 40, 43, 44, 46, 48-62, 64-67, 70, 71, 73-75, 78]
	Length and timing was checked	8 (16%)	[30, 48-50, 54, 56, 59, 67]
	Piloting was performed but purpose not stated	2 (4%)	[41, 47, 68]
	No piloting reported	8 (16%)	[29, 31, 38, 39, 42, 45, 72, 77]
Number of choice tasks per respondent	8≥	5 (10%)	[46, 56, 58, 63, 77]
	9-16	37 (76%)	[29, 30, 35-42, 44, 45, 47, 48, 50-55, 57, 59-62, 64-74, 78]
	16<	7 (14%)	[31-34, 43, 49, 75]
Rationality/ validity checks*	Stability (test-retest)	5 (10%)	[33, 41, 57, 59, 71]
	Non-satiation (dominant choice task)	18 (37%)	[31, 37, 41-44, 48-50, 55-59, 61, 67, 72, 78]
	Transitivity	1 (2%)	[73]
	Dominant attribute	4 (8%)	[31, 33, 41, 78]
	No variation in alternative chosen (flat-lining)	1 (2%)	[64]
	External validity	1 (2%)	[54]
	Other	4 (8%)	[30, 51, 52, 70]
	None/ not reported	22 (45%)	[29, 32, 34-36, 38-40, 45-47, 53, 60, 62, 63, 65, 66, 68, 69, 74, 75, 77]
*Some studies fall into multiple categories so n>49.			

3.3 Identification and selection of attributes and levels

The number of attributes per choice task ranged from 2 to 13, with an average 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, 71% (35/49) of studies took a multi-method approach, most commonly combining literature reviews (39/49; 80%) with an additional qualitative and/or quantitative source such as expert opinion (17/49; 35%) [37, 41, 43, 45, 48-50, 54, 55, 61-63, 67, 70, 71, 77, 78]. Over half the studies (27/49) reported consulting the target population throughout the attribute selection process, typically using qualitative methods such as interviews (15/49; 31%) [37, 44, 47-50, 60, 61, 64, 66, 68, 71, 73-75] and focus groups (10/49; 20%) [29, 31, 41, 54, 55, 66, 69, 70, 73, 78]. 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 [56-59].

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 [29, 41, 43, 47, 68, 70, 72, 77] . The number of levels ranged from two to 12, averaging at three per attribute. Where described, literature reviews (28/49; 57%) were the most common source for obtaining attribute levels, with existing policy (8/49; 16%) [32, 34, 54, 56, 60, 65, 66, 78], qualitative methods (5/49; 8%) [42, 44, 50, 73, 74], trial data (5/49; 10%) [31, 62, 63, 67, 69] and expert opinion (4/49; 8%) [30, 50, 54, 71] also cited. Levels were modelled categorically using dummy or effects coding for half of all attributes across all studies (139/280; 50%) and twenty studies used categorically coded attributes exclusively. Alternatively, twenty-one studies applied a combination of categorical and continuous attribute specifications and 9 studies modelled all attributes as continuous [33, 35, 37, 48, 50, 56, 61, 67, 69]. Thirty-four studies (69%) explicitly acknowledged checking the coverage of attributes and levels during an iterative piloting stage. A full summary of attribute levels can be found in ESM 4.

3.4 Attribute classification

In total, 280 attributes were included across all studies. Similarities between attributes meant they could be grouped into 28 general attributes that were then classified into the four previously outlined categories: 1) test-specific; 2) outcomes; 3) service delivery; and 4) monetary.

Fig. 2 provides a summary of the results of the attribute classification analysis. Each dot represents the inclusion of one of the 28 identified general attributes. The dots are colour coded to indicate the type of information contained in the attribute (i.e. risk/time) and the type of dot indicates the significance of each attribute. Information about the complexity, importance and marginal rate of substitution are also displayed. Several studies included single attributes that contained information relating to two or more of the distinct 'general' attributes. We refer to these attributes as 'compound' attributes as indicated by 'c'. Further explanation can be found in the detailed key at the bottom of the table. A full summary of the attributes in each study can be found in the ESM 4.

Studies generally examined trade-offs across combinations of attributes spanning the four categories, with some categories occurring more frequently. Test-specific attributes included sensitivity, specificity and test procedure and were the most common category overall, appearing in almost 90% (43/49) of studies and accounting for 35% (102/280) of all attributes. Service delivery attributes such as test interval, test location and type of HCP were also very common, accounting for 36% (100/280) of all attributes and occurring at least once in 36/49 (73%) studies. Outcome attributes were less common, appearing in just 25/49 (51%) studies and accounting for 17% (48/280) of all attributes. Despite occurring least frequently overall (37/289; 13%), monetary attributes were well-dispersed across studies with 31/49 (63%) including at least one monetary attribute. The individual attributes occurring most frequently across studies spanned multiple categories and included; costs (31/49; 63%); specificity (28/49; 57%); sensitivity (24/49; 49%) and mortality reduction (22/49; 45%).

3.5 Significance of attributes

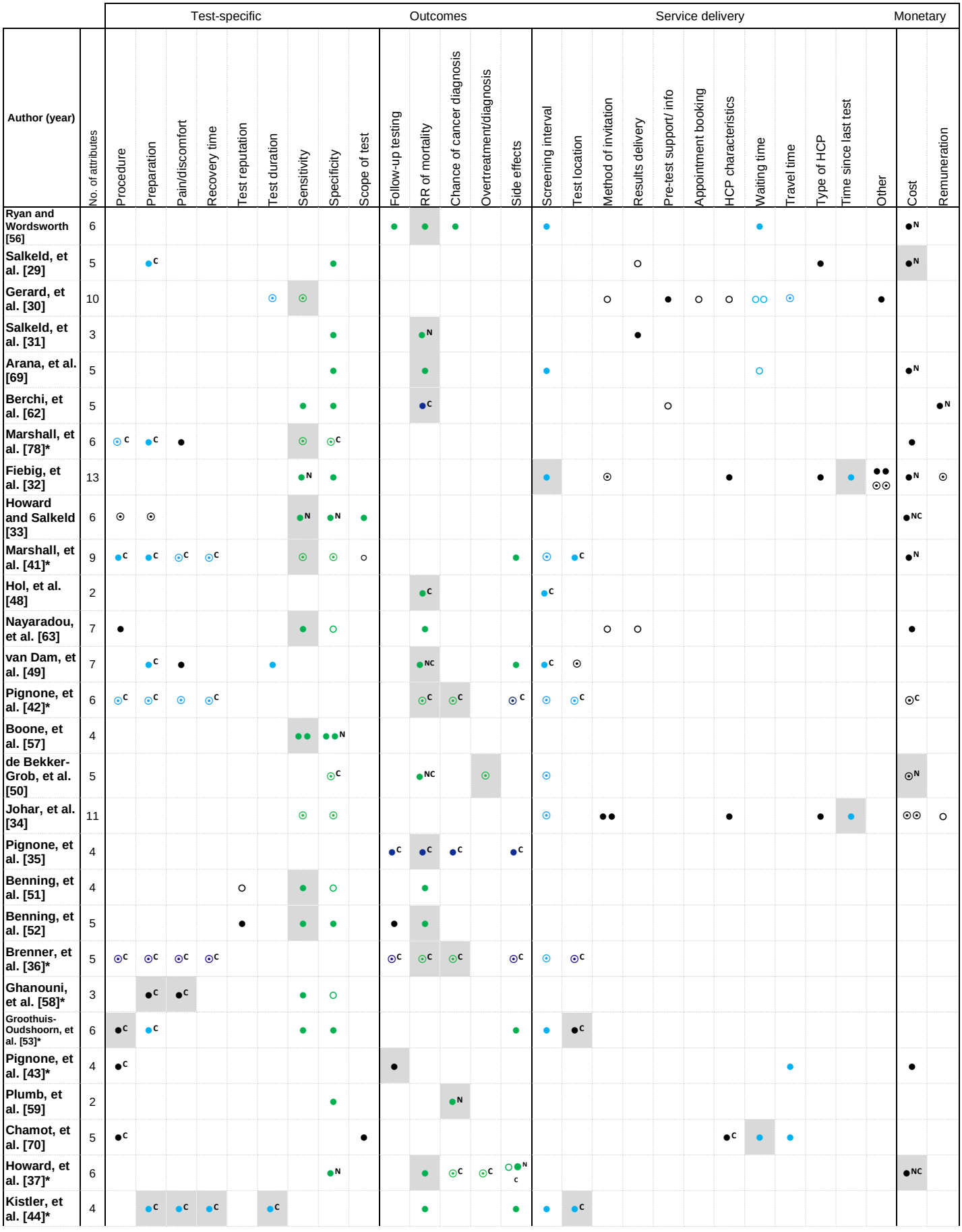
All but two studies [36, 42] reported significance of attributes. Attributes were generally found to be statistically significant (i.e. important to respondents), with just 10% found to be insignificant (26 out of 269 attributes with reported significance levels). However, variations in significance across attribute levels were not uncommon (70/269, 26%). Five attributes were consistently significant across all studies where they appeared; sensitivity, overdiagnosis/treatment, travel time, type of HCP and time since last test. The significance of attributes did not vary greatly across the four attribute categories.

