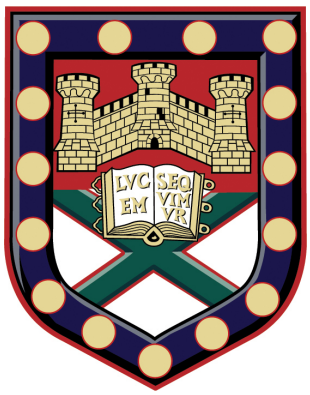
Thrombocytosis: an important marker of cancer in primary care



Sarah Elizabeth Rose Bailey

University of Exeter Medical School

University of Exeter

A thesis submitted for the degree of

Doctor of Philosophy

July 2016

Thrombocytosis: an important marker of cancer in primary care

Submitted by Sarah Elizabeth Rose Bailey to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Medical Studies
July 2016

I certify that all material in this thesis that is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Sarah Elizabeth Rose Bailey
July 2016

Acknowledgements

Without the outstanding supervision of Willie Hamilton, Obi Ukoumunne, and Elizabeth Shephard, the completion of this PhD would not have been possible.

I am sincerely grateful to Willie Hamilton for giving me the opportunity, means, and encouragement to pursue a PhD in the first place, and for his continued supervision throughout the process. His support has extended far beyond the traditional supervisory role and has included allowing me to store all of my worldly possessions in his stables for a month in summer 2015, and teaching me how to drive a tractor. Thank you for making this possible.

I am extremely fortunate to have worked with Obi Ukoumunne, who was my first choice statistician for my supervisory team. Obi, thank you for all of your stats wisdom, your gift for explaining complex statistics in a way I can understand, and for helping me to solve my trickiest Stata puzzles.

Elizabeth Shephard is an invaluable source of help and guidance and has always been on hand to talk things through. Without her, I would not have spent a month in Australia at the end of my second year. Thank you for everything you have done for me - you make our little office a lovely place to work!

I would also like to express my gratitude to Arlene Gallagher, Wilhelmine Meeraus, Grant Lee, and everyone else at the CPRD for providing the study data and follow up support throughout the project. I am also thankful for support from the Policy Research Unit and PenCLAHRC.

The entire DISCO team deserve thanks for many fantastic cake meetings, and for being there to help me whenever I've hit a block. I am incredibly lucky to work with such a friendly group. Sarah W, Sophie, Tani, Sarah P, Jen, and Sal - thank you! I am also grateful to Sal, Willie, Sarah W, and Liz for sharing data from their studies.

Outside of work I am fortunate to be surrounded by some extremely kind, supportive, and hilarious friends and I am grateful for the encouragement and distractions they have provided. Thank you to my team of proof readers: Lisa Price, Helen Rance, James Bailey, and Jon Bailey. Lisa Price - you have been there for me at some of the highest (and lowest!) points of this project. Thank you for all the wine, puppy therapy, and for your voice of reason. To all of the physicists in my life - thank you for encouraging me and telling me that it will all be okay.

My family have been there for me from the start - the very start! I am grateful for their continued love and guidance. Thank you to my parents Alison and Jon for everything they have done for me to get me to this stage. I hope I've made you proud.

Finally, to my husband Tim Starkey - a man who has read every word of this thesis (several times over) and converted my original Word copy into LaTeX (including the tables). You were a serious project threat at the start but as it turns out, you contributed significantly to the successful completion of this PhD. This would have been an entirely different experience without you beside me. Thank you for your encouragement, your support, and for not laughing in my face when I naïvely proclaimed 'I am never going to get crazy over a PhD'. I will be keeping you for the foreseeable future.

Abstract

Thrombocytosis (raised platelet count) has recently been identified as a pre-diagnostic risk marker of cancer; however, the association has not been fully investigated. This thesis aimed to explore the relationship between thrombocytosis and a future diagnosis of cancer through three complementary pieces of research.

Firstly, a systematic review was carried out which aimed to identify studies that had investigated thrombocytosis as a diagnostic marker of cancer. Four case-control studies were identified that had found thrombocytosis to be a significant predictor of lung, oesophago-gastric, uterine, and renal cancer. A further four studies found that thrombocytosis did not predict pancreatic, breast, ovarian, or colorectal cancer. One further study had collected, but not analysed, platelet count data. Data from all nine studies were included in a meta-analysis. The findings of the review suggest that thrombocytosis is a marker of some, but not all, types of cancer.

The second study used data from the Clinical Practice Research Datalink (CPRD) and the English cancer registry. This cohort study examined the relationship between thrombocytosis and cancer using two groups of patients. The first included $40,000$ patients with a raised platelet count (a platelet count of $> 400 \times 10^9$/L). The second cohort included $10,000$ patients with a normal platelet count (150 - 400×10^9/L) who were age, sex, and practice matched to a random quarter of the first cohort. This study found that the risk of cancer was greater in patients with thrombocytosis compared to those with a normal platelet count. The one year cancer incidence was 11.6% (95% CI 11.0 - 12.3) for male patients with thrombocytosis, and 4.1% (95% CI 3.4 - 4.9) in males with a normal platelet count. In female patients, the one year cancer incidence was 6.2% (95% CI 5.9 - 6.5) for those with thrombocytosis and 2.2% (95% CI 1.8 - 2.6) for those with a normal platelet count. Lung and colorectal cancer were more likely to be diagnosed in patients with thrombocytosis than in patients with a normal platelet count, and breast and prostate cancer less likely. In patients with a sustained increase in platelet count over six months, the risk of cancer increased to 18.1% in males (95% CI 15.9 - 20.5) and 10.1% in females (95% CI 9.0 - 11.3). Around a third of patients with lung or colorectal cancer and thrombocytosis had no other symptoms prior to diagnosis that would have prompted investigation for cancer as per current NICE guidance.

The third study compared cancer recording in the CPRD and in the English cancer registry. The aim of this study was to examine the validity of cancer

recording in the CPRD using cancer registry recording as the gold standard, and to estimate predictors of concordance between the two data sources. A sensitivity analysis repeated the primary analysis from the second study to estimate the effect of including unverified CPRD cancer diagnoses. The CPRD identified 5,924 of 7,785 cancers recorded in the cancer registry (sensitivity 76.1%, 95% CI 75.1 - 77.0). 36,255 patients with no record of cancer in the CPRD also had no cancer record in the cancer registry (specificity 97.0%, 95% CI 96.1 - 97.2). 5,924 of 7,028 CPRD cancer diagnoses were confirmed by the cancer registry data; the positive predictive value (PPV) of a CPRD recorded diagnosis was 84.3% (95% CI 83.4 - 85.1). Male cancers, those in younger patients, and those recorded from 2005 onwards were more likely to be recorded in both sources. In a sensitivity analysis, the exclusion of cancer diagnoses that were only recorded in the CPRD did not significantly alter findings from the cohort study described above.

The findings from this thesis show that thrombocytosis is an important predictor of undiagnosed cancer in adults aged 40 years and over. Patients with thrombocytosis are more likely to be diagnosed with lung and colorectal cancer than other types. These results suggest that cancer should be considered as an underlying diagnosis in patients with unexpectedly raised platelets, even if cancer was not suspected at the time that the blood test was ordered. For at least a third of patients with thrombocytosis and cancer, there will be no other clinical features of malignancy; for this proportion, thrombocytosis has great potential to expedite diagnosis and improved survival.

Contents

List of Figures

List of Tables

List of Abbreviations

CI	Confidence Interval
CPRD	Clinical Practice Research Datalink
EMBASE	Excerpta Medical Database
ENCORE	English National Cancer Online Registration Environment
ES	Elizabeth Shephard
FBC	Full Blood Count
GP	General Practitioner
GPRD	General Practice Research Database
ICBP	International Cancer Benchmarking Partnership
ICD	International Statistical Classification of Diseases and Related Health Problems (codes)
ID	Identification
IL-6	Interleukin-6
IOM	Institute of Medicine
IQR	Interquartile Range
ISAC	Independent Scientific Advisory Committee
L	Litre (unit of volume)
LR	Likelihood Ratio
NAEDI	National Awareness and Early Diagnosis Initiative
NCRS	National Cancer Registration Service
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NPC	Normal Platelet Count
NPV	Negative Predictive Value
Ovid/OvidSP	Ovid Technologies Database Collection
PhD	Doctor of Philosophy
PPV	Positive Predictive Value
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
RevMan	Review Manager (software)

TH	Thrombocytosis
TNM	Tumour, Nodes, Metastasis
UK	United Kingdom
WH	Willie Hamilton

Chapter 1

Introduction

1.1 Thesis introduction

This thesis investigates thrombocytosis, or raised platelet count, as a possible risk marker of malignancy in adults, which could be used to identify undiagnosed cancer. This is the first newly discovered marker of cancer in years, one which is commonly measured in primary care as part of a full blood count. Previous primary care-based studies have identified thrombocytosis as a marker of some types of cancer (namely lung, uterine, renal, and oesophago-gastric cancer); secondary care studies have found thrombocytosis to be a marker of poor prognosis; and biological studies have proposed several theories that could explain the link between thrombocytosis and cancer. However, no study to date has comprehensively investigated thrombocytosis across all types of cancer, or examined the relationship within the context of other patient factors. This PhD aimed to address that gap. The work presented in this thesis uses data from the Clinical Practice Research Datalink (CPRD). The accuracy and reliability of the work is only as good as the data source; therefore this PhD also presents a validation study of cancer recording in the CPRD. The structure of the thesis is outlined below, and a brief description of the contents of each chapter is given.

1.2 Aims and objectives

Here, the overall aims and objectives of this PhD are set out. The thesis includes three pieces of work that address two main aims; the first relates to the investigation of the relationship between thrombocytosis and cancer diagnosis. This aim is addressed in two parts: firstly with a systematic review of studies that have examined the association between thrombocytosis and cancer, and secondly with a cohort study using CPRD and cancer registry data. Results from the systematic review feed into and inform the

cohort study.

The second aim takes a methodological perspective. Using data from the cohort study, it provides an essential estimate of the accuracy and validity of cancer recording in the CPRD, and a sensitivity analysis to determine how certain we can be of results using CPRD data.

Overall PhD aims:

1. **To investigate the relationship between thrombocytosis and cancer diagnosis in adults aged 40 years and over.**

Objectives:

i. To carry out a systematic review of studies that have examined the association between platelet count and cancer in a diagnostic context to identify what is currently known about thrombocytosis as a risk marker of cancer.

ii. To examine the incidence of cancer in two cohorts of patients; those with thrombocytosis and those with a normal platelet count, to determine the risk of cancer in each cohort.

iii. To compare the cancer incidence between these two cohorts to determine the absolute increase in risk associated with thrombocytosis.

iv. To examine how the relationship between thrombocytosis and cancer differs across subgroups defined by age, sex, and smoking status.

v. To determine whether some types of cancer are more likely to be diagnosed than others in patients with thrombocytosis compared to patients with a normal platelet count.

vi. To investigate how the risk of cancer changes depending on how the patient's platelet count changes over time.

vii. To investigate the risk of cancer in patients who report symptoms in addition to thrombocytosis.

viii. To investigate the stage at which cancers are diagnosed in patients with thrombocytosis and with a normal platelet count.

ix. To estimate the potential impact of the recognition of thrombocytosis as a marker of cancer in UK suspected cancer guidance by examining the proportion of patients who have thrombocytosis but no other cancer symptoms or markers.

2. To assess the validity of cancer recording in the Clinical Practice Research Datalink used in this thesis, using cancer recording in the English cancer registry as the 'gold standard'.

Objectives:

i. To compare cancer recording in the CPRD and in the cancer registry to determine the level of concordance between the two sources.

ii. To compare the age and sex of patients recorded in both, or either, source.

iii. For cancers recorded in both sources, to compare the date of recording between the two.

iv. To estimate predictors of concordance between the two data sources.

v. To examine the extent to which the inclusion of unverified CPRD-recorded cancer diagnoses causes overestimates in incidence figures from CPRD data, by repeating the primary analysis from Chapter 4 including only cancer registry recorded diagnoses.

1.3 Thesis structure

Chapter 2 - Thesis background

Chapter 2 sets out the background and context for this work, and justification in terms of what is already known about thrombocytosis and cancer. The chapter begins by outlining normal platelet production and function and describing thrombocytosis. It sets out the justification for this work by reviewing the existing evidence supporting the theory of a relationship between thrombocytosis and cancer. This is split in to three areas: primary care studies that have examined thrombocytosis as a diagnostic marker of cancer; secondary care studies that have examined thrombocytosis as a prognostic marker in patients already diagnosed with cancer; and studies of the physiological mechanisms that could underlie the relationship between thrombocytosis and cancer. Although the research presented in this thesis primarily concerns thrombocytosis as a diagnostic rather than prognostic marker, a section on evidence from secondary care is included because it strongly indicates a relationship between platelet count and cancer in a broader sense.

In the second part of the chapter, the context for this thesis is given by providing an overview of the burden of cancer in the UK by reviewing cancer incidence, mortality, and survival. The current issues surrounding primary care cancer diagnosis are considered,

and a section describes how thrombocytosis fits in here and could potentially contribute to the research area and to clinical practice.

In the final part of the chapter, the two data sources used in this PhD, the Clinical Practice Research Datalink (CPRD) and the English cancer registry, are introduced and the accuracy and validity of the CPRD as a data source are explored.

Chapter 3 - Systematic review of evidence of a relationship between thrombocytosis and cancer diagnosis

Chapter 3 presents a systematic review and meta-analysis to address objective i of the first aim of the PhD. The systematic review identifies and critically examines existing research that has investigated the diagnostic potential of the platelet count in patients diagnosed with cancer. The systematic review determines what evidence already exists concerning the effects of age, sex, and change in platelet count over time on the relationship. The types of cancer diagnosed in patients with thrombocytosis are examined to inform the analyses presented later in this thesis. The key findings from the systematic review which partly informed the cohort study presented in Chapter 4 are described at the end of the chapter.

Chapter 4 - The association between thrombocytosis and cancer

Chapter 4 begins with an introduction to the cohort study on the association between thrombocytosis and cancer, and a description of the methods used. The main body presents the results from the analyses which address objectives ii to ix of the first aim of the thesis. The characteristics of the cohort are examined and compared to the UK general population. The one year cancer incidence for patients with thrombocytosis and patients with a normal platelet count is presented. The effects of age, sex, and smoking status on the relationship are examined, and the time interval between blood test results and cancer diagnosis is calculated for patients in each of the two cohorts. The risk of cancer is estimated for patients with thrombocytosis plus symptoms indicative of malignancy. The impact of changes in platelet count over time on the risk of cancer is examined. The potential impact of this work on cancer diagnosis in the UK is estimated by examining the proportion of cancer patients who had thrombocytosis, but did not have symptoms that warranted investigation for cancer under the UK suspected cancer guidance for clinicians.

Chapter 5 - CPRD validation study and main study sensitivity analysis

Chapter 5 presents the results of a study to examine the validity of cancer diagnosis recording in the CPRD, to address the second aim of the PhD. This is done by comparing cancer recording in the CPRD to recording in the cancer registry; considered the 'gold standard' of cancer diagnosis registration. A sensitivity analysis is reported to judge the reliability of the results presented in Chapter 4, given the results of the CPRD validation study.

Chapter 6 - Discussion and conclusions

In Chapter 6, the results from the three main pieces of work in this thesis are drawn together to address the overall aims of the thesis and to draw conclusions. The strengths and limitations are critically examined. The implications of the findings for research and for clinical practices are presented. Finally, future directions and recommendations for ongoing research are discussed.

1.4 Chapter summary

This chapter has briefly outlined the aims and objectives of this thesis, and has outlined the content that will be presented in each of the subsequent chapters. The next chapter presents the justification for this work in terms of what is already known about the association between thrombocytosis and cancer, and where this work fits in. The context is set for the work in terms of the burden of cancer in the UK and the issues surrounding the diagnosis of cancer in primary care, and the two data sources used in the final two studies are presented.

Chapter 2

Thesis background

2.1 Chapter introduction

In this chapter, the justification and context for this thesis is presented. This begins in Section 2.2 with an exploration of normal platelet function and the physiological processes underlying thrombocytosis, including usual causes. The proposed biological mechanisms underlying the thrombocytosis-cancer relationship are explored. Although the main focus of this thesis is thrombocytosis as a risk marker of cancer within a primary care setting, examining the physiological theories supporting the association is contextually valuable in providing useful background knowledge and understanding to support the clinical evidence. Following this, Section 2.3 examines the burden of cancer in the UK in terms of incidence, survival, and mortality. Cancer outcomes in the UK are compared to those in other countries, and reasons for the differences in outcomes are explored. The issues surrounding cancer diagnosis in UK primary care are discussed within the context of earlier diagnosis, where the main body of work from this thesis aims to have an impact. Finally, Section 2.5 introduces the two data sources used in this PhD study: the Clinical Practice Research Datalink (CPRD) and the UK cancer registry, and the strengths and limitations of these are discussed.

2.2 Platelets, thrombocytosis, and their link with cancer

2.2.1 The roles of platelets

Platelets are small, anucleate cells which circulate in the blood in an inactive state. They are activated by a number of factors relating to their roles in haemostasis, most notably in the clotting process when membrane damage is detected, and in inflammatory response (Daly, 2011). A normal platelet count in an adult is in the range of 150

- 400×10^9/L (Biino *et al.*, 2013; Giles, 1981). The platelet count is on average higher in women than in men, most likely mediated by hormonal differences; it declines with age (Bain, 1996; Daly, 2011). When cellular or epithelial lining damage is detected, platelets are involved in the subsequent cascade of cell activity that results in a mesh being formed over the damaged area, and a thrombus forming to allow the membrane to heal (Gay & Felding-Habermann, 2011). Platelets are also involved in inflammatory processes, but their exact role is the subject of ongoing research and debate. Upon activation, platelets release a number of pro-inflammatory cytokines and express cell surface receptors which are activated by key factors in the inflammatory process (Herter *et al.*, 2014). This process was summarised recently in Nature Immunology (Mantovani *et al.*, 2008).

2.2.2 Thrombopoiesis

Thrombopoiesis, the process of platelet production, begins with megakaryocytes (Deutsch & Tomer, 2006). These cells are found primarily in the bone marrow, but have also been isolated from secondary lymphoid tissue. The process of platelet formation is not fully understood, but it is known to start with megakaryocytes responding to thrombopoeitin produced in the liver (Kuter, 1996). Upon stimulation by thrombopoietin, megakaryocytes generate the organelles needed to form platelets within their matrix. The megakaryocytes extend long processes known as 'proplatelets' through the cell wall of the bone marrow into capillaries that lie alongside the marrow (Machlus & Italiano, 2013). Internal cellular structures are assembled within the megakaryocytes and along the proplatelets to carry the component parts needed for platelet formation to the tip of the structure. In response to shear forces within blood vessels, these proplatelets break off into the blood circulation (Geddis, 2009). These are then known as preplatelets. Preplatelets are barbell shaped and once within the blood, they split to become two platelets. The exact cause of proplatelet formation is unknown, but it is believed that megakaryocyte migration during proplatelet formation is driven by differences in chemical gradients between the osteoblastic and vascular niches.

The process of platelet production is believed to happen in locations all over the body but there is evidence to suggest that the lungs may be the site of particular significance. Megakaryocytes are found at higher concentration in central venous arteries compared to peripheral arteries. Higher platelet counts are found in post-pulmonary blood vessels compared to pre-pulmonary (Zucker-Franklin & Philipp, 2000). Rat models with lung damage were found to have decreased platelet counts (Xiao da *et al.*, 2006). There is also evidence from human studies to support this hypothesis. Low platelet count (thrombocytopenia) is common in patients with lung disease, with the degree of

reduction correlating to the severity of disease (Xiao da *et al.*, 2006) and thrombocytosis has been identified as an early marker of lung cancer (Hamilton *et al.*, 2005a).

Thrombopoeitin is the strongest influence on thrombopoiesis. Knockout of thrombopoeitin or thrombopoeitin cell surface receptors on megakaryocytes leads to low levels of platelets, but does not eliminate them altogether (Kaushansky, 2005). This suggests that although thrombopoeitin is a strong mediator of platelet production, it is not the only influencing factor. Up to 60% of thrombopoeitin is produced in parenchymal cells in the liver, but it is also produced in the kidneys and in bone marrow. The production of thrombopoeitin is strongly influenced by interleukin-6 (IL-6), the number of circulating platelets in the blood, and other factors (Kaushansky, 2005). IL-6 is an inflammatory cytokine which has been linked to malignancy (summarised in Schafer & Brugge (2007)); IL-6 is released by tumours in a number of cancer types including lung (Gao *et al.*, 2007), colorectal (Waldner *et al.*, 2012), and ovarian (Offner *et al.*, 1995). IL-6 has been linked to tumour stem cell renewal (Sansone *et al.*, 2007), and elevated levels of IL-6 are associated with poorer prognosis in several types of cancer including myeloma, lymphoma, ovarian, prostate, and renal. (Hong *et al.*, 2007)

2.2.3 Thrombocytosis

Thrombocytosis is a condition in which the platelet count exceeds the normal range ($> 400 \times 10^9$/L). Thrombocytosis can be classified as primary or secondary; primary thrombocytosis occurs as a result of a genetic or myoproliferative disorders, whereas secondary or reactive thrombocytosis occurs as a result of another underlying condition (Schafer, 2002). Most commonly, thrombocytosis is secondary in response to another underlying disorder. These disorders include anaemia, acute blood loss, inflammatory and infectious conditions, and recovery from low platelet count. A platelet count of over $1,000 \times 10^9$/L is usually associated with primary thrombocytosis; commonly myoproliferative disorders or splenectomy (Griesshammer *et al.*, 1999).

2.2.4 Diagnosing thrombocytosis

Thrombocytosis is rarely a diagnosis in itself but is more commonly used as an indicator of underlying disease. The platelet count is one component of the full blood count (FBC); this involves the measurement of red blood cell numbers and cellular characteristics, white blood cells, and platelets in the blood (Cancer Research UK, 2015a). Clinically, such blood tests are carried out for a wide range of reasons, either to aid with diagnosis or routinely to check general health. The platelet count *per se* is rarely used to diagnose an underlying disease in practice but can be used with other symptoms and test results to determine the likelihood of a disease being present. Blood tests are rarely

carried out with the sole intention of measuring platelet count, and thrombocytosis is often discovered as an incidental finding (Bleeker & Hogan, 2011; Khan *et al.*, 2009).

2.2.5 Evidence supporting a link between thrombocytosis and cancer

The link between platelet count and cancer has long been acknowledged (Levin & Conley, 1964). In this early study, Levin and Conley examined 82 patients with thrombocytosis within a hospital setting (already selected for investigation) and found that 31 (38%) had an underlying cancer. They subsequently examined platelet counts in 268 patients with cancer of any type, and found that 40% had thrombocytosis. A number of primary and secondary care-based studies have found evidence to support this link, although few studies have investigated thrombocytosis as a predictor of cancer, and only some cancer sites have been studied. In secondary care, thrombocytosis is an established marker of poor cancer prognosis, although the relationship between thrombocytosis and prognosis varies by cancer site (Hauser *et al.*, 2006). There is also a range of research exploring the biological theories for the processes that underlie the thrombocytosis-cancer association. The three key bodies of evidence from primary care, secondary care, and biological studies are summarised here.

1. Evidence from primary care studies

A number of case-control studies have aimed to identify the early markers of certain types of cancer and to quantify the risk of cancer in patients with these markers, to enable general practitioners to direct patients to appropriate investigative services. In these studies, 'early markers' are defined as symptoms that patients subsequently diagnosed with cancer report to their GP in the year prior to their cancer diagnosis. The systematic review presented in Chapter 3 of this thesis found four matched case-control studies that had identified thrombocytosis as a significant predictor of cancer in multivariable models. (Hamilton *et al.*, 2005a,b; Shephard *et al.*, 2013; Walker *et al.*, 2013). A further four studies were identified in which thrombocytosis did not predict cancer (Hamilton, 2009; Stapley *et al.*, 2012, 2013; Walker *et al.*, 2014). One study had collected platelet count data but did not include thrombocytosis in any analysis, as was present in less than 5% of bladder cancer cases (Shephard *et al.*, 2012). Only features present in more than 5% of cases were included in the analysis in that study.

A study in 2012 examined paraneoplastic thrombocytosis in 619 patients with ovarian cancer (here, thrombocytosis was defined as a platelet count over 450×10^9/L) (Stone *et al.*, 2012). Although that study was set in secondary care, it reported that 31% of patients had thrombocytosis at the time they were diagnosed. The median platelet count was 542×10^9/L in patients with thrombocytosis, and 318×10^9/L in

patients without thrombocytosis. Patients with ovarian cancer and thrombocytosis also had shorter survival and more advanced disease at diagnosis than those without thrombocytosis. Stone *et al.* (2012) also investigated the mechanisms underlying the association between platelets and ovarian cancer using mouse models. Evidence presented in the same study points to excess tumour-derived IL-6 promoting thrombopoeitin production, which in turn increases platelet production and number which supports tumour growth and development.

2. Evidence from secondary care studies

Although evidence of a post-diagnosis association between platelet count and cancer is of little use in judging the value of thrombocytosis in detecting malignancy, it does strengthen the theory that a link between cancer and platelet count exists. It can also be useful to determine which cancer sites may be linked with thrombocytosis, and which may not. Thrombocytosis has been found to be a marker of poor prognosis in several cancers (Buergy *et al.*, 2012) including gastrointestinal (Ikeda *et al.*, 2002; Voutsadakis, 2014), lung (Aoe *et al.*, 2004; Pedersen & Milman, 1996), ovarian (Allensworth *et al.*, 2013; Stone *et al.*, 2012), endometrial (Gücer *et al.*, 1998; Njølstad *et al.*, 2013), bladder (Todenhöfer *et al.*, 2012), and renal (Suppiah *et al.*, 2006) cancers.

The study by Pedersen & Milman (1996), although set in secondary care, examined 126 patients who were admitted for further investigation for cancer following an abnormal chest x-ray. Thrombocytosis was observed in 35 of 61 patients (57%) who went on to be diagnosed with lung cancer and only 5 of 65 patients (8%) who did not.

A study in 1972 examined the clinical utility of thrombocytosis in predicting malignancy in 100 consecutive patients of any age (60 female and 40 male) seen in hospital for a variety of reasons (Davis & Mendez Ross, 1972). In this study, thrombocytosis was defined as a platelet count of $> 500 \times 10^9$/L. 36 of 100 patients with thrombocytosis (36%) were found to have cancer; the most commonly diagnosed types were ovarian ($n = 7$), lymphoma ($n = 7$), breast ($n = 5$), lung ($n = 4$), and colon ($n = 4$). The secondary care setting of this study, in which patients have already been selected for investigation, means that a greater proportion of patients are likely to have cancer than in the primary care population. Patients are also more likely to have another condition which would mean that thrombocytosis is an expected finding. In this sample, there were 19 such patients; either post-surgery or post-splenectomy, and a further eight with myoproliferative disorders. Therefore, thrombocytosis could be considered 'unexpected' in the remaining 75 patients. All 36 cancers were in this subgroup; so the cancer incidence in patients with unexpected thrombocytosis was 48%.

A cross-sectional survey of 1,007 patients enrolled in the Vermont Diabetes Information System investigated the association between anti-platelet drug use and history of

cancer (Holmes *et al.*, 2010). 50% of included patients were using anti-platelet drugs; these were associated with a significant reduction in self-reported cancer diagnoses (odds ratio 0.66, 95% CI 0.44-0.99; $p = 0.045$). This model was adjusted for age, sex, body mass index, co-morbidities, and other medications. Although self-reported data are considered low quality evidence, this supports the theory of a relationship between cancer and platelets as anti-platelet drugs appeared to have an anti-cancer effect, in patients with diabetes. This study did not distinguish between type I and type II diabetes, diagnoses were unconfirmed, and there were no data on cancer site.

3. Evidence from biological studies

Despite the large body of evidence from primary and secondary care studies supporting the theory of a relationship between platelets and cancer, the exact mechanisms perpetrating the association is unknown. There are three main theories that could explain the relationship (Buergy *et al.*, 2012):

- Platelets augment tumour growth
- Platelets promote tumour metastasis
- Tumour enhances platelet production

The first two of these theories are supported by findings from secondary care studies showing that thrombocytosis is associated with poorer survival outcomes in patients diagnosed with cancer. They imply that a raised platelet count is occurring independently of cancer and promoting its development. The last of the three is strongly supported by findings from primary care studies that thrombocytosis is a marker of an undiagnosed cancer; it implies that the malignancy precedes and causes the subsequent rise in platelet count, whether directly or indirectly. In a review published in 2012, Buergy *et al.* (2012) present these three theories and suggest that all are likely to be occurring in a cancer patient in a cycle of platelet production and cancer development. The evidence supporting each of these theories is summarised below.

Platelets augment tumour growth:
The process of platelet production involves a number of cells and factors acting in cascade (Deutsch & Tomer, 2006), described in brief earlier in this chapter, throughout which platelets secrete various cytokines. This theory purports that thrombocytosis occurs independently of cancer in response to another condition, and that proangiogenic cytokines secreted by platelets as part of the body's response to that condition promote the development of blood vessels in the growing tumour (Jain *et al.*, 2010). Therefore, patients with independently occurring thrombocytosis and cancer will have

poorer prognosis as their high number of circulating activated platelets promotes cancer development and hastens disease progression (Buergy *et al.*, 2012).

Platelets promote tumour metastasis

Similarly, in this second theory, another unrelated condition independent of malignancy is causing thrombocytosis in patients with cancer. There are three ways platelets could be promoting metastasis: 1) by protecting and stabilising circulating tumour cells from the body's immune response (Borsig *et al.*, 2001); 2) by stimulating tumour cell proliferation (Buergy *et al.*, 2012); 3) by promoting tumour cell angiogenesis (Gay & Felding-Habermann, 2011). This theory is supported by evidence of thrombocytosis being associated with metastases and more advanced disease in secondary care studies (Yu *et al.*, 2015).

Tumour enhances platelet production

This theory suggests that tumours secrete a number of factors which interfere with and promote platelet production. The primary factor believed to be involved in this is interleukin-6 (IL-6), which is involved in inflammatory processes and platelet production (Scheller *et al.*, 2011). As described earlier in this introduction, the primary mediator of platelet production is thrombopoietin; although this is not directly produced by tumour cells, its production is mediated by a number of cytokines that include IL-6 (Kaser *et al.*, 2001). Elevated IL-6 levels have been found in a number of cancers including gastro-intestinal (De Vita *et al.*, 2001), renal (Blay *et al.*, 1992), prostate (Nakashima *et al.*, 2000), ovarian (Plante *et al.*, 1994), and lung (Takeuchi *et al.*, 1996). Elevated IL-6 has also been found to be associated with poorer outcomes and later disease stage in cancer patients (Hong *et al.*, 2007), but it is not clear whether this is a direct effect or mediated by increased platelet count.

2.2.6 Platelet activation in cancer

As described previously, platelets circulate in the blood in an inactive state, and are activated in response to various stimuli. The platelet membrane contains glycoprotein receptors which, when stimulated, result in the activation and aggregation of platelets, usually in response to tissue or blood vessel damage. Several studies have found advanced cancer patients to have increased levels of activated platelets in their circulation (Boneu *et al.*, 1984). Mechanisms of tumour-related platelet activation are likely to involve tumour-derived thrombin, which stimulates tumour cell growth in addition to platelet activation; adenosine diphosphate, another tumour-derived factor which activates platelets; and direct activation of platelets through contact with tumour cell surface. Interestingly, thrombocytosis has been found to decline following solid tumour

resection (Nash *et al.*, 2002) lending further evidence to the theory that solid tumours increase platelet production.

2.2.7 Section summary

In the first section of this chapter, platelet production and function were described and the background evidence from primary care, secondary care, and biological studies supporting a link between platelet count and cancer has been presented. Evidence from primary care is promising but insufficient to draw conclusions about the nature of the relationship in a clinical setting; no studies have fully investigated the significance of the relationship for all cancer sites, or investigated how the relationship might vary with age, sex, or other factors. It is here that this PhD aims to contribute to the body of knowledge and have clinical utility. Secondary care evidence, although prognostic rather than diagnostic, clearly supports a link between platelet count and cancer, as does the evidence from biological studies. From a biological perspective, there is still insufficient evidence to support a single unifying theory of the processes mediating the association between platelet count and cancer; it is possible that elements of each of the proposed theories could be correct.

2.3 Cancer burden and cancer diagnosis in the UK

In the second section of this chapter, the burden of cancer in the UK is evaluated in terms of incidence, survival, and mortality. UK cancer outcomes are compared to those in other countries, and the reasons for differences are explored. The diagnostic and primary care factors influencing cancer outcomes in the UK are discussed, and the causes of delays in the diagnostic process are examined. This section also considers where the work presented in this thesis fits in and potentially advances the area.

2.3.1 UK cancer incidence

Cancer incidence statistics in the UK are compiled by the Office of National Statistics, Cancer Research UK, the Welsh Cancer Intelligence and Surveillance Unit, and the Information Services Division in Scotland. Since the late 1970s, cancer incidence has risen steadily in England - by 35% in females and by 15% in males. Similar figures are seen in Wales and Scotland; between 2004 and 2013 cancer incidence increased 12% in Wales. There were 352,000 newly registered cancer diagnoses in England, 19,000 in Wales, and 42,000 in Scotland in 2013 (Cancer Research UK, 2014b; Information Services Division Scotland, 2014; Welsh Cancer Intelligence and Surveillance Unit, 2014). Across all three regions, the greatest incidence is in adults in their late 60s and early

70s. Although an overall increase in cancer incidence is evident throughout the UK, the change in incidence varies considerably by the type of cancer diagnosed; for example, breast cancer incidence has risen 9% in the last decade in Scotland (although this could be an artefact of increased screening), whereas prostate cancer incidence has remained constant in the same time period. Lung cancer incidence in males has decreased in the last decade by 15%, whereas it increased in females in the same time period by 13%; this may reflect changes in the epidemiology of smoking during this time. The four most commonly diagnosed cancers, accounting for around a half of all new cases in the UK, are breast, prostate, lung, and colorectal cancer. In Scotland, the incidence of these types of cancer have decreased 4% for men and increased 7% for women in the last decade.

Changes in cancer incidence over time also vary by age and sex: the greatest increases in incidence over the last ten years are in adults aged 65-74 years, and the age-specific cancer incidence increases more sharply in men than in women in those aged 70 years and over. In this age group, incidence rates are over 50% higher in men than in women. In England, the majority of new cancer cases recorded between 2011 and 2013 were diagnosed in those aged 50-74 years, but the greatest incidence is still in those aged over 75 years. This is partly due to age being a risk factor for cancer, and partly due to the larger number of people in this age group.

The steady increases in cancer incidence observed over the last few decades can be attributed to several factors; although there has been an increase in risk factors associated with cancer such as alcohol consumption and obesity, there is also greater awareness of early cancer symptoms and uptake of screening programmes in the general public due to public health campaigns, and a decrease in smoking.

2.3.2 UK cancer mortality and survival

Cancer of any kind is still the leading cause of death in England and Wales (Office of National Statistics, 2015) although in the last decade, the number of deaths attributed to cancer in England has fallen by about 10%. In 2013, there were approximately 160,000 deaths from cancer; the four most common cancers (breast, prostate, colorectal, and lung) account for around half of these. Three quarters of cancer deaths were in those aged 65 and over (Cancer Research UK, 2014b). In Wales, cancer mortality is similar for men and women up to the ages of 55 to 59 years, after which it is greater in men than in women (Welsh Cancer Intelligence and Surveillance Unit, 2014). There were around 15,000 deaths from cancer in Scotland in 2014, and lung cancer accounted for around a quarter of these (Information Services Division Scotland, 2014). Mortality has decreased in Scotland in the last ten years in men (14.9% decrease) and women

(5.7% decrease). Most types of cancer show a decrease in mortality in recent years, with the exception of liver cancer, for which mortality has increased 42% in males and 44% in women in the last decade.