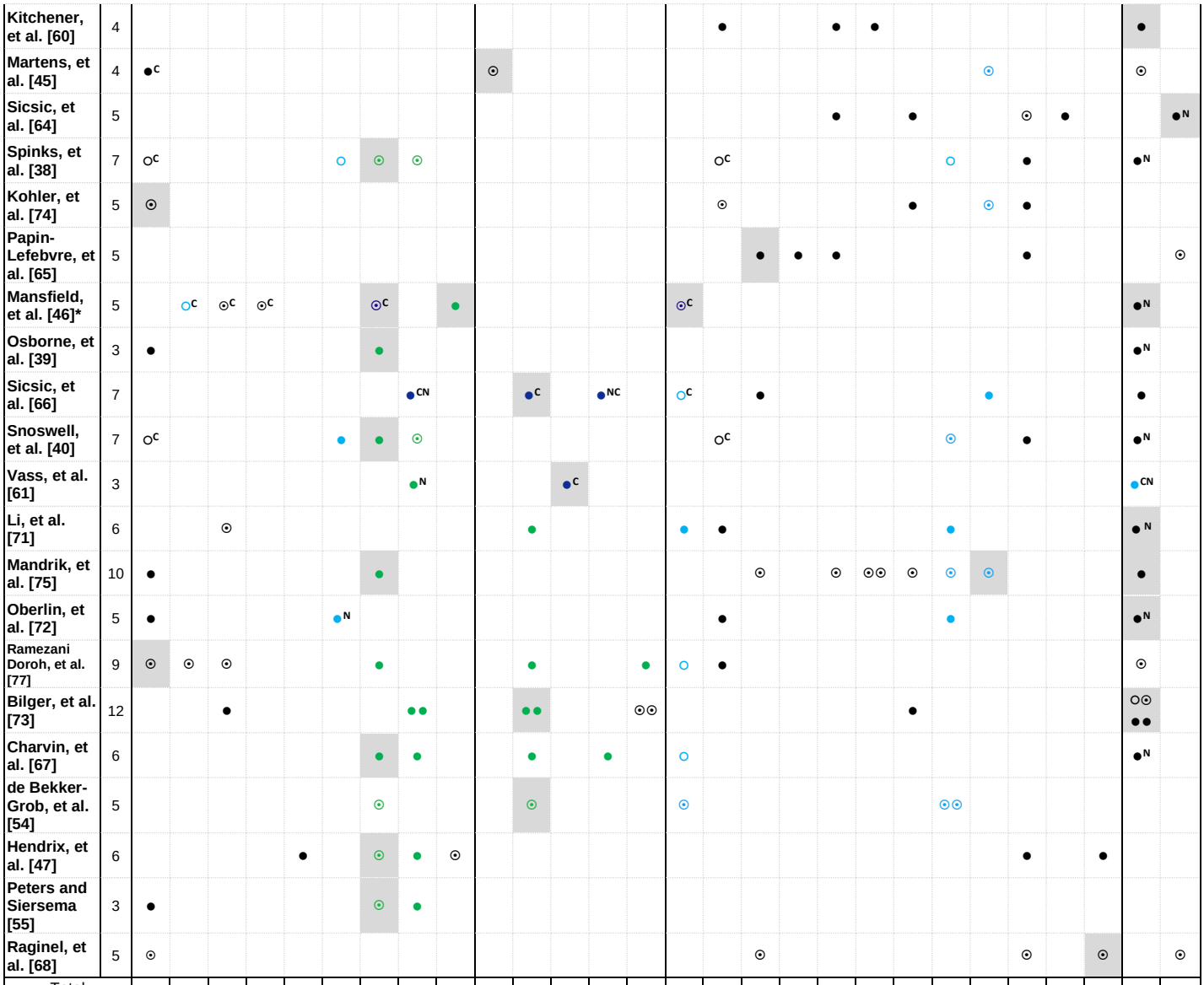
3.6 Complexity of attributes

The complexity of attributes varied across studies in several ways. Firstly, the inclusion of risk of cancer was prominent, with over 80% (40/49) studies including at least one risk attribute and overall 35% (97/280) of all attributes incorporating risk or probability. Attributes including risk were exclusive to the test-specific and outcome categories and were most frequently incorporated into attributes relating to sensitivity, specificity and mortality reduction. Risks were typically communicated using frequencies (57/90; 63%), percentages (29/90; 32%) or a combination of the two (9/90; 10%). Visual aids to help respondents interpret risk information were rarely used (18/90; 20%).

The interpretation of risk information particularly in a medical context is cognitively challenging. In twenty-two studies (55%) this was further complicated by inconsistencies in presentation. For instance, risks were typically negatively framed (60/90; 67%), however, inconsistent framing across attributes within studies was common (16/40; 40%) [31, 36, 38, 40, 42, 44, 47, 49, 50, 53, 54, 57-59, 62, 63]. For example, expressing some attributes positively (e.g. mortality change expressed as “lives saved” or “chance of survival”) and other attributes negatively (e.g. sensitivity expressed as “cases missed” or “chance of false-negative”). Other issues included inconsistencies in denominator when risks were displayed as frequencies (5/40; 13%) [32, 34, 38, 40, 63] (e.g. 1 in 100 and 1 in 1000 as opposed to 10 in 1000 and 1 in 1000) and variation in the risk presentation format across attributes, for instance switching between percentages and frequencies (9/40; 23%) [36, 38, 40, 42, 49, 51-53, 63].