Overall, improvements in cancer survival have outweighed increases in incidence and this means that, for most cancers, mortality is improving. In England, survival for all types of cancer has doubled in the last decade, and in 2014 the ten year survival rate was 50%. Although survival is improving year on year, there is still a discrepancy in survival between different cancer types. For men, this ranges from a ten year survival rate of 1% for pancreatic cancer to 98% for testicular cancer. Survival is generally higher in women than in men, and mostly poorer with older age, although the four most common types of cancer have their best survival rates in middle aged adults (Cancer Research UK, 2014a). Survival has also shown a consistent improvement over the last decade in Wales, and is also higher in women than in men in this principality. In Scotland, survival is also increasing; all cancer five year survival based on data from 2007 to 2011 is 54% for women and 48% for men. Here, pancreatic cancer has the poorest survival (3.6% in men and 5.5% in women) but at the other end of the scale, testicular cancer has a 93% five year survival rate.

2.3.3 Comparing UK cancer survival to other countries

Whilst UK cancer survival is improving, data from the EUROCARE studies shows that there are still substantial differences in survival rates between the UK and the rest of Europe, with the UK performing poorly (Berrino *et al.*, 2007; De Angelis *et al.*, 2013; Quaresma *et al.*, 2015; Thomson & Forman, 2009; Walters *et al.*, 2015).

The International Cancer Benchmarking Partnership (ICBP) (Butler *et al.*, 2013) was established between Australia, Canada, Denmark, Norway, Sweden, and the UK to estimate survival trends for selected types of cancer, and to investigate the causes of differences in these rates between countries. The ICBP has five modules: epidemiology; population awareness and beliefs; beliefs, behaviours and systems in primary care; root cause of treatment and diagnosis delays; and exploration of early deaths.

The first publication from the ICBP in 2011 reported on population-based trends in survival for colorectal, lung, breast, and ovarian cancer between 1995 and 2007. The study found that, although there were improvements in survival in all countries during this time, the one year survival for lung, colorectal, and ovarian cancer was consistently poorer in the UK and Denmark compared to the other included countries (Coleman *et al.*, 2011). Although surgical and treatment differences were considered as potential causes of this variance, the paper concluded that late stage diagnosis in the UK and Denmark compared to the other countries was the root cause of much of the variation.

An updated comparison of cancer survival in England and Australia, Canada, Norway, and Sweden was published four years later (Walters *et al.*, 2015). This study found that, although survival in England improved from 1999-2005, it was still lower than in the other countries included in the study. Abdel-Rahman *et al.* (2009) used EUROCARE data to estimate the avoidable premature cancer mortality in the UK as a result of poorer survival in this country compared to better survival rates seen in Europe. The study estimated that 6-7% of cancer mortality in the UK is 'avoidable' in the sense that it would not exist if UK survival was as good as the best in Europe. Richards (2009) estimated that at least 5,000 deaths within five years of a cancer diagnosis could be prevented annually in England with better diagnostic and treatment services.

Differences in data reporting may account for some of the variation between countries (Butler *et al.*, 2006) but the main factors identified as reasons for the differences in survival are differences in patient awareness and attitudes (Forbes *et al.*, 2013); primary care (Rose *et al.*, 2015); and stage-related differences (Maringe *et al.*, 2012, 2013; Walters *et al.*, 2013a,b). Lung and colorectal cancer, for example, are generally diagnosed at a later stage in the UK than in the rest of Europe and their survival is generally poorer. Although we cannot infer causation from this correlation, this along with other evidence presented here strongly suggests a link between the stage at which cancer is diagnosed and patient survival.

Overall, later stage diagnosis in the UK is believed to be one of the key contributory factors to poor disease outcomes. Reducing the interval from symptom reporting to diagnosis is generally accepted to be a key strategy to identify cancers at an earlier stage of disease progression, which in turn should result in better outcomes, achieving the ultimate aim of reduced premature cancer mortality and improved survival. However, the evidence base to support these ideas is varied and dominated by observational study design. Differences in how studies define and measure delay and outcomes, combined with differences in how types of cancers act (both different sites and different cancers within the same primary site) can make it difficult to compare and contrast different studies.

2.4 Primary care cancer diagnosis in the UK

As the majority of cancers are diagnosed in patients who present to their general practitioners (GP) with symptoms, primary care is a crucial setting for diagnosing cancer. In the UK, 98% of the population are registered with a NHS primary care practice. GPs are the first point of contact for most health issues and act as gatekeepers to secondary care services, accessed through referrals. Each patient has an individual electronic medical record and a unique NHS identification number. Patient data are

routinely collected in practice, including consultations with GPs, signs and symptoms of disease, tests and immunisations and any prescribed medications. Data on diagnoses and treatments are fed back to practices from secondary care.

In a retrospective analysis of primary care consultations, the crude annual consultation rate per person was found to be 5.16 consultations per year (Hobbs *et al.*, 2016). Consultation rates were found to increase with age, and generally women consulted more often than men. Independently calculated ratios using data presented in Hobbs *et al.* (2016) show that the male to female consulting ratio increases from 0.71 in those aged 45 to 64 years to 0.93 in those aged 65 to 74 years; 0.96 in those aged 75 to 84 years; and 1.01 in those aged 85 years and over. This is supported by a cohort study from 2013 which examined gender differences in consulting and found that the crude consultation rate was 32% lower in men compared to women of all ages, but the consultation rate was more similar between men and women in older age. In those aged 40 to 57 years, the rate ratio of male to female consultations was 0.62 (0.62-0.63) and in patients aged over 58 years, the rate ratio narrowed to 0.92 (0.91-0.93) (Wang *et al.*, 2013).

2.4.1 The difficulties of diagnosing cancer in primary care

Although the majority of cancers are diagnosed in this setting, cancer is still a relatively rare diagnosis for a general practitioner. Identifying symptomatic cancer is difficult because even the most well-known 'alarm' symptoms have low relative predictive values and are most likely to be caused by benign disease. Most cancer symptoms are vague, such as feeling tired or weight loss. For GPs, identifying which patients have undiagnosed cancer from vague, common symptoms is a challenge. Identifying the symptoms, or combination of symptoms, that are most likely to be indicative of underlying malignancy can aid GPs in deciding which patients to refer to further investigation.

2.4.2 Policy and guidance for diagnosing cancer in general practice

In the UK, the guidance underlying the investigation and diagnosis of suspected cancer in primary care is developed by the National Institute for Health and Care Excellence (NICE). The current version of the guidance (Suspected cancer: recognition and referral, NG12), an update of the 2005 version, was published in 2015. The evidence-based recommendations advise clinicians on the appropriate course of action depending on the patient's symptoms; any patient with a symptom profile with a greater than 3% chance of cancer is to be referred for further investigation (NICE, 2015).

The Department of Health's Cancer Reform Strategy (2007) stated that more work was needed to enable earlier diagnosis in symptomatic patients, as a key strategy to

improve cancer outcomes in the UK and bring them in line with cancer survival in other countries. The National Awareness and Early Diagnosis Initiative (NAEDI) was launched by the UK Department of Health and the charity Cancer Research UK in 2008 in response to the Cancer Reform Strategy, with the aim of improving earlier diagnosis of cancer through the coordination and support of activities and research to improve survival and reduce mortality (Richards, 2009). In 2011, the Department of Health's 'Improving outcomes: a strategy for cancer' policy stated that avoidable deaths from cancer in the UK would be prevented 'mainly through earlier diagnosis' (Department of Health, 2011).

The NAEDI pathway shows factors that influence cancer outcomes and survival. Any one of these features can result in a delay in the pathway to diagnosis (Hiom, 2015) (see Figure 2.1). The work presented in this thesis aims to influence and improve the 'delays in primary care interval' box (2); the interval between patients presenting to their GP and being referred to hospital services for diagnostics, by identifying a new risk marker of cancer that could prompt GPs to suspect cancer (and so refer) sooner. The primary care interval forms part of the diagnostic interval; the time from first symptomatic presentation of the patient to diagnosis (Weller *et al.*, 2012).

2.4.3 Is earlier diagnosis better?

Earlier diagnosis is hailed as the key to improving cancer outcomes for the UK, through the identification of cancers at an earlier and therefore more treatable disease stage. Earlier diagnosis is sometimes used interchangeably to mean either diagnosing cancer at an earlier stage in the progression of the disease, or to mean a shorter diagnostic interval (shorter time from first symptomatic presentation of the patient to diagnosis). The evidence for whether earlier diagnosis is better within each of these two contexts is summarised below.

2.4.3.1 Earlier diagnosis: earlier disease stage

Previous research has examined the causes of delays in diagnosis and developed strategies to promote earlier diagnosis. It is generally assumed that diagnosis at an earlier stage will improve cancer outcomes. However, the evidence supporting this assumption is mixed, and there are potential harms from earlier diagnosis which should be considered. It can be difficult to separate the true effects of earlier diagnosis from confounding and bias; particularly symptom lead time bias (whereby patients diagnosed with symptoms at an earlier stage appear to have greater survival time because of the earlier diagnosis date; they may not actually survive longer than they would have if they had been diagnosed later).

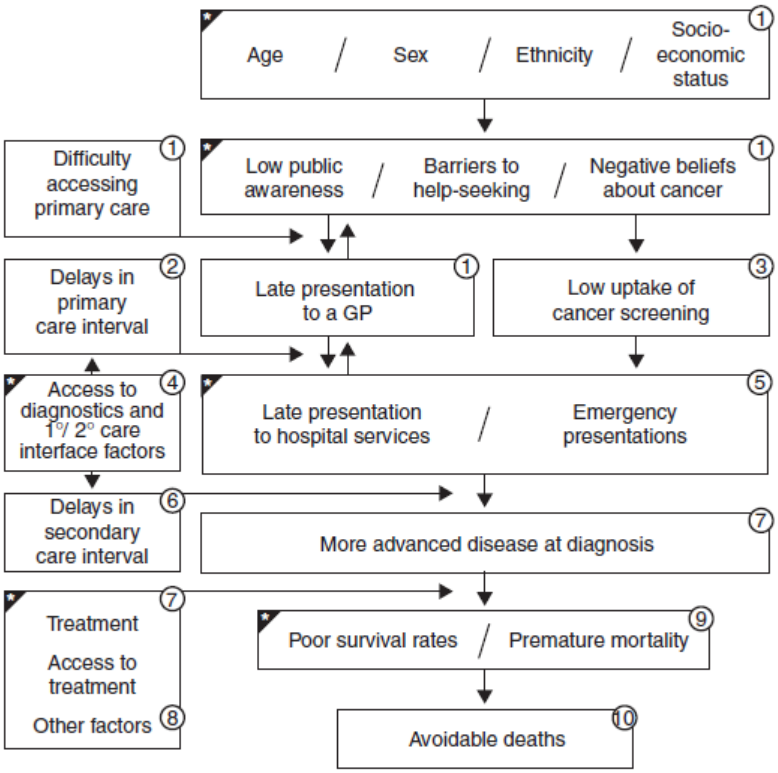


Figure 2.1: Factors influencing cancer survival and premature mortality (Hiom, 2015).

Studies that have investigated the causes of poorer cancer outcomes in the UK compared to the rest of Europe (described above) have concluded that late stage diagnosis is one important contributing factor. Certainly, late stage diagnosis is associated with poorer outcomes, and there is evidence to suggest that lower rates of investigation for cancer result in poorer survival: Møller *et al.* (2015) found low use of urgent referral pathways in GP practices to be associated with increased risk of death in cancer patients (hazard ratio 1.07; 95% CI 1.05-1.08). A survey of GPs in countries involved in the ICPB found increased diagnostics to be associated with better outcomes; there was strong correlation between readiness to investigate for cancer and cancer survival rates (Rose *et al.*, 2015). Although there could be other explanations for this association, such as greater cancer awareness in GPs who investigate more often, or the influence of ascertainment bias (as the population under study is not representative of the general population, being at greater risk of disease), overall there is strong evidence that greater investigation for suspected cancer is associated with better survival. Localised, smaller, earlier stage cancers are easier (and less expensive) (Insicive Health, 2014) to treat than more advanced disease. Patients want earlier diagnosis (and want investigation for cancer at lower risk thresholds when faced with hypothetical situations), as do other stakeholders including the government and the media (Banks *et al.*, 2014). There are harms associated with later diagnosis; poorer survival and prognosis, greater morbidity from the disease and from treatment, and greater psychological trauma for patients and their families.

However, there are other considerations which mean that earlier diagnosis may not always be better. Earlier diagnosis could result in the identification of small, slow-developing cancers that may not have caused any additional morbidity or mortality over the patient's life course (this is known as over-diagnosis) (Marcus *et al.*, 2015). This will however result in the patient being subjected to otherwise unnecessary treatment. It is not always clear whether survival actually improves with earlier diagnosis or whether improved survival time is an artefact of symptom lead time bias. The issue is further confused by the fact that more aggressive, late stage cancers presenting with clearly identifiable symptoms will often have a very short diagnostic interval, and have poor outcomes and short survival time; this is known as the waiting time paradox, defined by Crawford *et al.* (2002).

2.4.3.2 Earlier diagnosis: shorter diagnostic interval

The previous section described earlier diagnosis in terms of aiming to diagnose patients at an earlier disease stage. Earlier diagnosis can also refer to a diagnosis happening more quickly after symptomatic presentation, reducing the time known as the diagnostic

interval. The diagnostic interval is closely related to stage at diagnosis and also has an impact on cancer outcomes. It could be assumed that a shorter diagnostic interval, and therefore shorter time to treatment, would mean better survival but the evidence surrounding this issue is mixed.

In a cohort study in 2002, the relationship between treatment delay and survival was explored for 703 women with endometrial cancer (Crawford *et al.*, 2002). The study found delayed time to treatment to be inversely related to survival; patients with the longest wait for surgical treatment survived significantly longer than those with a shorter time to wait. This may be the result of more serious, advanced disease cases being prioritised for treatment over less serious cases. Although this study focusses on the time from GP referral to surgical treatment, the concept of the waiting time paradox has been applied in other studies. Rupassara *et al.* (2006) studied this effect further in 154 patients with colorectal cancer. In this study, the interval of interest was the time from GP referral to a diagnosis of cancer being made. In a stratified analysis, 44 patients with a longer diagnostic interval (more than or equal to 50 days from referral to diagnosis) had a 93.7% 5-year survival rate; 110 patients with a shorter diagnostic interval (less than 50 days from referral to diagnosis) had a 5-year survival rate of 65.3%. The longer diagnostic interval group also had smaller tumours, more low-risk symptoms, and more early stage disease than patients with a shorter diagnostic interval.

2.4.3.3 The association between diagnostic interval and cancer outcomes

The relationship between longer diagnostic interval and poorer cancer outcomes has been investigated in a number of observational studies. It can be difficult to compare or combine results from studies of this type due to heterogeneity in measures of 'diagnostic interval'; in a systematic review of diagnostic intervals discussed later in this section, 15 different definitions for diagnostic interval were used in 177 articles. Studies also use varying measures of outcomes.

Neal *et al.* (2015) carried out a systematic review to assess the evidence for whether a more timely diagnosis improves cancer outcomes. Of the 1,036 full text papers assessed for inclusion, 177 articles were included in their narrative synthesis, which reported on 209 studies. They found that only seven of these addressed the waiting time paradox (described by Crawford *et al.* (2002)). 28 different cancer sites were investigated by the included papers, and 15 different time intervals were described as the 'diagnostic interval'. The methodology of the included papers was too heterogeneous for meta-analysis.

The results from the included studies were mixed, and there was variation in the

methodological quality. Even within specific types of cancer the authors found contradictory evidence about whether a shorter diagnostic interval had no effect on survival, improved survival, or in some cases decreased survival. An extensive analysis of the findings is presented in the original publication; the results for the four most common cancers are summarised here. For lung cancer, studies supporting a positive association between diagnostic interval and outcomes (shorter interval meaning better outcomes), no association, and a negative association (shorter interval meaning poorer outcomes) were found. Only one of these accounted for the waiting time paradox in the analysis; in that study, a positive association was found. For colorectal cancer, more studies showing a positive association than a negative association were found. Three of the four colorectal cancer studies that accounted for the waiting time paradox had found a positive association. For breast cancer, all included studies reported either a positive association, or no association. For prostate cancer, two studies reported a positive association (one of which accounted for the waiting time paradox) and the remaining four found no association. There was also strong evidence for benefits of a shorter diagnostic interval for head and neck, testicular, and melanoma. To a lesser extent, the same was also true for pancreatic and bladder cancer. Of the seven studies that had taken the waiting time paradox into consideration, most reported poorer outcomes associated with shorter diagnostic intervals.

Overall, this systematic review concluded that it is 'reasonable to assume' that shorter diagnostic intervals are associated with better survival, reduced mortality, and improved outcomes. This takes into account the effects of late stage disease with symptoms very strongly suggestive of cancer having short diagnostic intervals and poor outcomes. The methodological heterogeneity makes it difficult to sum or directly compare papers, but the recently published Aarhus statement which promotes greater precision and transparency in definitions and methods in early diagnosis research should mean that this improves with time (Weller *et al.*, 2012). Aside from the effects on outcomes, shorter diagnostic intervals are also desirable as patients and other stakeholders want quicker diagnosis (Banks *et al.*, 2014), and earlier stage cancers are easier and less expensive to treat (Insicive Health, 2014).

2.4.3.4 What factors contribute to delays in diagnosis?

The process of diagnosing cancer is complex and multi-factorial and there are many potential causes of delay throughout the process (see the NAEDI pathway discussed in Section 2.4.2); these fit in to three broad areas: the patient interval (delay in presenting to primary care); the primary care interval (delay in diagnosis); and the secondary care interval (delays to treatment). Delays at any of these stages can result in poorer

outcomes, and delays in the patient and primary care intervals can contribute to delays in diagnosis.