Fig. 2: Attribute classification by study. Each circle represents an occurrence of an attribute within each study. The shape of the circle represents information about the significance of the attribute and the colour indicates the type of information contain within each attribute, as described in the key at the bottom of the table.





Kitchener, et al. [60]	4																														
Martens, et al. [45]	4	●C										⊙																			
Sicsic, et al. [64]	5																													●N	
Spinks, et al. [38]	7	○C					○	⊙	⊙								○C						○			●			●N		
Kohler, et al. [74]	5	⊙															⊙					●		⊙	●					
Papin-Lefebvre, et al. [65]	5																	●	●	●					●					⊙	
Mansfield, et al. [46]*	5		○C	⊙C	⊙C			⊙C		●							⊙C												●N		
Osborne, et al. [39]	3	●						●																					●N		
Sicsic, et al. [66]	7								●CN			●C		●NC			○C		●					●					●		
Snoswell, et al. [40]	7	○C					●	●	⊙									○C						⊙		●			●N		
Vass, et al. [61]	3								●N				●C																●CN		
Li, et al. [71]	6			⊙								●					●	●						●					●N		
Mandrik, et al. [75]	10	●						●											⊙		⊙	⊙⊙	⊙	⊙	⊙				●		
Oberlin, et al. [72]	5	●					●N											●						●					●N		
Ramezani Doroh, et al. [77]	9	⊙	⊙	⊙				●				●			●		○	●											⊙		
Bilger, et al. [73]	12			●					●●			●●			⊙⊙								●						○⊙ ●●		
Charvin, et al. [67]	6							●	●			●		●			○												●N		
de Bekker-Grob, et al. [54]	5							⊙				⊙					⊙							⊙⊙							
Hendrix, et al. [47]	6					●		⊙	●	⊙																●		●			
Peters and Siersema [55]	3	●						⊙	●																						
Raginel, et al. [68]	5	⊙																	⊙							⊙		⊙		⊙	
Total occurrences (% of studies)		**19** (39%)	**12** (24%)	**11** (22%)	**5** (10%)	**3** (6%)	**6** (12%)	**24** (47%)	**28** (53%)	**5** (10%)	**6** (12%)	**21** (41%)	**7** (14%)	**4** (8%)	**12** (20%)	**18** (37%)	**13** (27%)	**9** (16%)	**4** (8%)	**6** (12%)	**4** (6%)	**7** (14%)	**12** (20%)	**7** (14%)	**10** (20%)	**3** (6%)	**7** (8%)	**31** (55%)	**6** (12%)

Colour of the circle corresponds to the type(s) information included in attributes: ●○⊙ no risk or time, ○●⊙ risk only, ○●⊙ time only, ○●⊙risk *and* time
Type of dot corresponds to the significance of attributes at a 95% confidence level: ● significant, ○ not significant, ⊙unclear or mixed significance across levels
C Compound attribute N Numeraire in MRS estimates *Study contains compound attributes that span multiple categories.
The most important attribute(s) from each study is shaded grey- some studies have more than one most important attribute (e.g. where using latent class analysis, compound attributes or including more than one population in a single study)

Attributes were further complicated through the inclusion of time (e.g. costs over a lifetime) with 32/49 (65%) studies and 25% (69/280) of all attributes including a time element. The inclusion of risk and time frequently overlapped, not only within a choice task (28/49; 57%) but also within a single attribute (e.g. describing a risk over time) in several instances (13/280; 5%).

The combination of risk and time within a single attribute is an example of the frequent use of 'compound' attributes across the studies within this review (23/49; 47% studies and 49/280; 18% attributes). Compound attributes add to the complexity of choice tasks by merging multiple aspects of a screening test into a single attribute; for example authors frequently combined details on test procedure, test duration, pain/discomfort and/or recovery time in a single attribute relating to 'process'/'procedure'.

3.7 Most important attributes for screening

The importance of attributes varied greatly (Fig. 3). In total there were 58 'most important' attributes across 49 studies due to method of analysis (e.g. latent class analysis) or multiple populations included within some studies. Additionally, some attributes spanned multiple categories within our analysis because they were classified as compound. Time since last test (2/3; 67%) [32, 34], sensitivity (16/24; 67%) [30, 33, 38-41, 46, 47, 51, 52, 55, 57, 63, 67, 75, 78] and mortality risk reduction (14/22; 64%) [31, 35, 37, 42, 48, 49, 51, 52, 54, 56, 62, 66, 69, 73] were most frequently identified as the most important attributes. Ten attributes including side effects, specificity and results delivery were never rated most important.

When comparing attribute categories, outcome attributes were rated most important in nineteen of the 25 studies (76%) where they were featured [31, 35-37, 42, 43, 45, 48, 49, 52, 54, 56, 57, 59, 61, 62, 66, 69, 73]. Test-specific attributes were also preferred and found to be most important in twenty-two out of 43 studies including at least one test-specific attribute (51%). Monetary attributes were less of a priority, being rated most important in ten out of 31 studies (32%) [29, 37, 46, 50, 60, 64, 71-73, 75] and service delivery attributes were top priority

in just nine of the thirty-six studies (25%) where they were featured [32, 34, 44, 46, 65, 68, 70, 75].

When considering complexity, over half (33/58; 57%) of the most important attributes contained risk information. Twelve attributes (21%) included a time element and fourteen (24%) were compound attributes.

Level ranges also appeared to play a role in the significance of attributes. For example, for the fifteen studies where sensitivity was the most important attribute, the average level range covered 34 percentage points with an average range from 54 to 88%. For studies where sensitivity was not ranked the most important the average level range covered 24 percentage points on average with an average range of 64-89% (ESM 4).

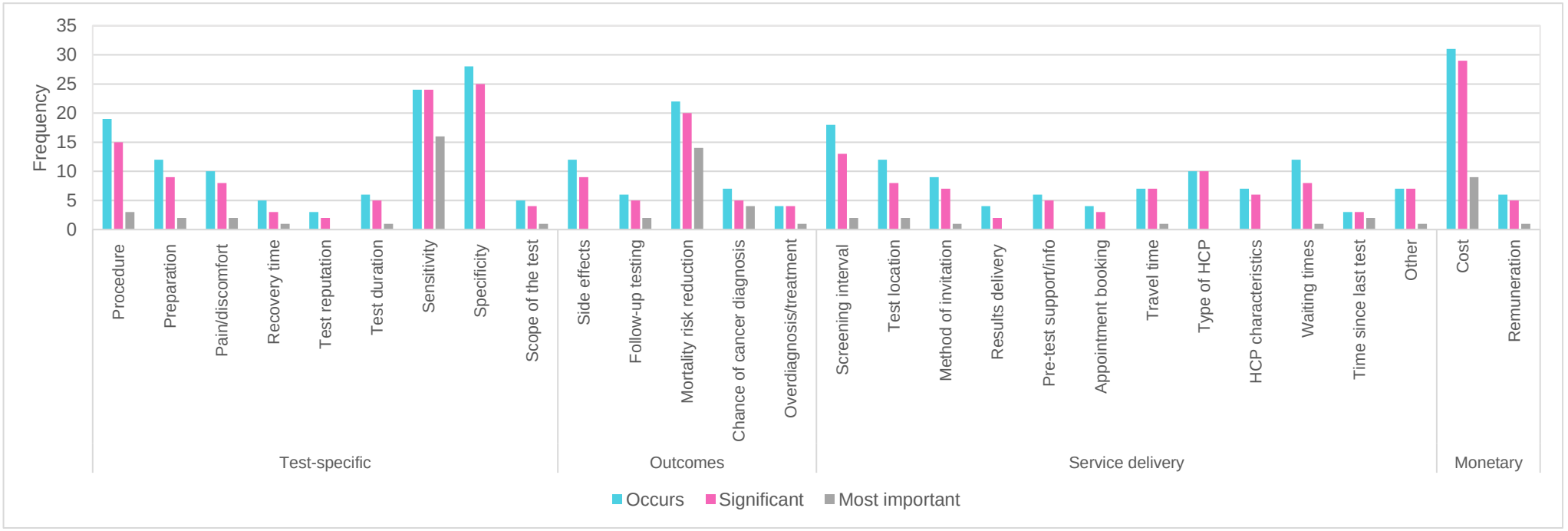
3.8 Trade-offs between attributes

Twenty-three studies (47%) attempted to quantify the relative importance of attributes and the willingness to trade between different components of cancer screening (i.e. marginal rates of substitution) using a common metric (e.g. costs, false-positives). Most commonly, studies employed costs associated with screening as a proxy for marginal utility of income, allowing for an indirect measurement of the marginal willingness-to-pay (mWTP) for a unit improvement in the attribute level. In total, twenty-seven out of 49 studies (55%) included a screening-related cost attribute, with sixteen of these studies calculating WTP [29, 32, 33, 37-41, 46, 50, 56, 61, 67, 69, 71, 72]. Costs were most commonly expressed as the cost of the test/screening (15/27; 56%) [29, 32-34, 39, 41, 56, 60, 63, 68, 69, 72, 75, 78] or ‘out of pocket costs’ (12/27; 37%) [37, 38, 40, 42, 46, 50, 61, 66, 67, 71, 73]. Generally studies described costs on a ‘per test’ basis (24/27; 89%); however, three studies introduced further cognitive complexity by expressing costs over a longer period such as 10-years [37, 42] or lifetime [61], thereby introducing the need to consider concepts of discounting and delayed payments.

Thirteen studies (27%) used an alternative numeraire alongside or instead of mWTP [31-33, 37, 49, 50, 57, 59, 61, 62, 64, 66, 72]. Alternative numeraires included: willingness to accept

measures using reductions in remuneration pay for doctors [62, 64]; cancer deaths prevented [49, 50]; false-positive rates [33, 37, 59, 61]; false-negative rates [32, 66], rates of overdiagnosis [66]; and side effects [37]. In such instances, the marginal rates of substitution were expressed in the natural units of the alternatively specified numeraire. For example, de Bekker-Grob, et al. [50] reported that men were willing to pay €37.60 a year or willing to forego a 1.8% risk reduction in cancer mortality for a 10% reduction in the risk of unnecessary biopsies associated with prostate cancer screening

Fig. 3: Results of attribute analysis. This figure demonstrates how often an attribute appeared, was significant and was rated most important across all 49 studies



4. Discussion

This review demonstrates that the number of DCEs for cancer screening continues to grow, with colorectal cancer screening studies accounting for over half of the included studies.

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. Studies mostly included a variety of attributes but there was a clear emphasis on the inclusion of test-specific and service delivery characteristics. The prioritisation of these types of attributes reflects the nature of population-level screening which requires proactive and voluntary participation from asymptomatic individuals meaning issues relating to the convenience and burden of the intervention are likely to relevant to uptake and adherence. Despite high rates of inclusion, service delivery attributes were rarely prioritised by respondents. Instead, outcome attributes ranked highest most often, followed by test-specific characteristics, especially test sensitivity.

We have identified a number of methodological considerations relating to the specification and interpretation of attributes which have impacts on the comparability of findings between studies. We discuss these considerations below under the five broad headings: 1) differences in level range; 2) oversimplification; 3) compound attributes; 4) risk information and 5) marginal rates of substitution and choice of numeraire. Addressing these methodological considerations with future studies may help to reduce the current gap between stated choices and actual screening behaviour [17, 19, 21].

4.1 Differences in level range

Level range should be wide enough to allow levels to be meaningfully different from each other whilst also being clinical realistic and relevant to policy [14]. Although, level ranges were generally realistic (even where descriptions of selection methods were scarce), variations across studies examining the same tests were common. The choice of attribute level has an

impact on the overall findings of a study, since attribute importance is a direct function of the level range of attributes included within a given study. Taking a closer look at studies including sensitivity appears to further support this argument, with studies covering wider ranges generally demonstrating higher relative importance (average level range was 34 percentage points for studies where sensitivity was ranked most important compared to 24 percentage points on average for studies where an alternative attribute was most important).

4.2 Oversimplification

The oversimplification of attributes represents a further challenge to the interpretation of findings, particularly where authors attempted to draw inferences about uptake. The complexity and nuances of medical decisions means participants are rarely, if ever, provided with a sufficient level of detail to make a truly informed choice, reflective of their true behaviour. Notable instances of oversimplification include explanations of “preparation”, particularly where dietary restrictions and/or laxative use are required [29, 46, 49, 77]; two aspects of testing that represent some of highest burdens placed on test users and are therefore likely to affect uptake/adherence. The reduction of a complex, multidimensional test into a limited number of attributes further highlights the importance of a rigorous attribute selection process to capture the most influential and representative aspects of tests to the target population.

4.3 Compound attributes

In many instances, authors attempted to overcome the issue of oversimplification by employing the use of compound attributes. Compound attributes merge multiple aspects of a test into a single attribute, thereby increasing the level of information and complexity relayed to respondents within choice tasks whilst keeping the overall number of attributes low. The use of these types of attributes appears to be an enduring and expansive trend, making up 18% (49/280) of all attributes, occurring within 47% (23/49) of studies in this review. For instance, a single ‘procedure’ attribute containing information on test modality, duration and pain/discomfort may be equal or even more cognitively challenging, than including three

separate attributes capturing the same issues. Furthermore, with compound attributes it is not possible to disaggregate the characteristics to allow for independent evaluation, meaning the underlying driver of preferences within compound attributes will remain unclear. The complexity of compound attributes may dominate the other attributes, particularly in an otherwise straightforward choice task, since the disproportionate amount of information may unintentionally signal that the attribute is implicitly more important. There are certain instances where characteristics of a test are interrelated and therefore logically can/should be combined. For example, in the only reference to compound attributes we could identify, Mühlbacher, et al. [79] cite the example of side effects. The importance of the risk of a side effect cannot be fully evaluated without also considering its severity and/or duration. In instances where either of these characteristics are omitted, they may be incorrectly assumed leading to biased results. Similarly, expressing attributes over time is fairly standard and is actually recommended for risks and costs [80], but adds an additional layer of complexity for those specific attributes, introduces the complication of time preferences and likely leads to respondent making assumptions/interpretations that may alter the meaning of the attribute and demands greater attention (e.g. people may attempt to convert in to annual or per-screen risks/costs). Overall, the possible implications of compound attributes have been given little attention in the DCE literature to date: however, this study found around a third of the most influential attributes across studies were compound (29%; 14/49) [35-37, 42, 44, 46, 48, 49, 53, 58, 59, 61, 62, 66].