Within primary care, missed opportunities to diagnose can result in longer diagnostic intervals, longer time to treatment, and potentially later stage diagnosis (Lyratzopoulos *et al.*, 2015). A missed opportunity occurs when an alternative course of action would have resulted in the diagnosis being made sooner (McGlynn *et al.*, 2015). Missed opportunities for diagnosis often result in repeated primary care consultations all cause delay and increase the primary care interval; an increased number of consultations in primary care is associated with a longer primary care interval (Lyratzopoulos *et al.*, 2013), although not all of these consultations will reflect missed opportunities. Lyratzopoulos *et al.* (2013) found that one in five patients diagnosed with cancer consulted their GP $\geq$ 3 times before being referred; this varied by type of cancer from 7.4% of breast cancer patients to 50.6% of multiple myeloma patients. This study did not examine whether patients with more pre-referral consultations were diagnosed at a later disease stage, or had poorer survival.

2.4.4 Where does thrombocytosis fit in?

Thrombocytosis may have the potential to contribute to the earlier diagnosis of cancer and a reduction in diagnostic delay by prompting GPs to consider a cancer diagnosis earlier. This could be particularly useful in cases where multiple pre-referral consultations occur due to vague symptoms; one of the objectives of this thesis is to investigate the risk of cancer in patients who report symptoms in addition to thrombocytosis; two vague symptoms, weight loss and loss of appetite, are included as part of this. As thrombocytosis is often an incidental finding, one which is reported as part of a full blood count but rarely tested for for its own sake, the discovery and promotion of thrombocytosis as a previously unknown risk marker of cancer could prompt clinicians to investigate where they otherwise might not have done until the patient's symptoms worsened; this could take more time and repeated consultations. The patients who have the most potential for improvements to their cancer outcomes as a result of the recognition of thrombocytosis as a risk marker of cancer are those whose platelet counts are elevated at an early disease stage, and who do not have any classic alarm symptoms of cancer or do not have symptoms that match NICE guidance for urgent referral. The proportion of cancer patients who could fall in to this category is explored in this thesis.

2.5 Data sources used in this thesis

The final section of this introductory chapter describes the two data sources used in the thesis; the Clinical Practice Research Datalink (CPRD) and the English cancer registry.

2.5.1 The Clinical Practice Research Datalink (CPRD)

The CPRD is a government-funded organisation which collates anonymised electronic patient records into a longitudinal dataset for research purposes. It was founded in London in 1987 as the Value Added Medical Products dataset, which expanded to become the General Practice Research Database (GPRD) in 1993, and later the Clinical Practice Research Datalink (CPRD) in 2012. As of 2016, 674 UK practices were registered with the CPRD, contributing over 11.4 million patient records, a representative sample of around 8% of the UK population. 4.4 million patients in the database are 'active' (alive and currently registered with a CPRD practice) (Boggon *et al.*, 2013). Individual primary care practices register with the CPRD and all patient records from that practice contribute to the CPRD, unless patients individually opt out. Patient records are updated monthly, and are subject to strict quality control procedures (Herrett *et al.*, 2010). Practices contributing to the CPRD must follow set recording guidelines which outline precisely how to record patient events. Data quality is monitored at two levels; at practice level (with the practice required to meet certain data recording standards) and at patient level (with internal consistency required for sex, age, event recording, and registration details) (Khan *et al.*, 2010). All patients' consultations, laboratory results, and referrals are dated and coded with an internal coding system, mapped to the Read codes used in clinical practice (Williams *et al.*, 2012). Read codes are a standard set of coded medical terms used in the UK to record patient symptoms, diagnoses, findings and procedures (Booth, 1994). The observational data provided by the CPRD have been used for a wide range of studies.

2.5.2 Accuracy and validity of CPRD data

Despite the stringent quality controls, it cannot be overlooked that CPRD data are primarily recorded in practice for clinical use, not for research purposes, and therefore it is important to consider the quality of the data and their limitations for use in research. Some elements of the dataset, such as blood test results, have a high degree of accuracy due to being electronically transmitted to patient records directly from the laboratory. The accuracy of patient symptoms and diagnoses, recorded during consultations with GPs and retrospectively by practice staff, is subject to greater uncertainty.

A number of studies have aimed to assess the validity of the coded data in the CPRD, and these were recently summarised in a systematic review (Khan *et al.*, 2010). The authors carried out a comprehensive literature search of six bibliographies, which found 40 studies that had assessed the accuracy of data recording in the CPRD, and 12 that had evaluated the completeness of data recording. Studies that had assessed data accuracy compared CPRD diagnostic coding to a 'gold standard'; either a questionnaire to GPs ($n = 19$), comparison to hospital letters or records ($n = 16$) or both ($n = 5$). The systematic review found that the diagnostic accuracy of CPRD coding has been assessed for a range of acute and chronic conditions. The accuracy of CPRD coding was reported as the positive predictive value (PPV) of a code for a true diagnosis. The PPV varied dramatically between studies from 16.5% for acute liver injury to 100% for coarctation of the aorta, pressure ulcer, and non-organic psychosis. The majority of the 51 conditions assessed by the included papers had a PPV of over 50%, with 32 having a PPV of over 80%. The review did not find any studies that had assessed the validity of cancer diagnosis recording in the CPRD. The accuracy of the date of diagnosis was assessed by only three of the included studies for dementia, irritable bowel disease, and acute myocardial infarction. There was evidence of some differences in date of diagnosis, but these differences were generally small. Of the studies that had assessed completeness of data recording in the CPRD, there was generally evidence of under-reporting of disease incidence. However, although other sources were used as the 'gold standard', it is not always possible to say whether one database is 'better' or more complete than another. The systematic review included some papers that had assessed the validity of smoking status records in the CPRD. Compared to the Living in Britain National Household Survey of 1996, current smoking rates in the CPRD were 79% of expected rates. For previous smokers, the level was 29%. This suggests that previous smoking is considerably under reported in the CPRD. The review concluded that the validity and completeness of diagnostic coding in the CPRD is generally good, and recommended that researchers use several codes to identify diagnoses when working with CPRD data, or use internal validation methods such as looking for relevant referrals after particular diagnoses.

Other studies have looked specifically at cancer recording in the CPRD. Boggon *et al.* (2013) examined the concordance of cancer recording in the CPRD and the English cancer registry from 1997-2006. Both sources were searched for cancer-related codes, and where a diagnosis was recorded in the CPRD only, attempts were made to validate the record with hospital data or death certificates. Where only the cancer registry held a record of the diagnosis, the CPRD free text section was searched for selected text strings to confirm the diagnosis. Age, sex, and history of diabetes were investigated as potential predictors of non-concordance using logistic regression. The

study found 5,797 cancers recorded in the CPRD; 4,830 were also recorded in the cancer registry (83.3% agreement). Of the 967 CPRD cancer records that were not matched by a record in the cancer registry, 528 (54.6%) were confirmed using hospital statistics or death certificates. 307 of the 341 diagnoses recorded only in the cancer registry were confirmed in the same way. Age was found to be a predictor of non-concordance, which increased with increasing age. The majority of cancer registry diagnoses were recorded within one month of the CPRD record (63.1%). 23.5% were recorded within 1-3 months. 3.3% of cancer registry records were made more than a year before the first CPRD record. Conversely, 1.6% of cancer registry records were made more than one year after the first CPRD records. The authors suggest that more severe cases, where the patient dies soon after diagnosis, are less likely to be recorded in the CPRD as there is not enough time for the data to be fed back and obviously the patient does not attend again. They also suggest that cases missed from the cancer registry are more likely to be cancers for which diagnosis does not rely on histology (cancer registry recording being over reliant on histopathology in some areas). Other sources of discrepancy could be disagreements in coding between different clinicians, different coding patterns between the two sources, or errors in recording patient ID number. Overall, however, the paper concluded that the level of concordance between the two data sources is reasonably high.

Dregan *et al.* (2012) specifically evaluated the accuracy of lung, colorectal, gastro-oesophago, and urinary cancers in the CPRD compared to the cancer registry, from 2001-2007. Patient records were compared in the two sources and concordance was evaluated in terms of the positive predictive value (PPV) of a CPRD record (the proportion of CPRD diagnoses confirmed by the cancer registry), and the sensitivity (proportion of patients with a cancer record in the cancer registry also having cancer in the CPRD) and specificity (proportion of patients with no CPRD cancer record, who also had no cancer record in the cancer registry). During the study timeframe there were 5,429 cancers recorded in the CPRD and 5,710 in the cancer registry; they agreed on 5,216 of these (91%). 494 cancer registry cancers were not recorded in the CPRD, and 213 CPRD cancers were not recorded in the cancer registry. The PPV of a CPRD cancer record varied by cancer type: for lung cancer it was 96%, for urinary cancers it was 92%, for gastro-oesophageal cancers it was 97%, and for colorectal cancer it was 98%. The sensitivity varied from 85% for urinary cancers to 94% for lung cancers, and the specificity was generally high at 99% for all cancer types. The median difference in date of diagnosis between the two sources was 11 days (IQR -6 to 30); that is 11 days sooner in the cancer registry than in the CPRD, and the interquartile range is 30 days sooner in the cancer registry to six days later in the CPRD.

The authors explored possible reasons for discrepancies in diagnoses between the

two data sources. These included differences in the nature of the diagnosis, or recording of the diagnosis close to the end or just before the beginning of the study period. The patient may have transferred out of their CPRD-participating practice before their diagnosis could be coded in their CPRD record. It is possible that CPRD-only cancers were actually coded as 'suspected' cancer; if the suspicion was not confirmed and no diagnosis was made, there would be no accompanying cancer registry record: 77% of patients with a CPRD record but no cancer registry record had a cancer 'alarm' symptom which could have prompted a suspected cancer record. Different types of cancer may be recorded in each source if the 'suspected' cancer site was recorded first in the CPRD, then further investigation revealed the 'correct' cancer site, which is then recorded in the cancer registry. The patient's CPRD record is rarely (if ever) updated to reflect this. This study concluded that cancer diagnoses in the CPRD can be used with 'reasonable confidence' for the four types of cancer investigated. Some limitations of the study were noted, which will apply to other studies using CPRD data; namely that there may be differences in recording patterns between CPRD practices that do and do not subscribe to the cancer registry linkage system.

Overall, there is evidence to suggest that recording in the CPRD is of a reasonably high quality and that using linked cancer registry data when using cancer data from the CPRD can increase the degree of certainty in results. However, the results presented in Dregan *et al.* (2012) and Boggon *et al.* (2013) are based on data collected up to and including 2007. There are likely to have been improvements in data recording since this time as a result of policy initiatives in the UK in 2011 so we can assume that the quality of recording in the data used in the studies presented in this thesis is at least as good as the evidence described here.

2.5.3 The Cancer Registry

The National Cancer Registration Service for England (NCRS), the English cancer registry (hereafter referred to as the cancer registry), gathers patient data under strict collection, storage, and usage policies. It is run by Public Health England. The cancer registry dates back to 2000, when the Somerset Cancer Register was set up as part of the NHS Cancer Plan, before being extended across England where eight regional cancer registries operated. In 2013, the eight registries were combined into the National Cancer Registration Service for England (NCRS) which used the English National Cancer Online Registration Environment (ENCORE) database to collect data. The alternative services in other parts of the UK are the Information and Services Division in Scotland, the Welsh Cancer and Intelligence Surveillance Unit, and the Northern Ireland Cancer Registry. All four organisations feed data back to Public Health England. Data

are collected directly from healthcare providers including screening services, imaging services, Hospital Episode Data, pathology, chemotherapy and radiotherapy, secondary care patient administration systems, palliative care data, and death certificates. The single registration service aimed to standardise data collection for improved consistency and comparability. Data recording has improved steadily in the cancer registry since its inception. Currently, the database holds information on stage at diagnosis for about 80% of incident cases (100% coverage is not possible as not all cancers can be staged); this is a vast improvement on levels of staging recording in 2011, when levels were as low as 15% in some regions (Rashbass, 2014).

2.5.4 CPRD and cancer registry data linkage

A data linkage exists between the CPRD and the cancer registry. The primary function of the data linkage is to enable cancer registry records to supplement CPRD data. The linkage also allows cancer records in the CPRD to be validated by the cancer registry. The linkage is carried out by the Health and Social Care Information Centre (HSCIC), a trusted third party, using key identifying information such as date of birth, postcode, and NHS ID number (Boggon *et al.*, 2013). Around 75% of English CPRD practices and 58% of all CPRD practices consent to data linkage with the cancer registry.

2.6 Chapter summary

This chapter has presented a range of background information to enable a wider understanding of the context in which this PhD is set, and justification for the topic under study. The basic biology of platelet and thrombocytosis has been discussed, and the three pools of evidence underpinning the work done in the PhD have been presented. The burden of cancer in the UK has been described, and the processes in general practice around diagnosing cancer have been examined in terms of achieving earlier diagnosis to improve cancer outcomes. Finally, the data sources used in this PhD have been introduced. The next chapter presents the first study of the thesis; a systematic review and meta-analysis examining evidence of the relationship between thrombocytosis and cancer in a primary care setting.

Chapter 3

Systematic review of evidence of a relationship between thrombocytosis and cancer diagnosis

3.1 Chapter summary

In this chapter, the association between thrombocytosis and cancer is explored through a systematic review, the aim of which was to find evidence on this subject in a primary care setting. The methods used in undertaking the review are described, the results are presented, and the implications of the findings are considered.

3.2 Chapter introduction

Earlier diagnosis of cancer is a leading objective to improve cancer outcomes in the UK and address the gap in cancer survival between the UK and other European countries. It has been estimated that at least 5,000 cancer deaths annually could be prevented in England by improvements in early diagnosis (Richards, 2009). Screening can identify cancers at an early stage which may contribute to improved survival, although false negatives do occur, and screening programmes only exist for some types of cancer. Therefore, most diagnosed cancers (around 80% in the UK) present with symptoms to primary care (Hamilton, 2010). Although some symptoms of cancer are classic 'alarm' symptoms like a breast lump, many other symptoms are vague (such as tiredness and weight loss) and are more likely to be caused by other acute or chronic conditions. A

range of initiatives have been developed and implemented in UK healthcare systems to hasten cancer diagnosis, but most of these target the later stages of the diagnostic interval, after the general practitioner (GP) has considered the possibility of cancer. In the UK, these include two-week wait clinics, open access GP investigation and appropriate NICE guidance (NICE, 2015). Other work has identified the early symptoms and markers of cancer, which has resulted in the widespread use of tables of risk profiles for specific cancers in UK general practice (60,000 mousemats with risk assessment tables (RATs) printed on them were sent to general practices, and eRATs are being used in 1,000 practices) (Barrett *et al.*, 2010; Hamilton *et al.*, 2005a,b; Shephard *et al.*, 2013; Stapley *et al.*, 2006, 2012, 2013; Walker *et al.*, 2014). So far, thrombocytosis has received little attention in these studies, although some have considered it. As described in Chapter 2, there is evidence to suggest that earlier diagnosis leads to better outcomes and better survival. Earlier diagnosis is likely to identify cancer at an earlier pathological stage, where it is not only easier to treat, but also also more likely to be followed by longer survival than late stage cancers. Early stage cancers are also less expensive to treat than late stage (Insicive Health, 2014). All of this supports the notion that diagnosing cancer earlier is beneficial. However, achieving this in practice can be difficult.

Thrombocytosis has recently been identified as a marker of cancer present before diagnosis. Previous studies have reported on the usefulness of platelet count as a prognostic tool in secondary care and have described the mechanisms that could underlie this association (see Chapter 2) but there is a relative dearth of evidence from primary care, where the initial suspicion of cancer is generally first made. A range of case-control studies have been carried out with patients with particular types of cancer, and control patients with no cancer diagnosis, to compare the symptoms these patients presented with to their GP in the year prior to their diagnosis. Some of these have investigated thrombocytosis as a potential risk marker of malignancy.

In order to answer the general question of whether thrombocytosis is a marker of undiagnosed cancer, this systematic review aimed to identify studies that have investigated whether adults aged ≥ 40 presenting with thrombocytosis in primary care are at greater risk of having a currently undiagnosed cancer than those with a normal platelet count and to bring the results together in a narrative synthesis and meta-analysis.

3.2.1 Systematic reviews

A systematic review is a method of identifying and collating evidence from previously published studies which enables the gathering of a large body of evidence to answer a research question. The review is designed to identify studies with a degree of homogene-

ity within the research question so that it is possible and valid to draw their individual results together and base conclusions on the overall picture. Collation of results takes the strength of the evidence in to account. It is sometimes possible as part of a systematic review to carry out a meta-analysis; this is a method of statistically combining numerical results from individual studies to get an larger overall result, usually with a greater degree of precision than results from individual studies with smaller sample sizes can provide (Deeks *et al.*, 2008). In a meta-analysis, greater weight is generally given to larger studies. The degree of heterogeneity between studies is assessed within the meta-analysis using the I^2 statistic which is the proportion of variation between studies that can be accounted for by heterogeneity (Higgins *et al.*, 2003).

3.2.2 Chapter research questions

This systematic review addresses the first objective outlined in Chapter 1: to carry out a systematic review of studies that have reported on the association between platelet count and cancer in a diagnostic context to identify what is currently known about thrombocytosis as a risk marker of cancer. The research questions for this review are:

- Are adults aged 40 years and over with thrombocytosis at greater risk of cancer than those with normal platelet counts?
- Which cancer sites have been found to be associated with thrombocytosis in primary care (as risk markers of cancer), and which have not?

3.3 Methods

The protocol for this systematic review is available online and as an appendix (see Appendix A).

3.3.1 Scoping search

Before the full search strategy for this systematic review was finalised, a scoping search was carried out. This included informal, unstructured searching using internet search engines and Pubmed, and key terms relating to thrombocytosis, cancer, and primary care. A few key papers were known prior to beginning the review. The scoping search tested the search strategy to ensure it was effective at identifying the known key papers. If they were found by the search strategy, it is possible to have more confidence in the search strategy's ability to identify unknown yet relevant papers.

3.3.2 Search strategy

The search strategy was developed using three key terms: "thrombocytosis", "cancer", and "primary care". A range of additional search terms were derived from each of these; for thrombocytosis the following terms were used: "thrombocytosis" or "platelet" or "thrombocyte" or "thrombocyte count". For cancer, "neoplasm" was also included. Additional terms for primary care included "primary medical care" or "family practice" or "family medicine". The full search strategy is included in Appendix A.

The search was limited to English language papers due to a lack of available translating facilities. The search was also limited to results published in the last 30 years (the scoping search found no results beyond 30 years). The following databases were searched: EMBASE (OvidSP); Medline (Ovid); Web of Science, The Cochrane Library. Forwards and backwards citation searching was carried out on included papers. This involved checking the papers referenced in included studies for eligibility for inclusion, and examining later published works for eligibility which had referenced the included papers. In addition to searching electronic databases, a range of national and international experts were contacted to ask if they knew of any other relevant studies. The need for this strategy became apparent as, prior to starting the systematic review, I was aware of four studies that had collected data relevant to the systematic review but had not published the results because they were negative, and one further study which had collected but not analysed platelet count data. Publication bias is a well-documented limitation of systematic reviews, and it was a strength of this review that experts were contacted to ask about what data they may have that were relevant to the review, even if the results were negative. In addition to approaching other researchers directly, I attended a number of national and international conferences throughout the term of this PhD, delivered several oral and poster presentations, and had numerous discussions with other researchers at these events. Through this process I was able to determine if anyone else within the cancer diagnostics field was pursuing this line of investigation, or had collected relevant data.

All of the relevant papers identified by the literature search were exported to Endnote X5 and de-duplicated. Relevant papers identified through non-literature search were included.

3.3.3 Study selection

The search aimed to identify any study that had investigated the association between thrombocytosis and a new diagnosis of cancer of any type in a primary care setting.

Inclusion criteria were:

- Adults aged $\geq$ 40 years;
- Primary care setting;
- Observational, case-control or cohort study, or any literature review.

Exclusion criteria were:

- Adults under 40 years of age. These were excluded because cancers in this age group tend to be familial or atypical.
- Studies that had investigated platelet count as a prognostic tool or guide for cancer therapy. These were excluded as the systematic review specifically aimed to find studies that had used thrombocytosis in a diagnostic capacity.

Titles and abstracts were screened by myself and ES, a supervisor of this PhD and co-author on the subsequent publication, and full text articles were retrieved and assessed for inclusion.

3.3.4 Study quality assessment

The quality of included studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool (Whiting *et al.*, 2011). This is a tool for assessing the quality of diagnostic test studies. Although the overall aim of this PhD is to investigate thrombocytosis as a marker of cancer, not as a diagnostic test for cancer, this quality assessment tool was chosen as it best fits the type of studies that the review expected to identify and the types of studies that were found in the scoping search. The QUADAS-2 tool assesses studies for risk of bias and applicability concerns. The domains within these two areas are:

Risk of bias:

- Patient selection - could the patient selection procedure have introduced bias?
- Index test - could the conduct or interpretation of the index test have introduced bias?
- Reference standard - could the reference standard, its conduct, or interpretation have introduced bias?
- Flow and timing - could the patient flow have introduced bias? Are there any concerns about the time interval between the index test and the reference standard? Are all patients receiving the same reference standard and are all patients included in the analysis?

Applicability concerns:

- Patient selection - do the included patients and the setting match the review question?
- Index test - are there any concerns that the index test differs from that in the review question?
- Reference standard - are there any concerns that the target condition, as defined by the reference standard, does not match the review question?

3.3.5 Data extraction

A range of data relating to the number of patients with and without the cancers of interest, and the number of these with and without thrombocytosis prior to diagnosis were extracted from papers or from personal communications with the authors onto custom-made data extraction forms (see Appendix B). These forms also held details of the study design, patient or data sources, and patient characteristics. Data extraction was primarily carried out by myself, and was checked by WH, a supervisor of this PhD and co-author on the publication which resulted from this systematic review.

3.3.6 Data analysis: narrative synthesis

Two methods were used to collate and analyse data. Firstly, the extracted data were drawn together in a narrative synthesis following the general framework set out in the Centre for Reviews and Dissemination's Guidance for Undertaking Reviews in Healthcare (Centre for Reviews and Dissemination, 2009). In the first instance, a preliminary synthesis of findings was carried out which involved a textual description of studies, tabulating raw data and key features of studies.

The raw data extracted from each study were used to calculate the likelihood ratio (LR: the probability of raised platelets in patients with cancer divided by the probability of raised platelets in patients without the disease) and the positive predictive value (PPV) using Bayes' Theorem. The PPV reflects the probability that a patient with a positive test result has the disease in question. It is calculated as the proportion of patients with a positive test result who have the disease in question. This approach to calculating PPVs is used with case-control study data as it accommodates the lower prevalence in the general population, a necessary adjustment as prevalence is artificially high in case-control studies. Thus, PPVs were calculated as the LR multiplied by the prior odds of the disease (Knottnerus, 2002); in this case incidence data were used to calculate prior odds.

The analysis examined two subgroups which were defined based on (1) whether the cancer being studied affects both sexes or only one (this latter group transpired to be females only as no male-only cancer sites were identified by the searches), and (2) the data sources used by the studies. The relationships within and between studies were explored, and the strength and robustness of the evidence was assessed.

3.3.7 Data analysis: meta-analysis

Where available, the odds ratios from relevant studies were pooled using the random effects meta-analysis model in RevMan 5.3. (RevMan, 2014). The primary meta-analysis included all studies, and two sensitivity analyses examined the effect of pooling results from the two subgroups defined above. Heterogeneity in these models was examined using I-squared (I^2) statistic. The primary meta-analysis included several studies which had each investigated a specific cancer site. There were no studies that had investigated all or any type of cancer (two studies that have investigated the risk of any type of cancer in patients (Hippisley-Cox & Coupland, 2013a,b) did not include platelet count data) It is possible that the strength of any association between thrombocytosis and cancer may vary between different cancer sites. Furthermore, not all cancer sites had been investigated by studies identified by this review. 95% prediction intervals were calculated for pooled odds. There may not be one true single effect for the relationship between thrombocytosis and cancer; it may vary greatly between different types of cancer giving multiple different true effects. The 95% prediction interval quantifies the range of odds ratios across different cancer sites to provide an estimate of the interval in which the measure of association for another, as yet unreported, cancer site may lie. Rather than simply presenting the mean of the different effect of different cancer sites, the prediction interval reflects the range in which most of the different true effects lie.

3.4 Results

The process of selecting papers for inclusion in the review is summarised in Figure 3.1. A total of 98 papers were identified by the literature search. After duplicates were removed, 79 titles and abstracts were screened for inclusion. This screening was carried out by myself, and by ES, a supervisor of this PhD and a co-author on the resulting publication. WH was nominated to adjudicate with any papers that were a source of disagreement. Five full text papers were retrieved; both screening authors agreed on these five. The reasons for exclusion are shown in the study selection flow diagram; to summarise, the reasons are:

- Studies not being primary care-based ($n = 21$)

- Study did not include adults aged 40 years and over ($n = 13$)
- Studies not having a focus on cancer ($n = 32$)
- Not relevant study design ($n = 4$)
- Study did not investigate thrombocytosis ($n = 4$)

After reviewing full texts, four of the five papers identified through the literature search met the inclusion criteria (Hamilton *et al.*, 2005a; Shephard *et al.*, 2013; Stapley *et al.*, 2013; Walker *et al.*, 2013). All were case-control studies. A further five case-control studies were included through other channels described in the methods.(Hamilton *et al.*, 2005b, 2009; Shephard *et al.*, 2012; Stapley *et al.*, 2012; Walker *et al.*, 2014). All of these studies were carried out within the same research group, led by WH. This is a limitation of this review, but unavoidable; as yet (autumn 2016), no other research group has published work investigating thrombocytosis, although relevant work is ongoing and not yet published. Others who have investigated markers of cancer have not included thrombocytosis. A group in Sweden are collecting data on thrombocytosis and cancer, and work in Nottingham on the association has been initiated by the results of this thesis. Due to reporting bias four of these studies were not found by the literature search because, although they had collected and analysed platelet count data in relation to cancer diagnosis, they did not report this in the publication as the results were negative. As they did not report platelet data, the papers did not contain any keywords related to thrombocytosis or platelets. The fifth study, Shephard *et al.* (2012), had collected but not analysed platelet count data so was also not identified by the literature search. It was known that these studies had investigated a range of early markers of cancer and following discussion with the authors of these studies it became apparent that they had collected platelet data, and were able and willing to contribute raw data for this systematic review. In total, therefore, nine case-control studies were included, each investigating a different cancer site. Six of these studies had used data from the Clinical Practice Research Datalink (CPRD). The remainder of studies had used general practice records in a specific geographical location.

3.4.1 Study quality

The results of the QUADAS-2 assessment are shown in Table 3.1; the overall study quality was judged to be high.

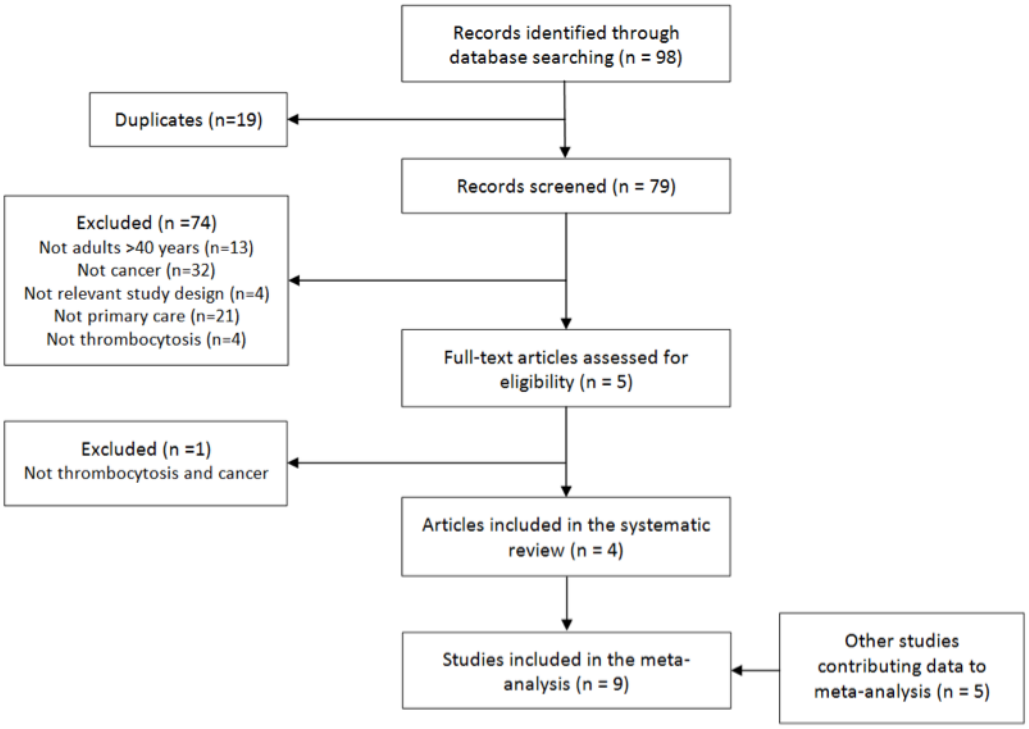


Figure 3.1: Study selection process including excluded papers and reasons for exclusion.

3.4.1.1 Patient selection

For all included studies, the methods of patient selection were well described. The studies that had used CPRD data had included all patients in the database with the specific cancer type being investigated. For those studies using general practice records, all patients with the cancer of interest identified from the records were included. The control patients were randomly selected age, sex, and practice matched patients from the CPRD, or age and sex matched patients from the same practice in the non-CPRD studies. In all cases, appropriate exclusions were applied (described in full in the next section). The selection of patients in a case-control study often introduces a degree of bias; in the CPRD studies found by the review, all existing eligible cases in the data source were included, reducing this selection bias (some diagnosed cases may be missing from the data source, and there could be non-random differences between those that are and are not recorded). There were no concerns that the included patients did not match the review question.

3.4.1.2 Index test

The index test, referred to in QUADAS-2, is the diagnostic test under investigation. It is compared to the *reference standard*. All of the studies included in this review aimed to estimate the diagnostic value of signs and symptoms in identifying patients with cancer. This review specifically sought data on platelet count as a diagnostic marker of cancer. Therefore, the index test was platelet count, typically taken as part of a full blood count. In all included studies, a blood test was used to determine platelet count, and all blood tests were carried out prospectively, prior to patients being diagnosed with cancer, so without knowledge of the reference standard. The threshold for thrombocytosis was pre-defined and taken as the local laboratory thresholds, generally $> 400 \times 10^9$/L or $> 450 \times 10^9$/L. As blood tests were ordered by GPs in practice prior to patients being diagnosed with cancer, there is little chance of any degree of bias being introduced; the control patients also had blood tests and platelet counts measured. The CPRD studies included patients with no platelet count available as having a normal platelet count. This may introduce a degree of bias, and could dilute any true association between thrombocytosis and cancer.

3.4.1.3 Reference standard

The reference standard determines whether or not a patient is classed as having the disease in question. For CPRD studies, patients were classified as having cancer if they had a code in their electronic medical records identifying them as being diagnosed

Table 3.1: Results of QUADAS-2 study quality assessment.

Citation	Risk of Bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow & timing	Patient selection	Index test	Reference standard
Walker *et al.* (2014)	✓	✓	?	?	✓	✓	✓
Shephard *et al.* (2013)	✓	✓	?	?	✓	✓	✓
Stapley *et al.* (2013)	✓	✓	?	?	✓	✓	✓
Walker *et al.* (2014)	✓	✓	?	?	✓	✓	✓
Shephard *et al.* (2012)	✓	✓	?	?	✓	✓	✓
Stapley *et al.* (2012)	✓	✓	?	?	✓	✓	✓
Hamilton *et al.* (2009)	✓	✓	✓	?	✓	✓	✓
Hamilton *et al.* (2005a)	✓	✓	✓	✗	✓	✓	✓
Hamilton *et al.* (2005b)	✓	✓	✓	?	✓	✓	✓

with the particular type of cancer being investigated. Histological confirmation was not sought; cancer recording in the CPRD is considered to be of high quality (discussed in Chapter 2 and confirmed by the validation study in Chapter 5), but recording errors do occur, and the quality of cancer recording in the CPRD is known to vary by cancer type (Dregan *et al.*, 2012). For studies that used GP practice data, cancer diagnoses were confirmed with histological reports. It is very likely that patients are correctly classified as histological confirmation is considered the gold standard of reference standards. All included studies used observational data in which cancer diagnoses were made or recorded by practice staff with no knowledge of the research study so there is unlikely to be any introduction of bias by prior knowledge of the results of the index test.

3.4.1.4 Flow and timing

It is not clear whether platelet count was available for all patients in the studies. If thrombocytosis was present it was coded as such. Normal platelet counts and missing platelet counts were grouped together into a 'no thrombocytosis', assumed normal platelet count group. In Hamilton *et al.* (2005a), the number of patients with a platelet count was available. Of 1,482 patients in the study, 528 had a platelet count available. This may have introduced a degree of bias if the patients with a platelet count available were somehow different from those with no platelet count data. In all studies, platelet counts taken up to a year before a cancer diagnosis were included. Older data were excluded to ensure that any symptoms and test results were related to the cancer

diagnosis which defined the cases.

3.4.2 Study characteristics

The included studies are summarised in Table 3.2. Six of the nine case-control studies used data from the Clinical Practice Research Datalink (CPRD) database. Cases were all patients in the CPRD database with oesophago-gastric, pancreatic, bladder, breast, kidney, or uterine cancer diagnosed from 2000 to the year of study. These patients were classified as cases by the identification of a site-specific cancer code in their records: no histology was available to confirm this diagnosis. Up to five age, sex, and practice matched controls were included per case. These were randomly selected from the CPRD database.

The remaining three studies had included all patients diagnosed with the cancer of interest (ovarian, colorectal or lung) within particular geographical areas, identified from cancer registry data (including histology) and from searching general practice records. Controls were age- and sex-matched patients randomly selected from the same practices.

3.4.3 Study exclusions

All studies excluded patients with metastatic cancer from a different primary site (although included patients with metastatic disease from the cancer of interest at the time of diagnosis), patients for whom no matching controls were available, patients with no data available in the year prior to diagnosis, duplicates, or controls whose matching case had been excluded. Other cancer-specific exclusions were applied that varied between studies; uterine cancer controls with a hysterectomy prior to the cancer diagnosis date of their associated cases were excluded from Walker *et al.* (2013) (uterine cancer) and patients with a mastectomy more than three months prior to their cancer diagnosis date in Walker *et al.* (2014) (breast cancer). All studies had included only adults aged 40 years and over.

Table 3.2: Characteristics of included studies, including study type, cancer site, source of data used in the study, and the number of cases and controls and proportion of females included in the study.