4.4 Risk information

Over a third of all attributes across included studies incorporated a risk or probability element (35%; 97/280). 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 specifically focused on how variations risk presentation may affect results. For example, Howard and Salkeld [33] compared the impact

of positive or negative framing of risks in the context of colorectal cancer screening and found attribute framing significantly influenced WTP estimates in several instances, particularly when considering accuracy. A notable finding given 16 of the 40 DCEs containing risk in this review contained inconsistencies in framing across attributes and many DCEs did not use visual aids to communicate risk. Following best practice guidelines on risk presentation seems an important step to allow comparability between studies [80].

4.5 Marginal rates of substitution- choice of numeraire

Whilst the number of studies relating to cancer screening is increasing, differences across studies means there is limited ability to combine findings and draw overall conclusions about the importance of attributes for different cancer sites or using different tests. One way to increase comparability is by calculating marginal rate of substitution (MRS). However, despite many authors referencing the ability to calculate the willingness-to-trade between attributes as a motivation for their study, in practice only 47% (23/49) explicitly calculated MRS.

For studies that formalised the willingness-to-trade between attributes, costs were the dominant choice of numeraire. Willingness-to-pay (WTP) estimates are a common and widely accepted way of providing decision makers a transparent metric to compare the importance of different interventions or aspects of interventions when thinking about prioritisation. However, inclusion of costs for countries where screening tests are free at the point of use through universal access to national health services (e.g. Australia, UK) or subsidisation within private (USA) and state-managed (Netherlands) insurance-funded systems may be problematic. In such instances, the inclusion of costs, particularly when expressed as direct out-of-pocket payments may be viewed as inappropriate or unrelatable by respondents, increasing the likelihood of protest responses, refused responses and dishonest or tactical answers [81]. It is therefore important to consider the most appropriate payment vehicle under these circumstances.

When considering the inclusion of monetary aspects associated with cancer screening, perhaps a more relevant, and currently underexplored issue for countries without universal healthcare coverage, are follow-up costs (e.g. in the USA screening is often fully covered by insurance companies but individuals are usually responsible for a share of any follow-up testing or subsequent treatments). The consideration of follow-up costs is likely to have important implications for the other commonly occurring attributes such as overtreatment/diagnosis and in particular, test specificity which were found to be inconsistently significant/important across studies. This may partially be due to a failure to fully consider the financial consequences of a false-positive result as studies typically focus on the emotional or physical consequences of follow-up testing. Follow up costs were included in only two studies [43, 45] but were found to be significant and the most important attribute in both instances.

Alternative numeraire were used in a number of studies, however, with the exception of mortality reduction. However, these alternatives tend to be specific to a screening/testing setting (e.g. false-positives, overdiagnosis, false-negatives) meaning that the ability to compare MRS across interventions is limited if broader prioritisation questions are of interest.

4.6 Recommendations for future research

Consistently low or even declining rates of screening suggest that understanding the preferences and decisions around screening remain an important area of research and studies attempting to identify possible barriers to uptake may be particularly important [54]. However, despite the increased number of cancer screening DCEs publications, findings remain under-utilised within clinical and policy settings, an issue highlighted in previous systematic reviews [17, 19, 21].

Previous sections have highlighted aspects of attribute development and specification which may unintentionally add to the complexity of choice tasks making comparability of studies and integration into policy potentially harder. Overlooking these methodological considerations may potentially reduce response rates, exclude specific populations (e.g. elderly or cognitively

impaired) and/or increase the use of heuristics leading to non-compensatory decision making [82-85].However, in reality decisions relating to screening are complex and multi-faceted requiring a balanced consideration of the potential benefits and harms. Raising the question: what can we do better in attribute development and specification to make choices clearer to respondents and more directly applicable to policy issues?

To answer this question, we recommend the use of qualitative methods, such as cogitative interviews, during the piloting stage to help to manage and further understand the impact of attribute and level selection, and choice task complexity in future preference-based studies on cancer screening [86]. Cognitive or 'think aloud' interviews can be used to understand how respondents complete choice tasks including areas of misinterpretation, difficulty and poor task attendance [87]. Despite, thirty-seven of the studies within this review reported conducting some form of piloting prior data collection, the use of qualitative methods during the pre-testing stage was under-reported and potentially underutilised by authors. Future studies would also benefit from greater transparency around the piloting process including the methods used, sample size and implications of the findings to the final design of the experiment.

More generally, the incorporation of qualitative methods used to understand the views and experiences of the target audience and ensure relevance of included attributes during the development stage should remain a priority, with less than half of the studies within this review reporting the use of qualitative methods. Additionally, greater transparency in reporting methods and findings of qualitative elements is also needed, an issue that has been discussed in depth by Coast, et al. [16]. We acknowledge detailed reporting is often limited by word limits of journals. More recently authors have sought to overcome this challenge through the separate publication of attribute development processes, piloting or study protocols in addition to the final results [88-93].

The inclusion of patient and public involvement (PPI) throughout the design process is another way to maximise relevance and ensure understanding of attributes, task instructions and survey instruments for the target audience. The use of PPI in preference-based studies has

become increasingly popular in recent years with case studies and guidance papers being recently published [94, 95]. PPI and stakeholder engagement from the outset of future DCEs may help to tailor research and bridge the gap between research and adoption by decision makers increasing the impact of findings. Preferences surrounding emerging technologies (e.g. tele-screening and self-screening modalities [38, 40, 60]) or potential future screening programs currently in the trial stages (e.g. prostate cancer, lung cancer) [96, 97] are key areas where collaboration with stakeholders may be particularly beneficial.

Finally, given the growth in both DCEs and systematic reviews of DCEs, we recommend the development of a risk of bias measure for DCEs covering the methods used to identify, select and construct attributes, as well as the challenges identified in section 4.1-4.6. alongside other methodological aspects. Whilst a number of invaluable DCE-specific best practice guides covering a range of methodological aspects exist [14-16, 27, 98], this measure should seek to incorporate guidance from outside of the discipline, particularly for more nuanced aspects of attribute development including the design and conduct of qualitative research [99], PPI [100] and risk communication [101-103].

4.8 Limitations

This review was conducted using a rigorous search strategy, performed and reported with full transparency. However, a few limitations remain. Firstly, we limited the search to published, peer-reviewed studies meaning some unpublished studies that add to the knowledge base may have been excluded. Secondly, 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 was unavailable this may have led to an unduly critical assessment of a study. Finally, quality assessment was performed using a methodological checklist [15] due to the unavailability of an established risk of bias assessment tool to evaluate discrete choice studies. This enabled us to discuss the quality and limitations of included studies in a broad sense, specifically in relation to the quality of reporting within DCEs but 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. In general, studies scored highly across items relating to attribute and level development (2.1-2.3; 3.1; 6.2) indicating methods were reported and well-justified, however, the methodological quality cannot be fully inferred.

5. Conclusion

Our review provides an in-depth summary of the attributes used to elicit preferences for screening interventions within discrete choice experiments which serves as a useful starting point for researchers designing future DCEs in this area. In general, test sensitivity and mortality reduction are prioritised over service delivery characteristics by most respondents.