Citation	Cancer Site	Study Type	Data Source	Cases	Controls
Walker *et al.* (2014)	Breast	Case-control	All women in the CPRD database diagnosed with breast cancer from 2000-2009.	$n = 4,407$ 100% female	$n = 21,755$ 100% female
Shephard *et al.* (2013)	Kidney	Case-control	All patients in the CPRD database diagnosed with kidney cancer from 2000-2009.	$n = 3,183$ 38.7% female	$n = 15,707$ 40.2% female
Stapley *et al.* (2013)	Oesophago-gastric	Case-control	All patients in the CPRD database diagnosed with oesophago-gastric cancer from 2000-2009.	$n = 7,657$ 34.6% female	$n = 37,699$ 34.6% female
Walker *et al.* (2013)	Uterine	Case-control	All women in the CPRD database diagnosed with uterine cancer from 2000-2009.	$n = 3,166$ 100% female	$n = 9,537$ 100% female
Shephard *et al.* (2012)	Bladder	Case-control	All patients in the CPRD database diagnosed with bladder cancer from 2000-2009.	$n = 4,935$ 27.5% female	$n = 24,098$ 28.9% female
Stapley *et al.* (2012)	Pancreatic	Case-control	All patients in the CPRD database diagnosed with pancreatic cancer from 2000-2009.	$n = 3,635$ 52.0% female	$n = 16,459$ 52.0% female
Hamilton *et al.* (2009)	Ovarian	Case-control	All women in Exeter and mid- and east Devon diagnosed with ovarian cancer from 2000-2007, identified from the Royal Devon and Exeter hospital Cancer Registry, and electronic records of all patients at 39 general practices in Exeter, mid-Devon and east Devon, UK.	$n = 212$ 100% female	$n = 1,060$ 100% female
Hamilton *et al.* (2005b)	Colorectal	Case-control	All patients in Exeter, Devon, diagnosed with colorectal cancer from 1998-2002, identified from the Royal Devon and Exeter hospital Cancer Registry, and electronic records of all patients at 21 general practices in Exeter, UK.	$n = 349$ 49.3% female	$n = 1,744$ 49.3% female
Hamilton *et al.* (2005a)	Lung	Case-control	All patients in Exeter, Devon, diagnosed with lung cancer from 1998-2002, identified from the Royal Devon and Exeter hospital Cancer Registry, and electronic records of all patients at 21 general practices in Exeter, UK.	$n = 247$ 31.9% female	$n = 1,235$ 31.9% female

3.4.4 Association between thrombocytosis and cancer

The four studies identified by the literature search had reported multivariable (adjusted for potential confounders) analyses and found a statistically significant association between thrombocytosis and their specific cancer site (Hamilton *et al.*, 2005a; Shephard *et al.*, 2013; Stapley *et al.*, 2013; Walker *et al.*, 2013). These had investigated lung, kidney, oesophago-gastric, and uterine cancer. The adjusted odds ratios as presented in the publications were:

- Lung cancer 9.3 (95% CI 3.4 to 26)
- Kidney cancer 2.2 (95% CI 1.7 to 2.7)
- Oesophago-gastric cancer 2.4 (95% CI 2.0 to 2.9)
- Uterine cancer 1.50 (95% CI 1.00 to 2.25)

Of the remaining five studies, four reported no significant association between thrombocytosis and cancer and the fifth did not include thrombocytosis in the analysis; these were the studies identified through channels other than the literature search.

My independently calculated likelihood ratios (LRs) and positive predictive values (PPVs) for each cancer site are presented in Table 3.3; all with the exception of breast cancer indicate an increased probability of malignancy in patients with a blood test showing thrombocytosis. LRs were highest in ovarian, lung, renal, colorectal, and oesophago-gastric cancer, but due to differences in the prevalence of these cancers in the consulting population, lung and colorectal (the most common cancers of the nine) had the highest PPVs for cancer. Ovarian cancer, with the greatest LR of 14.61, had a relatively small PPV of 0.65 which reflects the relative rarity of this diagnosis.

Generally, studies that used the CPRD as their primary data source achieved larger sample sizes than those that used cancer registry and general practice records for patients within defined practice areas: CPRD studies included patient numbers in the thousands, with the smallest including 2,732 cases and 9,537 controls, up to the largest which included 7,481 cases and 32,877 controls. This is because all patients diagnosed with the cancer of interest were included as cases, from any CPRD-registered practice in the UK, rather than being restricted to cases within a set geographical area. In contrast, the non-CPRD studies included patients diagnosed with the cancer of interest from set geographical areas within South-West England and were much smaller; the largest recruited 349 cases and 1,744 controls. The confidence intervals were generally narrower and showed less overlap in CPRD studies when compared to the non-CPRD studies, reflecting their larger sample sizes. The lack of overlap between confidence

intervals for individual cancer sites suggests that thrombocytosis is more strongly predictive of some types of cancer than others.

The sex of included patients was another key difference between studies; three included only female patients due to the type of cancer (breast, ovarian and uterine) whilst the other studies included both males and females, it was not clear from the mixed-sex studies how many males and females were identified as having thrombocytosis. Breast cancer was the only one of the three female only cancers to show no statistically significant association with thrombocytosis (LR 1.22, 95% CI 0.97 to 1.53). This was one of the larger studies which used the CPRD as a data source; reflected in the narrow confidence interval. Thrombocytosis was predictive of the other two female cancers, uterine and ovarian; although the ovarian cancer study had the greatest LR of 14.61, it also had the widest confidence interval (95% CI 6.94 to 30.73) reflecting the smaller sample size. The LR for uterine cancer was smaller at 1.60, and the confidence interval narrower (95% CI 1.27 to 2.01); this study also used the CPRD as its primary data source and achieved a larger sample size.

Table 3.3: Analysis of data from included studies. Table details the number of cases and controls in each study and how many of these had a blood test result showing thrombocytosis. Independently calculated likelihood ratios (LR) and positive predictive values (PPV) are presented along with 95% confidence intervals (CI).

Reference	**Study source**	**Cancer site**	**Cases** n**,** n **(%) with thrombocytosis**	**Controls** n**,** n **(%) with thrombocytosis**	**LR (95% CI)**	**PPV % (95% CI)**
Hamilton *et al.* (2005a)	Literature search	Lung	247 34 (13.8)	1,235 19 (1.5)	8.9 (5.19 - 15.41)	1.63 (0.92 - 2.90)
Shephard *et al.* (2013)	Literature search	Kidney	3,183 348 (10.9)	15,707 251 (1.6)	6.20 (5.3 - 7.3)	0.17 (0.15 - 0.20)
Stapley *et al.* (2013)	Literature search	Oesophago-gastric	7,657 707 (9.2)	37,699 568 (1.5)	5.28 (4.73 - 5.90)	0.47 (0.42 - 0.52)
Walker *et al.* (2013)	Literature search	Uterine	3,166 110 (3.5)	9,537 207 (2.2)	1.60 (1.27 - 2.01)	0.08 (0.07 - 0.11)
Walker *et al.* (2014)	Contact with experts	Breast	4,407 91 (2.1)	21,755 369 (1.7)	1.22 (0.97 - 1.53)	0.38 (0.31 - 0.48)
Shephard *et al.* (2012)	Contact with experts	Bladder	4,935 156 (3.2)	24,098 247 (1.0)	3.08 (2.53 - 3.76)	0.10 (0.08 - 0.12)
Stapley *et al.* (2012)	Contact with experts	Pancreatic	3,635 214 (5.9)	16,459 222 (1.3)	4.36 (3.63 - 5.25)	0.13 (0.11 - 0.15)
Hamilton *et al.* (2009)	Contact with experts	Ovarian	212 26 (12.3)	1,060 9 (0.8)	14.61 (6.94 - 30.73)	0.65 (0.31 - 1.36)
Hamilton *et al.* (2005b)	Contact with experts	Colorectal	349 48 (13.8)	1,744 42 (2.4)	5.71 (3.84 - 8.50)	1.39 (0.94 - 2.09)

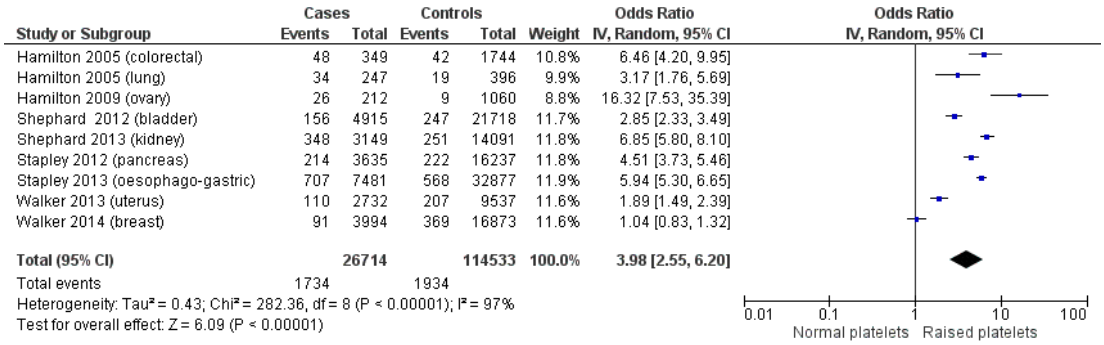


Figure 3.2: Forest plot and pooled result for meta-analysis including all nine studies identified by the review.

3.4.5 Meta-analysis

Data from all nine studies were pooled in the primary meta-analysis (Figure 3.2). This gave an overall odds ratio of 3.98 (95% CI: 2.55 to 6.20). The prediction interval for the odds ratios was 0.77 to 20.5; this indicates the range in which the odds ratio for a new cancer site not included in the meta-analysis might fall. Heterogeneity was high in this model, with an I^2 of 97%.

The high heterogeneity in this model prompted two sensitivity analyses which aimed to identify and (if suitable) eliminate the primary source of heterogeneity. It is possible that the heterogeneity is caused by differences between cancer sites in their relationship with thrombocytosis. It could also be caused by differences in the platelet-cancer relationship between men and women, or differences between data sources used by the studies - some used CPRD data and others used paper records in practices.

The first sensitivity analysis repeated the meta-analysis excluding the female-specific cancers from the model (breast, ovarian and uterine). The resulting OR was slightly higher than in the primary meta-analysis at 4.79 (95% CI: 3.59 to 6.38) and the prediction interval was 1.75 to 13.1. In this model heterogeneity was still high ($I^2 = 91\%$). The forest plot is shown in Figure 3.3.

The high heterogeneity in this sensitivity analysis suggested that differences between men and women in the relationship between thrombocytosis and cancer did not account (or at most accounted for only a small part) for the degree of difference between study results.

A second sensitivity analysis was carried out only including studies that had used CPRD data; and excluded data from the non-CPRD studies. The non-CPRD studies used primarily paper records, were much smaller, and focused on one geographical area (Hamilton *et al.*, 2005a,b, 2009). In this analysis, the OR was 3.19 (95% CI 1.87 - 5.45) with a prediction interval of 0.96 to 24.1, and I^2 value of 98%. The high heterogeneity

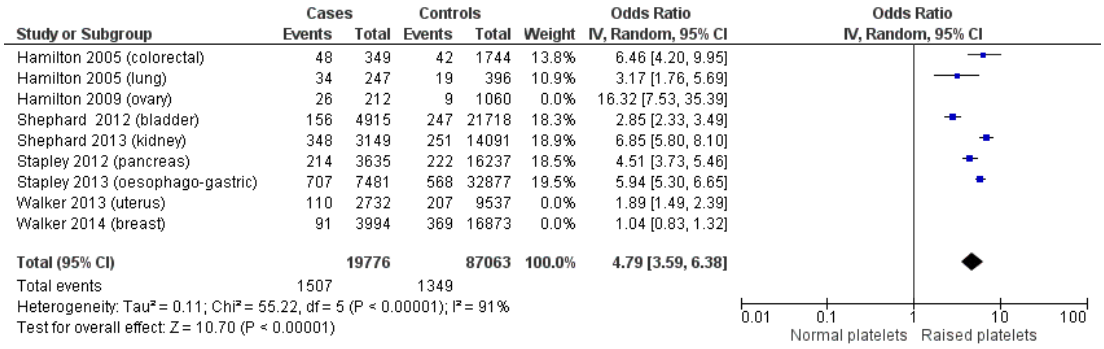


Figure 3.3: Forest plot and pooled result for meta-analysis including six studies identified by the literature review which included both male and female patients; female-specific cancers are excluded.

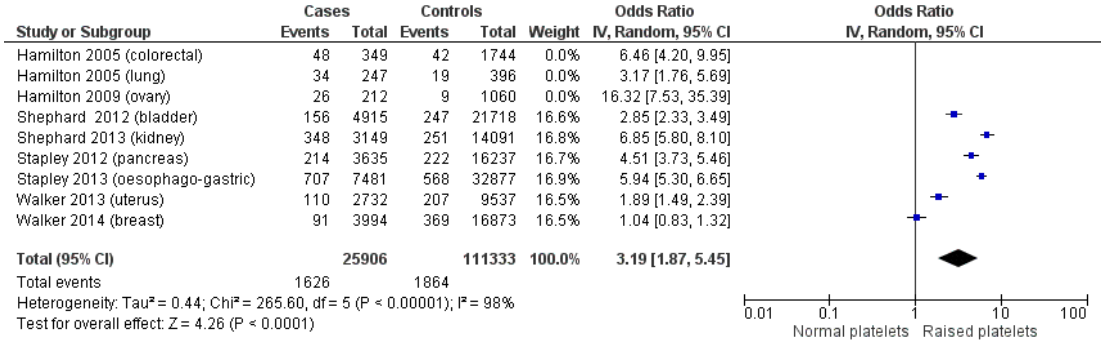


Figure 3.4: Forest plot and pooled result for meta-analysis including the five studies identified by the literature review which used CPRD data; non-CPRD studies are excluded.

indicates that the differences between studies again cannot be explained simply by differences in the primary data sources between studies. The results from the second sensitivity analysis are shown in Figure 3.4.

Overall, although the broad conclusions drawn for each study based on their individual ORs are generally in the same direction (with the exception of breast cancer), there are still differences between different cancer sites and their relationship with thrombocytosis that cannot be explained by different data sources or sex differences in the relationship.

3.5 Chapter discussion

3.5.1 Summary of results

This is the first systematic review to identify and collate results from studies investigating the association between thrombocytosis and diagnosis of cancer in primary

care. These results suggest that patients with thrombocytosis in primary care have an increased risk of harbouring an undiagnosed cancer, and that some types of cancer are more strongly associated with thrombocytosis than others.

The fact that all but one of the nine cancer sites investigated in studies identified by this review had conventionally significant positive likelihood ratios calculated from raw data adds strength to this overall conclusion; for four of the nine studies, however, the significant relationship between thrombocytosis and cancer was not retained in the multivariable models presented in the original publications. While this does not detract from the overall conclusion of an association between cancer and thrombocytosis, it does suggest that the results of this review should be interpreted with caution and supports the view that the association only exists for certain types of cancer.

There did not appear to be any biological link between those cancer sites that did and did not have a significant association with thrombocytosis in multivariable models. When considering the anatomy of cancer sites, there were counter-intuitive results; firstly that although colorectal and oesophago-gastric cancer, both of the digestive system, had similar odds ratios for cancer (6.5 and 5.9 respectively) presented in the original publications, only oesophago-gastric cancer retained thrombocytosis as a significant predictor of cancer in the published multivariable models. In the crude (unadjusted) analysis presented here, both sites had LRs of around 5, but the PPV for colorectal cancer was higher than that for oesophago-gastric cancer (oesophago-gastric 0.47% and colorectal 1.39%); this difference can be partly explained by the fact that colorectal cancer is more common than oesophago-gastric cancer. Secondly, uterine and ovarian cancer had somewhat disparate results despite both cancers being within the female reproductive system. Uterine cancer alone retained significance both in the original publication's multivariable models, and in my independent analysis (LR 14.61). Key differences in the way these types of cancer develop or manifest themselves may underlie these observations. Further investigation of this could provide evidence for the biological mechanisms that underlie the association between thrombocytosis and cancer.

The biological processes behind the thrombocytosis-cancer association have been studied in patients post-diagnosis (Arslan & Coskun (2005); Nash *et al.* (2002)) but the effect is not fully understood. It is also uncertain whether the mechanisms behind the association in secondary care would apply in primary care, before diagnosis is made.

3.5.2 Strengths and limitations

The quality of studies included in this review was judged to be high, and the majority used CPRD data. The merits of the CPRD as a data source are discussed extensively

elsewhere in this thesis. Briefly, the CPRD holds primary care records for patients from 684 GP practices, covering 8.8% of the UK population, with a representative geographical distribution. This largely representative sample can yield studies with widely applicable results. It is particularly relevant that the study data are taken from primary care, as it is in this environment that the study results should be clinically useful. The validity of CPRD data has been found to be high in one recent systematic review and two further validation studies (Boggon *et al.*, 2013; Dregan *et al.*, 2012; Khan *et al.*, 2010), although none of the studies included in those reviews addressed cancer data specifically. Laboratory results, including platelet counts, are directly transmitted to the CPRD, greatly reducing the chance for errors in data recording. The CPRD studies included all patients in the UK within CPRD registered practices who were diagnosed with the cancer of interest rather than patients within a restricted geographical area, as in the non-CPRD studies. This reduces the element of selection bias which is inherent in many case-control studies. It also enables much larger studies, and consequently more accurate and reliable results. However, CRPD cases are based on electronic records of cancer diagnoses alone; whereas in the non-CPRD studies, electronic records from the cancer registry were used as well as paper general practice records which had the added advantage of histology reports. In terms of the exposure variable, the platelet counts recorded in the CPRD are electronic and automatically transmitted from laboratory to the patient records, reducing the chance of human error and subsequent risk of misclassification bias. In contrast, the non-CPRD studies relied on manual checking and recording of platelet counts, more open to sources of bias and error. Overall, the CPRD studies can be considered higher quality than the non-CPRD studies; these show an overwhelmingly positive association between thrombocytosis and cancer, although there is marked variation between different cancer sites.

While these results can readily be applied to UK general practice, the lack of any non-UK studies limits the extent to which these findings can be generalised outside the UK. Although a wide range of search terms were used, alongside discussion with experts, supplemented by networking within the cancer diagnostic field, it is possible that some relevant studies were missed by the search. This is particularly likely if studies were published in non-English languages (language bias) or did not report non-significant findings (publication bias). It was not possible to widen our search to non-English language papers due to a lack of translating facilities, which may have introduced a degree of language bias. There is conflicting evidence on the extent of language bias in systematic reviews, summarised in the Cochrane handbook. The evidence cited in the Cochrane handbook includes a study by Moher *et al.* (2003) which found that excluding studies published in non-English language did not markedly alter the results of a meta-analysis, in a study of two meta-analyses (Moher *et al.*, 2003).