We identify several methodological challenges relating to attribute development that future researchers may wish to consider going forward. In particular, we highlight the need to consider the complexity of choice tasks when considering risk information or compound attributes. PPI and stakeholder engagement is highly recommended throughout the development process to ensure the relevance of the research question and to optimise understanding when faced with unavoidably complex choice tasks. Additionally, we argue research into the development of a risk of bias tool and increased transparency in the reporting of attribute development procedures are integral for ensuring quality and maximising comparability across studies and encouraging the application of findings.

References

1. Peirson L, Fitzpatrick-Lewis D, Ciliska D, Warren R. Screening for cervical cancer: a systematic review and meta-analysis. Systematic reviews. 2013;2(1):35.
2. Lin JS, Piper MA, Perdue LA, Rutter CM, Webber EM, O'Connor E, et al. Screening for colorectal cancer: updated evidence report and systematic review for the US Preventive Services Task Force. Jama. 2016;315(23):2576-94.
3. Hirst Y, Stoffel S, Baio G, McGregor L, von Wagner C. Uptake of the English Bowel (Colorectal) Cancer Screening Programme: an update 5 years after the full roll-out. European Journal of Cancer. 2018;103:267-73.
4. Shahidi N, Cheung WY. Colorectal cancer screening: Opportunities to improve uptake, outcomes, and disparities. World journal of gastrointestinal endoscopy. 2016;8(20):733.
5. NHS Digital. Cervical Screening Programme England 2017-182018.
6. Moons L, Mariman A, Vermeir P, Colemont L, Clays E, Van Vlierberghe H, et al. Sociodemographic factors and strategies in colorectal cancer screening: a narrative review and practical recommendations. Acta Clinica Belgica. 2020;75(1):33-41.
7. Petkeviciene J, Ivanauskiene R, Klumbiene J. Sociodemographic and lifestyle determinants of non-attendance for cervical cancer screening in Lithuania, 2006–2014. Public Health. 2018;156:79-86.
8. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. Journal of general internal medicine. 2012;27(10):1361-7.
9. Vass CM, Payne K. Using discrete choice experiments to inform the benefit-risk assessment of medicines: are we ready yet? Pharmacoeconomics. 2017;35(9):859-66.
10. Ran T, Cheng C-Y, Misselwitz B, Brenner H, Ubels J, Schlander M. Cost-effectiveness of colorectal cancer screening strategies—a systematic review. Clinical Gastroenterology and Hepatology. 2019;17(10):1969-81. e15.
11. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete choice experiments in health economics: past, present and future. Pharmacoeconomics. 2019;37(2):201-26.
12. Ali S, Ronaldson S. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. British medical bulletin. 2012;103(1):21-44.
13. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. Applied health economics and health policy. 2003;2(1):55-64.
14. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making. Pharmacoeconomics. 2008;26(8):661-77.
15. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value in health. 2011;14(4):403-13.
16. Coast J, Al-Janabi H, Sutton EJ, Horrocks SA, Vosper AJ, Swancutt DR, et al. Using qualitative methods for attribute development for discrete choice development experiments: issues and recommendations. Health Economics. 2012;21(6):730-41.
17. Ghanouni A, Smith SG, Halligan S, Plumb A, Boone D, Yao GL, et al. Public preferences for colorectal cancer screening tests: a review of conjoint analysis studies. Expert review of medical devices. 2013;10(4):489-99.
18. Marshall D, McGregor SE, Currie G. Measuring preferences for colorectal cancer screening: What are the implications for moving forward? Patient. 2010;3(2):79-89.
19. Phillips KA, Van Bebber S, Marshall D, Walsh J, Thabane L. A review of studies examining stated preferences for cancer screening. Preventing chronic disease. 2006;3(3):A75.

20. Wortley S, Wong G, Kieu A, Howard K. Assessing stated preferences for colorectal cancer screening: A critical systematic review of discrete choice experiments. Patient. 2014;7(3):271-82.
21. Mansfield C, Tangka FK, Ekwueme DU, Smith JL, Guy Jr GP, Li C, et al. Peer reviewed: stated preference for cancer screening: a systematic review of the literature, 1990–2013. Preventing chronic disease. 2016;13.
22. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Annals of internal medicine. 2009;151(4):264-9.
23. Brouwer WB, Culyer AJ, van Exel NJA, Rutten FF. Welfarism vs. extra-welfarism. Journal of health economics. 2008;27(2):325-38.
24. Brouwer WB, van Exel NJA, van den Berg B, van den Bos GA, Koopmanschap MA. Process utility from providing informal care: the benefit of caring. Health policy. 2005;74(1):85-99.
25. Bien DR, Danner M, Vennedey V, Civello D, Evers SM, Hiligsmann M. Patients' preferences for outcome, process and cost attributes in cancer treatment: A systematic review of discrete choice experiments. The Patient-Patient-Centered Outcomes Research. 2017;10(5):553-65.
26. Coast J, Smith RD, Lorgelly P. Welfarism, extra-welfarism and capability: the spread of ideas in health economics. Social science & medicine. 2008;67(7):1190-8.
27. Hauber AB, González JM, Groothuis-Oudshoorn CG, Prior T, Marshall DA, Cunningham C, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. Value in health. 2016;19(4):300-15.
28. Wortley S, Wong G, Kieu A, Howard K. Assessing stated preferences for colorectal cancer screening: a critical systematic review of discrete choice experiments. The Patient-Patient-Centered Outcomes Research. 2014;7(3):271-82.
29. Salkeld G, Ryan M, Short L. The Veil of Experience: Do Consumers Prefer What They Know Best? Health Economics. 2000;9(3):267-70.
30. Gerard K, Shanahan M, Louviere J. Using stated preference discrete choice modelling to inform health care decisionmaking: A pilot study of breast screening participation. Applied Economics. 2003 Jun;35(9):1073-85.
31. Salkeld G, Solomon M, Short L, Ryan M, Ward JE. Evidence-based consumer choice: A case study in colorectal cancer screening. Australian and New Zealand Journal of Public Health. 2003;27(4):449-55.
32. Fiebig DG, Haas M, Hossain I, Street DJ, Viney R. Decisions about Pap tests: What influences women and providers? Social Science and Medicine. 2009;68(10):1766-74.
33. Howard K, Salkeld G. 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. 2009;12(2):354-63.
34. Johar M, Fiebig DG, Haas M, Viney R. Using repeated choice experiments to evaluate the impact of policy changes on cervical screening. Applied Economics. 2013;45(14):1845-55.
35. Pignone MP, Howard K, Brenner AT, Crutchfield TM, Hawley ST, Lewis CL, et al. Comparing 3 techniques for eliciting patient values for decision making about prostate-specific antigen screening: A randomized controlled trial. JAMA Internal Medicine. 2013;173(5):362-8.
36. Brenner A, Howard K, Lewis C, Sheridan S, Crutchfield T, Hawley S, et al. Comparing 3 values clarification methods for colorectal cancer screening decision-making: A randomized trial in the US and Australia. Journal of General Internal Medicine. 2014;29(3):507-13.
37. Howard K, Salkeld GP, Patel MI, Mann GJ, Pignone MP. 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. 2015;18(6):3123-35.
38. Spinks J, Janda M, Soyer HP, Whitty JA. Consumer preferences for teledermoscopy screening to detect melanoma early. Journal of telemedicine and telecare. 2016;22(1):39-46.
39. Osborne JM, Flight I, Wilson CJ, Chen G, Ratcliffe J, Young GP. The impact of sample type and procedural attributes on relative acceptability of different colorectal cancer screening regimens. Patient Preference and Adherence. 2018;12:1825-36.

40. Snoswell CL, Whitty JA, Caffery LJ, Loescher LJ, Gillespie N, Janda M. Direct-to-consumer mobile teledermoscopy for skin cancer screening: Preliminary results demonstrating willingness-to-pay in Australia. Journal of telemedicine and telecare. 2018;24(10):683-9.
41. Marshall DA, Johnson R, Kulin NA, Ozdemir A, Walsh A, Marshall J, et al. 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. 2009;18(12):1420-39.
42. Pignone MP, Brenner AT, Hawley S, Sheridan SL, Lewis CL, Jonas DE, et al. Conjoint analysis versus rating and ranking for values elicitation and clarification in colorectal cancer screening. Journal of General Internal Medicine. 2012;27(1):45-50.
43. Pignone MP, Crutchfield TM, Brown PM, Hawley ST, Laping JL, Lewis CL, et al. 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. 2014;14:611.
44. Kistler CE, Hess TM, Howard K, Pignone MP, Crutchfield TM, Hawley ST, et al. Older adults' preferences for colorectal cancer-screening test attributes and test choice. Patient Preference and Adherence. 2015;9:1005-16.
45. Martens CE, Crutchfield TM, Laping JL, Perreras L, Reuland DS, Cubillos L, et al. 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. 2016;31(4):652-9.
46. Mansfield C, Ekwueme DU, Tangka FKL, Brown DS, Smith JL, Guy GP, et al. Colorectal Cancer Screening: Preferences, Past Behavior, and Future Intentions. Patient. 2018:1-13.
47. Hendrix N, Hauber B, Lee CI, Bansal A, Veenstra DL. Artificial intelligence in breast cancer screening: primary care provider preferences. Journal of the American Medical Informatics Association. 2020.
48. Hol L, De Bekker-Grob EW, Van Dam L, Donkers B, Kuipers EJ, Habbema JDF, et al. Preferences for colorectal cancer screening strategies: A discrete choice experiment. British Journal of Cancer. 2010;102(6):972-80.
49. van Dam L, Hol L, Bekker-Grob EWd, Steyerberg EW, Kuipers EJ, Habbema JDF, et al. What determines individuals' preferences for colorectal cancer screening programmes? A discrete choice experiment. European Journal of Cancer. 2010;46(1):150-9.
50. de Bekker-Grob E, Rose JM, Donkers B, Essink-Bot ML, Bangma CH, Steyerberg EW. Men's preferences for prostate cancer screening: A discrete choice experiment. British Journal of Cancer. 2013;108(3):533-41.
51. Benning TM, Dellaert BGC, Dirksen CD, Severens JL. Preferences for potential innovations in non-invasive colorectal cancer screening: A labeled discrete choice experiment for a Dutch screening campaign. Acta Oncologica. 2014;53(7):898-908.
52. Benning TM, Dellaert BGC, Severens JL, Dirksen CD. 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. 2014;17(5):578-87.
53. Groothuis-Oudshoorn CG, Fermont JM, van Til JA, Ijzerman MJ. Public stated preferences and predicted uptake for genome-based colorectal cancer screening. BMC medical informatics and decision making. 2014;14:18.
54. de Bekker-Grob EW, Donkers B, Veldwijk J, Jonker MF, Buis S, Huisman J, et al. What Factors Influence Non-Participation Most in Colorectal Cancer Screening? A Discrete Choice Experiment. The Patient-Patient-Centered Outcomes Research. 2020:1-13.
55. Peters Y, Siersema PD. Public Preferences and Predicted Uptake for Esophageal Cancer Screening Strategies: A Labeled Discrete Choice Experiment. Clinical and translational gastroenterology. 2020;11(11).
56. Ryan M, Wordsworth S. Sensitivity of willingness to pay estimates to the level of attributes in discrete choice experiments. Scottish Journal of Political Economy. 2000;47(5):504-24.

57. Boone D, Mallett S, Zhu S, Yao GL, Bell N, Ghanouni A, et al. Patients' & healthcare professionals' values regarding true- & false-positive diagnosis when colorectal cancer screening by CT colonography: Discrete choice experiment. PLoS ONE. 2013;8(12):e80767.
58. Ghanouni A, Halligan S, Taylor SA, Boone D, Plumb A, Stoffel S, et al. Quantifying public preferences for different bowel preparation options prior to screening CT colonography: A discrete choice experiment. BMJ Open. 2014;4(4):e004327.
59. Plumb AA, Boone D, Fitzke H, Helbren E, Mallett S, Zhu S, et al. Detection of extracolonic pathologic findings with CT colonography: a discrete choice experiment of perceived benefits versus harms. Radiology. 2014;273(1):144-52.
60. Kitchener HC, Gittins M, Rivero-Arias O, Tsiachristas A, Cruickshank M, Gray A, et al. A cluster randomised trial of strategies to increase cervical screening uptake at first invitation (STRATEGIC). Health Technology Assessment. 2016;20(68).
61. Vass CM, Rigby D, Payne K. Investigating the Heterogeneity in Women's Preferences for Breast Screening: Does the Communication of Risk Matter? Value in Health. 2018;21(2):219-28.
62. Berchi C, Dupuis J-M, Launoy G. The Reasons of General Practitioners for Promoting Colorectal Cancer Mass Screening in France. European Journal of Health Economics. 2006;7(2):91-8.
63. Nayaradou M, Berchi C, Dejardin O, Launoy G. Eliciting population preferences for mass colorectal cancer screening organization. Medical decision making : an international journal of the Society for Medical Decision Making. 2010;30(2):224-33.
64. Sicsic J, Krucien N, Franc C. What are GPs' preferences for financial and non-financial incentives in cancer screening? Evidence for breast, cervical, and colorectal cancers. Social Science and Medicine. 2016;167:116-27.
65. Papin-Lefebvre F, Guillaume E, Moutel G, Launoy G, Berchi C. General practitioners' preferences with regard to colorectal cancer screening organisation Colon cancer screening medico-legal aspects. Health Policy. 2017;121(10):1079-84.
66. Sicsic J, Pelletier-Fleury N, Moumjid N. Women's Benefits and Harms Trade-Offs in Breast Cancer Screening: Results from a Discrete-Choice Experiment. Value in Health. 2018;21(1):78-88.
67. Charvin M, Launoy G, Berchi C. The effect of information on prostate cancer screening decision process: a discrete choice experiment. BMC Health Services Research. 2020;20:1-10.
68. Raginel T, Grandazzi G, Launoy G, Trocmé M, Christophe V, Berchi C, et al. Social inequalities in cervical cancer screening: a discrete choice experiment among French general practitioners and gynaecologists. BMC health services research. 2020;20(1):1-10.
69. Arana JE, Leon CJ, Quevedo JL. The effect of medical experience on the economic evaluation of health policies. A discrete choice experiment. Social Science and Medicine. 2006;63(2):512-24.
70. Chamot E, Mulambia C, Kapambwe S, Shrestha S, Parham GP, Macwan'gi M, et al. Preference for human papillomavirus-based cervical cancer screening: Results of a choice-based conjoint study in Zambia. Journal of Lower Genital Tract Disease. 2015;19(2):119-23.
71. Li S, Liu S, Ratcliffe J, Gray A, Chen G. Preferences for cervical cancer screening service attributes in rural China: a discrete choice experiment. Patient preference and adherence. 2019;13:881.
72. Oberlin AM, Pasipamire T, Chibwesha CJ. Exploring women's preferences for HPV-based cervical cancer screening in South Africa. International Journal of Gynecology & Obstetrics. 2019.
73. Bilger M, Özdemir S, Finkelstein EA. Demand for Cancer Screening Services: Results From Randomized Controlled Discrete Choice Experiments. Value in Health. 2020;23(9):1246-55.
74. Kohler RE, Gopal S, Lee CN, Weiner BJ, Reeve BB, Wheeler SB. Breast Cancer Knowledge, Behaviors, and Preferences in Malawi: Implications for Early Detection Interventions From a Discrete Choice Experiment. Journal of global oncology. 2017;3(5):480-9.
75. Mandrik O, Yaumenenka A, Herrero R, Jonker MF. Population preferences for breast cancer screening policies: Discrete choice experiment in Belarus. PloS one. 2019;14(11).
76. Light A, Elhage O, Marconi L, Dasgupta P. Prostate cancer screening: where are we now? BJU international. 2019;123(6):916-7.