There is a chance that appropriate data may not have been identified due to publication bias. Four of the nine studies that contributed raw data to this study were not identified by the literature review because they did not report non-significant findings. Although relevant experts who have carried out similar studies were contacted to ask if they had collected platelet count data (they had not), it is possible that other researchers who were not contacted are carrying out similar work and have collected data, but that this has remained unpublished due to a lack of significant findings.

No formal method was used to assess publication bias. Publication bias is commonly assessed in meta-analyses using funnel plots; asymmetry in the funnel plot suggests the presence of publication bias (although other factors can also cause asymmetry; a more detailed factor analysis can be used to determine the exact causes). In the present chapter, publication bias was not formally assessed in this way; it is not recommended to use this approach when fewer than 10 studies are included in a meta-analysis, or when there is substantial heterogeneity (Sterne *et al.*, 2011). When there is heterogeneity evident in the meta-analysis, an even greater minimum number of studies is recommended for a funnel plot (Sterne *et al.*, 2011). Some publication bias was addressed in the present systematic review by the inclusion of data from studies that had failed to publish their negative (non-significant) findings for thrombocytosis.

A further limitation to this systematic review is that all of the included studies were carried out in the same research group (although by different researchers in the team), WH is an author on all, and also an author of the published paper of this systematic review. This could not be avoided; as far as I am aware, no other relevant studies have investigated platelet count prior to diagnosis. As described in the methods section of this chapter, attempts were made to contact researchers who had carried out similar work, even though they had not published anything regarding platelet count, in case they had collected data but not used it in any publications.

3.5.3 Implications for research

The results from this review indicate a link between thrombocytosis and cancer, but this link has not yet been fully investigated. The systematic review identifies significant gaps in the knowledge which future research should aim to address. Firstly, the association between thrombocytosis and all cancer sites needs to be investigated; only nine sites have been examined by existing literature. Secondly there is no indication from existing studies of how the relationship between thrombocytosis and cancer might differ with sex and age. Furthermore, existing studies are based on one single platelet count; in reality, clinicians often have multiple blood counts available or may test again if uncertain of the diagnosis. The risk of cancer with elevated platelet count over time is

unknown. Many cancers present in the early stages with vague symptoms and previous research has shown that the combined predictive value of two or more symptoms can be more clinically useful. Future research could also combine thrombocytosis with other symptoms or clinical features to provide greater estimates of cancer risk.

3.5.4 Clinical use of findings

This systematic review suggests that thrombocytosis is an early marker of some cancers in primary care and will raise awareness of this marker amongst general practitioners. This finding can be of use in primary care for general practitioners receiving blood results unexpectedly showing high platelet counts. Simply adding the PPVs from individual cancers suggests that the overall cancer PPV from thrombocytosis is at least 5% for any of the nine cancers (several cancer sites have not been reported and any association with thrombocytosis with them would increase this figure). The primary research in this thesis studies the relationship between thrombocytosis and cancer as a whole to give a more accurate estimate of the risk, and aims to determine which specific cancers present with raised platelets as an early marker to allow targeted investigation.

Crucially, this review does not suggest that platelet count should be specifically used as a diagnostic test for cancer, or as a screening tool. The evidence found in this systematic review only supports clinicians considering cancer as a diagnosis if a blood test result shows thrombocytosis. A greater understanding of the biological mechanisms underpinning the association would augment the epidemiological studies called for earlier. Likewise, knowledge of which cancer sites are most strongly associated with thrombocytosis could inform research on the mechanisms underlying the association. Research presented later in this thesis compares the risk for specific patient sub-groups, including men and women, those in different age groups, and at various levels of elevated platelets and combines thrombocytosis with other early markers of cancer to develop cancer-specific risk values for combinations of symptoms.

3.6 Chapter summary

This chapter has set out the current evidence base for the association between thrombocytosis and cancer, which clearly shows that further investigation is needed to fully understand the relationship. Recommendations for future research, derived from the findings of this review, are addressed in the next chapter - a prospective cohort study examining the association between thrombocytosis and cancer.

Chapter 4

Thrombocytosis as an early marker of cancer

4.1 Chapter introduction

This chapter presents a cohort study on the relationship between thrombocytosis and cancer. Background evidence presented in Chapter 2 supports a relationship between thrombocytosis and cancer, and the systematic review in Chapter 3 found that a raised platelet count is evident in some cancer patients prior to diagnosis; therefore, it could be diagnostically useful. However, the relationship has not been fully explored, and further research is needed to do so. Key factors that have yet to be studied are whether the link between thrombocytosis and cancer is stronger in some types of cancer than others, how the relationship changes with age, sex, and change in platelet count over time, and the usefulness of combining thrombocytosis with other symptoms. Understanding these important factors will maximise the clinical utility of the platelet-cancer association. The study presented in this chapter aims to address these factors. This chapter also examines the potential clinical impact of recognising thrombocytosis as a risk marker for cancer in UK suspected cancer guidance for clinicians by estimating the number of patients with thrombocytosis and cancer who had no other symptoms warranting cancer investigation prior to their diagnosis.

4.2 Chapter aim and objectives

The overarching aim of this chapter is to examine the relationship between thrombocytosis and cancer diagnosis, using individual patient data from the Clinical Practice Research Datalink (CPRD), and linked patient data from the English cancer registry. This chapter addresses objectives ii-ix set out in Chapter 1:

- To examine the incidence of cancer in two cohorts of patients; those with thrombocytosis and those with a normal platelet count, to determine the risk of cancer in each cohort.
- To compare the cancer incidence between these two cohorts to determine the absolute increase in risk associated with thrombocytosis.
- To examine how the relationship between thrombocytosis and cancer differs across subgroups defined by age, sex, and smoking status.
- To determine whether some types of cancer are more likely to be diagnosed than others in patients with thrombocytosis compared to patients with a normal platelet count.
- To investigate how the risk of cancer changes depending on how the patient's platelet count changes over time.
- To investigate the risk of cancer in patients who report symptoms in addition to thrombocytosis.
- To investigate the stage at which cancers are diagnosed in patients with thrombocytosis and with a normal platelet count.
- To estimate the potential impact of the recognition of thrombocytosis as a marker of cancer in UK suspected cancer guidance by examining the proportion of patients who have thrombocytosis but no other cancer symptoms or markers.

The methods used in this chapter are fully described in Section 4.3. The completeness of the dataset is explored and patient characteristics are described. The primary analysis presents the cancer incidence in patients with thrombocytosis; the proportion of patients diagnosed with cancer within one year of a full blood count showing thrombocytosis. This is compared to the incidence of cancer in patients with a full blood count showing a normal platelet count, in an age, sex, and practice matched sample.

A number of additional analyses are presented. The cancer incidence is estimated for subgroups defined by age (in 10 year age groups), sex, and primary site of cancer diagnosis. A crude (unadjusted) logistic regression model is fitted to estimate the chance of being diagnosed with cancer given a raised platelet count. Smoking status is tested as a potential moderator of the relationship between thrombocytosis and cancer incidence. Fractional polynomial models were fitted to examine platelet count as a continuous predictor of the risk of cancer. The relationship between the change in a patient's platelet count over a six month time period and the cancer incidence is

reported for patients with thrombocytosis. The cancer incidence the second year after index date is reported to examine whether cancer incidence returns to baseline levels in patients with thrombocytosis.

In a clinical setting, general practitioners will have additional patient information such as medical history and other symptoms to help them make a clinical judgement about whether or not to refer a patient for further investigation for cancer. Therefore, in this chapter, analyses are presented which investigate the risk of the two most common types of cancer in patients with thrombocytosis; colorectal and lung, and the most common symptoms of those cancers (rectal bleeding and change in bowel habit; and cough, respectively). Weight loss and loss of appetite, symptoms of many types of cancer, are also investigated.

The UK suspected cancer guidance (NICE, 2015) aims to guide clinicians on diagnosing cancer in primary care. The guidance was updated during the execution of this PhD; the updated version published in June 2015. Thrombocytosis was picked up as a marker for some, but not all, types of cancer in this guidance. Further analysis in this chapter examines the potential impact of the recognition of thrombocytosis as a risk marker of lung and colorectal cancer in the guidance. The methods are described below.

4.3 Methods

4.3.1 Ethical approval

Ethical approval was granted prior to the commencement of this study by the Independent Scientific Advisory Committee (ISAC) reference number 13-007.

4.3.2 Data sources

The study data came from two sources; the Clinical Practice Research Datalink (CPRD) and the English cancer registry.

4.3.2.1 The Clinical Practice Research Datalink

The CPRD is a government-funded organisation which collates anonymised electronic patient records from primary care into a longitudinal dataset. As of 2016, 684 UK practices are registered with the CPRD, contributing over 11.4 million patient records in total, representing around 8% of the UK population. 4.4 million patients in the database are 'active' (alive and currently registered with a CPRD practice) (Boggon

et al., 2013). The strengths and weaknesses of CPRD data including accuracy and validity are examined in Chapter 2.

4.3.2.2 The Cancer Registry

The National Cancer Registration Service for England (NCRS), the English cancer registry (hereafter referred to as the cancer registry), is run by Public Health England. It gathers patient data under strict collection, storage, and usage policies concerning all aspects of cancer diagnoses including staging data. The cancer registry is further described in Chapter 2. A linkage exists between the CPRD and the cancer registry to enable cancer registry records to supplement CPRD data. Linked data were requested from the CPRD for use in this study.

4.3.3 Study definitions

In this study, thrombocytosis is defined as a platelet count over 400×10^9/L. There is some regional variation in the cut-off point defining thrombocytosis between different laboratories; some take the upper limit as 450×10^9/L. A platelet count of over 400×10^9/L defining thrombocytosis was chosen for this study to encompass all possible definitions and to identify any association between thrombocytosis and cancer at only minimally raised platelet levels. An upper limit of $1,000 \times 10^9$/L platelets was set as a platelet count over this value is often associated with primary thrombocytosis (Syed *et al.*, 2007)(Griesshammer *et al.*, 1999) and would be routinely investigated in practice, with haematological malignancy the first consideration. There is also the possibility that values over 1000×10^9/L are an artefact; three patients had platelet counts apparently in the region of $200,000 - 300,000 \times 10^9$/L. These patients had been classed as having thrombocytosis when it is likely that they had normal platelet counts of $200 - 300 \times 10^9$/L.

4.3.4 Sample selection

Two cohorts of patients were randomly selected from CPRD records. The thrombocytosis cohort ($n = 40,000$) selection criteria were:

i. A blood test result with platelet count available, recorded between 2000 and 2013

ii. The blood test result showing a platelet count of $> 400 \times 10^9$/L for the first time in the patient's records

iii. Patient aged ≥ 40 years old at the time of that blood test

iv. Linked data available in the English cancer registry

The normal platelet count cohort (10,000) selection criteria were:

i. Patient to be age, sex, and practice matched to a patient randomly selected from the thrombocytosis cohort

ii. A blood test result with platelet count available, recorded between 2000 and 2013

iii. The blood test result showing a platelet count of 150 - 400 $\times 10^9$/L (within the normal adult range)

iv. Patient aged ≥ 40 years old at the time of that blood test

v. Linked data available in the English cancer registry

4.3.5 Justification of the two cohorts

The main cohort of $40,000$ patients was selected to address the primary research aim of investigating the risk of cancer in patients with thrombocytosis. The second cohort of $10,000$ patients with a normal platelet count was selected to estimate the difference in cancer risk between those who have a raised platelet count and those who have a normal platelet count. The total number of $50,000$ was chosen as the CPRD prefer to limit their datasets to this size. The estimation of the risk of cancer in patients with thrombocytosis is the main result, and the most clinically useful measure. However without some comparative measure of risk in patients with a normal platelet count, the true magnitude of the usefulness of thrombocytosis as a risk marker of cancer would not be evident. Comparing the risk of cancer in the thrombocytosis cohort to the risk of cancer in the general population is not a valid comparison because patients with thrombocytosis, ill enough to consult their GP and ill enough to have a full blood count ordered, are a select group of patients inherently different from the general population. Patients with a normal platelet count come from the same select group of patients (selected to have a full blood count) and are a superior comparator group. The difference in risk between the two cohorts will reflect increased risk associated with a raised platelet count, rather than the platelet count plus the risk associated with GP attendance and blood testing. A ratio of 4:1 patients was chosen because the main focus of this study was cancer incidence in patients with thrombocytosis; as large a sample as possible was required to investigate the different types of cancer diagnosed. The CPRD prefer to limit datasets to 50,000 patients. A ratio of 1:1 patients would have resulted in a greater degree of imprecision in the thrombocytosis results due to the smaller sample size. This is particularly important when investigating different types of cancer, many of which are fairly uncommonly diagnosed.

4.3.6 Justification of matching patients in the two cohorts

The two cohorts were defined based on their platelet count. The normal platelet count patients were matched to a random quarter of thrombocytosis patients. Matching is a technique usually used in case-control studies and it was used in this cohort study as age, sex, and various geographical factors are known to influence cancer risk. The main outcome from this study was to estimate the risk of cancer in patients with thrombocytosis and to compare this to the risk in patients with a normal platelet count. By matching on these three factors, their influence on the difference in effect size between the two cohorts can be eliminated.

4.3.7 Data preparation

The next section describes the data files received from each data source and how these were prepared for analysis.

4.3.8 Data files received from the CPRD

One set of raw data files[1] were received in text format from the CPRD, and were imported in to Stata for analysis. These include the following files:

- The *patient file* detailed the patients' personal and demographic factors including sex, year and month of birth, marital status, date of registration with their practice, and date of death (if applicable).

- The *clinical and referral files* both detail all events in the patients medical history: all symptoms, referrals, measurements, and the date that these were recorded in

[1] Actually, there were two. The original CPRD data file was received in February 2014. The cancer registry data file was received in February 2015. After a few weeks of data manipulation and analysis it became apparent that there was a very poor level of agreement between the two data sources on recorded cancer diagnoses. Of approximately 3,000 cancer records, the CPRD and the cancer registry agreed on 120. Around 2,880 CPRD-recorded cancers had no matching record in the cancer registry data. Further examination of the data revealed that there were around a thousand records in the cancer registry data for which no CPRD record existed at all (at least, not in my dataset).

Several weeks of checking and repeating work failed to find an explanatory error. There followed a chain of email correspondence with the CPRD in which I tried to explain the problem, and they made many helpful suggestions to try and find the solution. None of these revealed the error. In the end, the CPRD realised that they had sent me the wrong data file. My CPRD data extraction was completed in November 2013, and in December 2013 a second extraction was created for an unknown reason. I was mistakenly sent the December 2013 extraction; the cancer registry data had been matched on the November 2013 creation. Therefore there were different patients in each of the two datasets. The CPRD rapidly sent me the November 2013 extraction, and I set about repeating the work I had carried out on the original extraction. As I take detailed notes and use Stata do files and log files I was able to replicate the processes I had used with the December 2013 extraction. I estimate that this set the project back by three months.

the database. It is also possible to identify the job title of the staff member who entered the data.

- The *test file* contains all blood test results, references ranges, and the dates the test results were received.
- A *matching file* shows the randomly selected sub-sample of 10,000 thrombocytosis patients and the normal platelet count comparator to whom they were matched. This file also gives the sex, practice, and birth year of each patient, on which they were matched.
- *Therapy and immunisation files* detail any prescribed medications and immunisations for the sample, although these were not used in the analysis.

4.3.9 Data files received from the cancer registry

One data file was received from the cancer registry. This included the patient identification number (identical to the CPRD patient identification number), the patient's age at diagnosis, their ethnic background, their year of birth, and the date they started NHS treatment. The patients' diagnostic, tumour, and treatment details are also held. This includes the month and year of diagnosis, and date of diagnosis recorded as the number of days from the date the patient was registered with their practice. The basis of the diagnosis is noted; this is either recorded from a death certificate, a clinical investigation, a tumour marker, cytology, histology, or unknown. One variable identifies cancers that were diagnosed through a screening programme. The primary cancer site is recorded using ICD-9 or ICD-10 codes. A series of variables classify the grade, size, and morphology of the tumour, whether distant metastases were present at diagnosis, and whether the patient had positive nodes at diagnosis. A further series of variables note stage details using the TNM (tumour, node, metastasis) format. Where specific staging systems exist for particular types of cancer, these were also included. Further variables code treatment and death information. All of this information was received but only site, staging, and diagnosis date were used.

4.3.10 Data manipulation and work-up

The main dataset created for the analysis included some variables that were transferred directly from the 'raw' CPRD files and used in their original format. Many more variables were created from the raw data for the analysis. The following section describes the variables in the dataset, those which were used as supplied from the data source, and those which were created.

4.3.11 Raw variables used

Patient ID number, which uniquely identifies each patient in the dataset, was used in its original form. Sex was also used as supplied. The patients' platelet count at the date of their first reading was supplied as a continuous variable. The CPRD also supplied a binary variable which identified patients as either having thrombocytosis or a normal platelet count. The practice region variable was used as supplied.

4.3.12 Variables created

A number of additional variables were created during the data work-up stage; these are described below.

4.3.12.1 Patient index date

For patients with thrombocytosis, the index date was the date of their first raised platelet count within the study timeframe. This was supplied as a variable in the raw data file from the CPRD. For patients with a normal platelet count, using the date of their first normal platelet count within the study timeframe as their index date may have introduced bias if there was a considerable time difference between this and the index date of their matched thrombocytosis counterpart. Therefore the date of the platelet count nearest in time to the index date of their matched thrombocytosis counterpart was assigned as the index date for patients with a normal platelet count. Because this index date was sometimes later than the date used to define patients for inclusion in the cohort, 374 (3.7%) of these 'normal' patients actually had thrombocytosis at their index date and were excluded.

The two cohorts were age matched; patients from each matched pair were the same age on the index date of the thrombocytosis patient. The process to define index dates for the normal platelet count patients may have resulted in a difference in the median age at index date between the two cohorts. This was investigated in the analysis by comparing the time between index dates between the two cohorts. Due to this method of selecting index dates for the normal platelet count patients, the time difference between index dates for thrombocytosis and normal platelet count patients was estimated. This is because bias may have been introduced to the study if the normal platelet count patients' index dates were much later in time (and consequently, when the patients were older) than the thrombocytosis patients.