77. Ramezani Doroh V, Delavari A, Yaseri M, Sefiddashti SE, Akbarisari A. Preferences of Iranian average risk population for colorectal cancer screening tests. International journal of health care quality assurance. 2019.
78. Marshall DA, Johnson FR, Phillips KA, Marshall JK, Thabane L, Kulin NA. Measuring patient preferences for colorectal cancer screening using a choice-format survey. Value in Health. 2007;10(5):415-30.
79. Mühlbacher A, Bethge S, Sadler A. Compound Attributes For Side Effect In Discrete Choice Experiments: Risk or Severity-What Is More Important to Hepatitis C Patients? Value in Health. 2015;18(7):A629-A30.
80. Harrison M, Rigby D, Vass C, Flynn T, Louviere J, Payne K. Risk as an attribute in discrete choice experiments: a systematic review of the literature. The Patient-Patient-Centered Outcomes Research. 2014;7(2):151-70.
81. Freeman AM, Herriges JA, Kling CL. The measurement of environmental and resource values: theory and methods: Routledge; 2014.
82. Spinks J, Mortimer D. Lost in the crowd? Using eye-tracking to investigate the effect of complexity on attribute non-attendance in discrete choice experiments. BMC medical informatics and decision making. 2015;16(1):1-13.
83. Regier DA, Watson V, Burnett H, Ungar WJ. Task complexity and response certainty in discrete choice experiments: an application to drug treatments for juvenile idiopathic arthritis. Journal of Behavioral and Experimental Economics. 2014;50:40-9.
84. Flynn TN, Bilger M, Malhotra C, Finkelstein EA. Are efficient designs used in discrete choice experiments too difficult for some respondents? A case study eliciting preferences for end-of-life care. Pharmacoeconomics. 2016;34(3):273-84.
85. Jonker MF, Donkers B, de Bekker-Grob E, Stolk EA. Attribute level overlap (and color coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. Health economics. 2019;28(3):350-63.
86. Veldwijk J, Determann D, Lambooij MS, Van Til JA, Korfage IJ, de Bekker-Grob EW, et al. Exploring how individuals complete the choice tasks in a discrete choice experiment: an interview study. BMC medical research methodology. 2016;16(1):1-11.
87. Ericsson KA, Simon HA. How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. Mind, Culture, and Activity. 1998;5(3):178-86.
88. Aguiar M, Harrison M, Munro S, Burch T, Kaal KJ, Hudson M, et al. Designing discrete choice experiments using a patient-oriented approach. The Patient-Patient-Centered Outcomes Research. 2021;14(4):389-97.
89. Barber S, Bekker H, Marti J, Pavitt S, Khambay B, Meads D. Development of a discrete-choice experiment (DCE) to elicit adolescent and parent preferences for hypodontia treatment. The Patient-Patient-Centered Outcomes Research. 2019;12(1):137-48.
90. McCarthy MC, Lourenco RDA, McMillan LJ, Meshcheriakova E, Cao A, Gillam L. Finding out what matters in decision-making related to genomics and personalized medicine in pediatric oncology: developing attributes to include in a discrete choice experiment. The Patient-Patient-Centered Outcomes Research. 2020:1-15.
91. Sarikhani Y, Ostovar T, Rossi-Fedele G, Edirippulige S, Bastani P. A Protocol for Developing a Discrete Choice Experiment to Elicit Preferences of General Practitioners for the Choice of Specialty. Value in Health Regional Issues. 2021;25:80-9.
92. Vellinga A, Devine C, Ho MY, Clarke C, Leahy P, Bourke J, et al. What do patients value as incentives for participation in clinical trials? A pilot discrete choice experiment. Research Ethics. 2020;16(1-2):1-12.
93. Hollin IL, Craig BM, Coast J, Beusterien K, Vass C, DiSantostefano R, et al. Reporting formative qualitative research to support the development of quantitative preference study

protocols and corresponding survey instruments: guidelines for authors and reviewers. The Patient-Patient-Centered Outcomes Research. 2020;13(1):121-36.
94. Shields GE, Brown L, Wells A, Capobianco L, Vass C. Utilising patient and public involvement in stated preference research in health: learning from the existing literature and a case study. The Patient-Patient-Centered Outcomes Research. 2020:1-14.
95. Hawton A, Boddy K, Kandiyali R, Tatnell L, Gibson A, Goodwin E. Involving patients in health economics research:"The PACTS Principles". The Patient-Patient-Centered Outcomes Research. 2020:1-6.
96. Callender T, Emberton M, Morris S, Pharoah PD, Pashayan N. Benefit, Harm, and Cost-effectiveness Associated With Magnetic Resonance Imaging Before Biopsy in Age-based and Risk-stratified Screening for Prostate Cancer. JAMA network open. 2021;4(3):e2037657-e.
97. Griffin E, Hyde C, Long L, Varley-Campbell J, Coelho H, Robinson S, et al. Lung cancer screening by low-dose computed tomography: a cost-effectiveness analysis of alternative programmes in the UK using a newly developed natural history-based economic model. Diagnostic and prognostic research. 2020;4(1):1-11.
98. Johnson FR, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. Value in health. 2013;16(1):3-13.
99. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. International journal for quality in health care. 2007;19(6):349-57.
100. Pollard K, Donskoy A-L, Moule P, Donald C, Lima M, Rice C. Developing and evaluating guidelines for patient and public involvement (PPI) in research. International journal of health care quality assurance. 2015.
101. Edwards AG, Naik G, Ahmed H, Elwyn GJ, Pickles T, Hood K, et al. Personalised risk communication for informed decision making about taking screening tests. Cochrane database of systematic reviews. 2013(2).
102. Fagerlin A, Zikmund-Fisher BJ, Ubel PA. Helping patients decide: ten steps to better risk communication. Journal of the National Cancer Institute. 2011;103(19):1436-43.
103. Paling J. Strategies to help patients understand risks. Bmj. 2003;327(7417):745-8.