4.3.12.2 Date variables

Stata deals with dates numerically by converting all day-month-year formatted dates into a number: the number of days that have passed since 1st January 1960. In the raw data provided by the CPRD, date variables are given as the number of days that have passed since the patient registered with their practice. As all patients have been registered for a different number of days (some have been registered with the same practice since birth whereas others have moved recently), this coding format made comparisons between patients and variables difficult. Dealing with dates was further complicated by the fact that cancer registry dates are given as a month and year. Only the month and year of diagnosis are supplied to protect anonymity, so the first of the month was arbitrarily assigned to all diagnosis dates. To achieve consistency across all patients and data sources, new date variables were created for index date, all cancer diagnosis dates, and the date of any other symptoms. All dates were recorded as the number of days that had passed since 1st January 1900. This enabled easy comparison of dates.

Most patients diagnosed with cancer had a record of this in both the CPRD and the cancer registry, and in this case the first recorded date was taken as the date of diagnosis. Some patients had cancer recorded in the CPRD but not in the cancer registry, and vice versa. The date of the present record was taken as the date of diagnosis. An exploration and comparison of cancer recording in the CPRD and the cancer registry in Chapter 5, including an analysis of cases recorded in one source and missing from the other.

4.3.12.3 Age variables

To protect anonymity, only the month and year of birth is provided by the CPRD. The first of the month was assigned as the day of birth for all patients. Date of birth was converted to the number of days that had passed since 1st January 1900, and then subtracted from the 'days passed since' variables for index date, diagnosis date, and the date of other recorded symptoms to determine age at each of these dates. A categorical age variable was also created with 10 year age brackets; initially this included 40-49 years, 50-59 years, 60-69 years, 70-79 years, 80-89 years, and 90 and over, but there were too few people in the final category so a 80 and over group was used.

4.3.12.4 Smoking status

Smoking status and behaviour are recorded using a number of variables in the CPRD, and evidence of the relationship between smoking status and platelet count is mixed

(Butkiewicz *et al.*, 2006; Green *et al.*, 1992; Sloan *et al.*, 2015; Suwansaksri *et al.*, 2004). A binary 'ever smoked' variable was created (with a 'missing' option) using those raw variables. Smoking status is defined in the CPRD as current, past, or never. Additional variables record the type of smoking (pipe, cigarette etc.) and the number or amount smoked per day. Current or past smoking codes were used to define patients as having 'ever' smoked or not. This binary variable coded patients who had ever smoked as 1 and patients who had never smoked as 0. Where smoking status was missing for a patient but data were available on their smoking habits (type or frequency), this was used to enter data for the ever smoked variable (classified as ever smoked). Only patients who were coded as having never smoked in the raw CPRD smoking variable were classed as having never smoked.

4.3.12.5 Patient symptoms

One of the limitations of this study is that the reasons for patient blood tests being ordered are unknown. Blood tests are ordered in general practice for a variety of reasons; this can be in response to symptoms or in asymptomatic patients for routine or health check reasons. If cancer was already suspected in patients with thrombocytosis when blood tests were ordered, then the usefulness of thrombocytosis as a clinical prompt of suspected cancer is limited. In an attempt to address this, the symptoms reported by all patients in the month prior to their index date were compared for those with thrombocytosis and those with a normal platelet count. To do this, patients' electronic records were searched for all recorded medcodes within 28 days before index date blood test. The 100 most common medcodes in each group were listed. After excluding 'organisational' codes, the 10 most common symptom codes in the two groups were compared. Eight of these appeared in both cohorts.

4.3.12.6 Geographical region

The geographical region in which the patient's CPRD practice was based was supplied in the raw CPRD data. These were tabulated, and the number and percentage in each region were reported. Whilst the CPRD is commonly cited as a 'geographically representative sample', there may be some variation in the proportion in patients in each region.

4.3.13 Outcome variables

4.3.14 Cancer variables from the CPRD

All new cancer diagnoses, other than for non-melanoma skin cancer, in the two years after the patients' index dates were identified by searching the CPRD records for any of 2,134 cancer-related codes (see Appendix C). This validated list of cancer codes has been used in several studies, and is collated into 20 common cancer sites, mapped onto ICD-10 (World Health Organisation, 2012). A binary cancer variable was created (0 = no cancer diagnosis, 1 = a cancer diagnosis at any time point) by searching patients' clinical and referral files for any of these cancer related medcodes. Where more than one cancer code was found in patient records, the first record in time was taken as the primary cancer site and date of diagnosis. The 20 cancer sites are: bladder, breast, cervix, kidney, leukaemia, lymphoma, myeloma, oesophagus, pancreas, stomach, testis, uterus, brain, colorectal, lung, ovary, oral, melanoma, prostate, and miscellaneous. Any cancer records lacking an event date were excluded; in every case, patients with undated records had a second dated record.

4.3.15 Cancer variables from the English cancer registry

New cancer diagnosed recorded in the two years after the patient index date were also extracted from English cancer registry data. Date variables in the registry data were used to determine which cancers were recorded within the target time frame, and which were pre-existing (patients with pre-existing cancer were excluded).

4.3.16 Combining CPRD and cancer registry diagnoses

Patients were counted as having a newly diagnosed cancer if they had a record of one in either the CPRD or the cancer registry. Where they had a diagnosis recorded in both sources, the earlier recording was taken as the date of diagnosis.

A range of time-dependent analyses were planned, so a variable was created to record how many months after the index date the patient's cancer was diagnosed. From this, a binary variable was created which coded patients as either diagnosed with cancer in the first year after their index date or not (1 = patient diagnosed with cancer within one year of their index date; 0 = patient not diagnosed with cancer within one year of their index date).

A second year category variable was created for patients who had not been diagnosed with cancer in the first year: diagnosed in the second year (months 13-24 after the index date) or not (1=patient diagnosed with cancer within the second year after their index date; 0 = patient not diagnosed with cancer within the second year after their index

date). 'Second year after index date' reflects months 13-24 after the index date; not the two year 1-24 month period after index date. An exclusion variable was created and applied to all analyses relating to cancers in the second year after index date; this excluded all patients diagnosed with cancer in the first year from second year analyses.

A '4-12 months' variable was created for those not diagnosed in months 1, 2, and 3 which identified patients diagnosed with cancer in months 4-12 after their index date (1 = patient diagnosed with cancer in months 4-12 after their index date; 0 = patient not diagnosed with cancer within months 4-12 after their index date).

4.3.17 Staging data

I intended to address the objective of investigating the stage at which cancers are diagnosed in patients with thrombocytosis and with a normal platelet count using staging variables in the cancer registry data. However, staging data were only available for 50% of patients in the sample so this analysis was not completed. Of those 50%, half had early stage disease and half had late stage disease.

4.3.18 Exclusions

Patients were excluded from all analyses on three criteria. Although patients who met exclusion criteria could have been removed from the data file altogether, they were coded and excluded from each analysis so that it was possible to account for each of the $50,000$ patients initially selected.

4.3.18.1 Pre-existing cancer

The greatest exclusion in terms of the number of patients was on the basis of pre-existing cancer. Cancer was defined as pre-existing if a record of any cancer existed prior to the patient's index date.

4.3.18.2 Platelet count exclusions

Two platelet count exclusions were created; for patients in the thrombocytosis group who had a platelet count under the 400×10^9/L threshold or over $1,000 \times 10^9$/L and for patients in the normal platelet count group who had a platelet count of over 400×10^9/L.

4.3.19 Non-melanoma skin cancer

Any patients diagnosed with non-melanoma skin cancer within two years of index date were excluded. This decision was made because this type of cancer is more common than other types, it is rarely fatal, and very few cases reach the cancer registry. Many

non-melanoma skin cancers are managed solely in general practice, although histology could reach the cancer registry. Those cases that do reach the cancer registry are less likely to have complete registrations than other cancer types.

4.3.20 Patient selection process: summary

The patient selection process is summarised in Figure 4.1. This figure includes a diagram to show the 'timelines' used to select patients in the two cohorts, situations in which exclusions would have arisen, and potential biases that may have been introduced with different aspects of the patient selection process.

4.4 Statistical methods and analysis

4.4.1 Sex differences in the thrombocytosis-cancer relationship

Most of the analyses in this chapter are presented for men and women separately. This decision was made in response to early scoping work for this study which revealed differences in both thrombocytosis and cancer diagnoses between men and women, and was confirmed when the differences in cancer incidence between men and women were revealed.

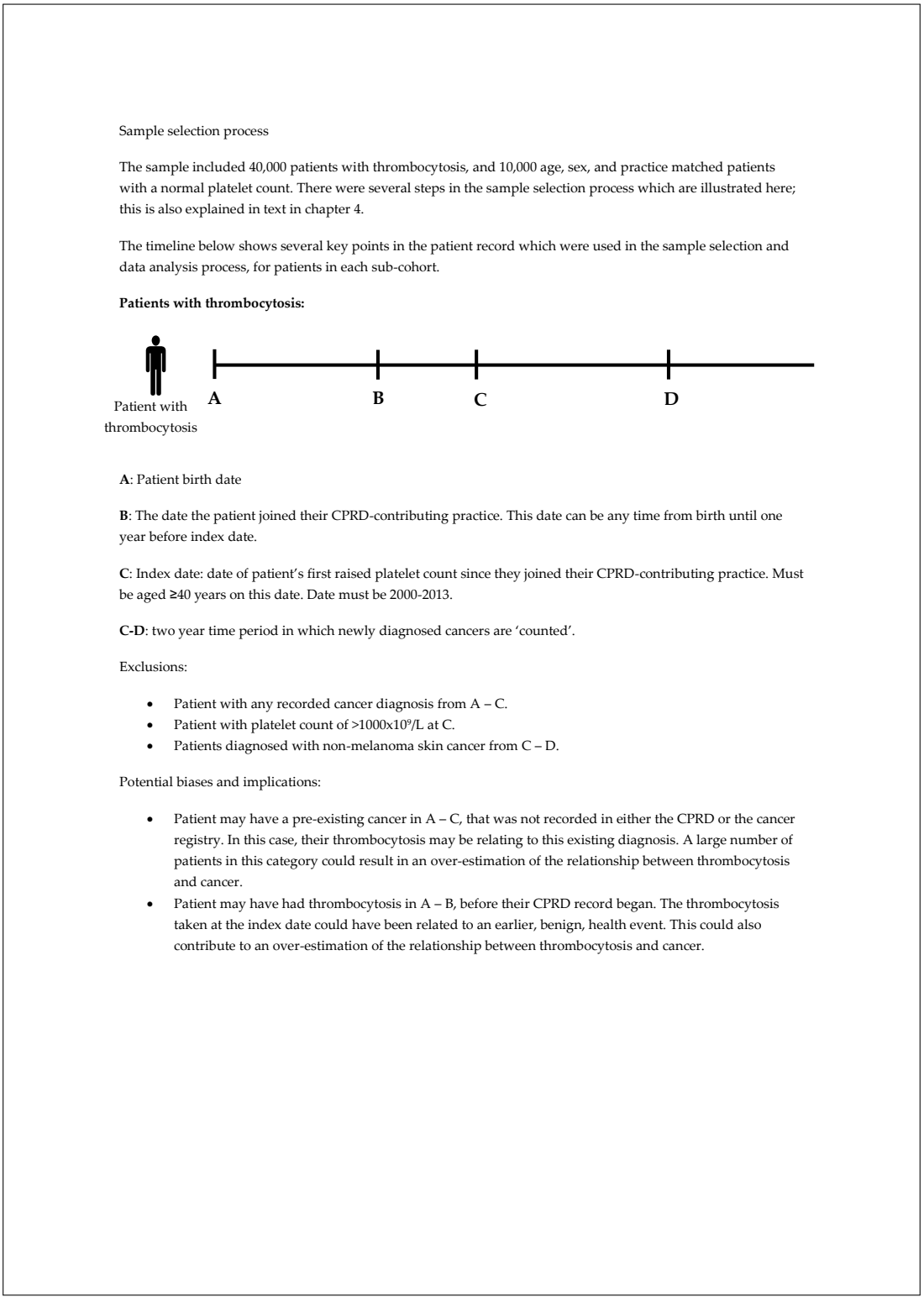
4.4.2 One year cancer incidence

The one year incidence of cancer was calculated for patients with thrombocytosis and patients with a normal platelet count as the number of new cancer diagnoses recorded within one year of index date, as a percentage of the number of patients in the cohort, reported with binomial 95% confidence intervals. The one year incidence was calculated for men and women separately, and compared between cohorts.

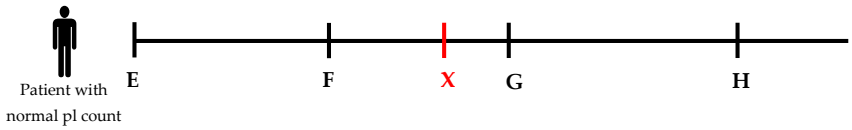
4.4.3 One year cancer incidence: epidemiological measures

A crude (unadjusted) logistic regression model was fitted to estimate the odds of a cancer diagnosis in patients with thrombocytosis compared to those with a normal platelet count. Odds ratios and 95% confidence intervals were reported. Smoking was included in the logistic regression model as an interaction term to investigate whether the size of this association differs between smokers and non-smokers. A fractional polynomial model was used to model the relationship between cancer incidence and platelet count as a continuous predictor, accommodating non-linearity in the relationship. Again, these analyses were performed separately for men and women.

Figure 4.1: Patient selection process summary



Patients with a normal platelet count:

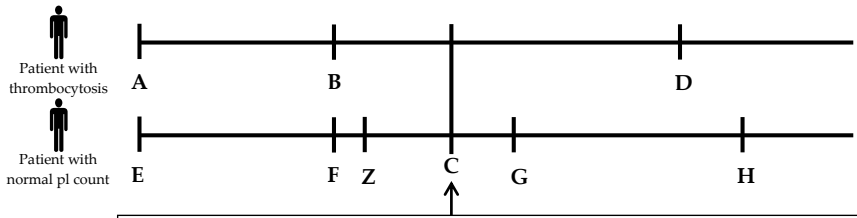


A comparison cohort of 10,000 patients with a normal platelet count was selected. These were age, sex, and practice matched to a random quarter of patients from the 40,000 thrombocytosis cohort.

E: Patient birth date – must be within five years of their thrombocytosis counterpart.
F: The date the patient joined their CPRD-contributing practice. This date can be any time from birth until one year before index date.
G: Index date: date of patient's blood test nearest in time to C (their matched patient's index date)
G – H: two year time period in which newly diagnosed cancers are 'counted'.
X: The index date of the matched thrombocytosis patient.
Additional selection criteria: to be considered eligible as a comparison patient, the patient must have had a platelet count from F – C which was within the normal range.

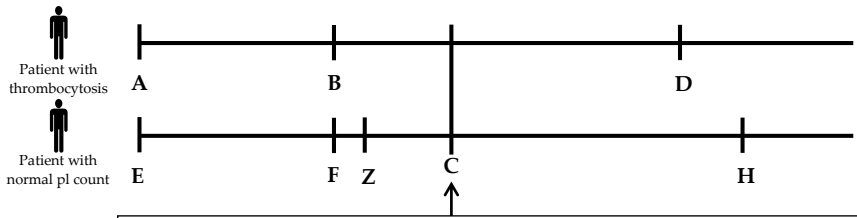
Exclusions:

- Patient with any recorded cancer diagnosis from E – G.
- Patients diagnosed with non-melanoma skin cancer from G – H.
- If G was later in time than C, some patients had thrombocytosis by G and were excluded.
  Example:



Date of matching. At this point, the platelet count is 'normal' as per the blood test at Z. But platelet count at G is nearer in time than Z, so is taken as the index date. By that point, the patient may have developed thrombocytosis.

- If G was earlier in time than C, some were aged <40 years at G and were excluded.
  Example:



Date of matching. Patient with normal platelet count may be aged 40 years on this date. In this case, they would be aged <40 years at Z, their index date.

Smoking was also investigated later in the analysis using a stratified comparison of cancer risk in different smoking status groups.

4.4.4 One year cancer incidence by age group

The effect of age on the risk of cancer in each of the two cohorts was examined using a stratified analysis with sub-groups defined by age at index date. The one year cancer incidence and corresponding 95% confidence intervals were estimated separately for men and women in each age group, for patients with thrombocytosis and with normal platelet count. As the risk of cancer increases with age, the rate of increase with age was compared between patients with thrombocytosis and patients with a normal platelet count. Scatter-plots were produced to show the incidence and 95% confidence intervals for each age group within each cohort.

4.4.5 Diagnostic interval for one year cancer incidence

The number of days between the index date (the date of the first thrombocytosis) and the cancer diagnosis date (the diagnostic interval) was determined for patients diagnosed with cancer in each cohort, to examine whether there were any differences between the two cohorts in the time to diagnosis following a blood test. This was expressed as the median number of days, with interquartile range. Histograms of the diagnostic interval in days were also produced for each of the cohorts. This analysis was carried out to determine whether the time to diagnosis differed between patients with thrombocytosis and their normal platelet count comparators.

4.4.6 Medium term risk of cancer

Additional investigation of the distribution of diagnoses in the time after index date was carried out to determine how useful thrombocytosis could be in terms of expediting diagnosis by more than three months. In this analysis, the diagnoses in each of the two cohorts were categorised depending on time after index date (measured in months). The data were then grouped according to whether or not the diagnosis was recorded within the first three months after index date, or the last nine months after index date. The rationale for this was to assess the potential for thrombocytosis to expedite cancer diagnoses by more than three months; if all cancers in patients with thrombocytosis were diagnosed within three months of a blood test, it is likely that other clinical signs and symptoms were present in patients which could have been triggering the suspicion of cancer in the patient's GP. Moreover, it is the patients whose diagnosis is made more than three months after their blood test result for whom the recognition of thrombocytosis as a marker of cancer would be the most clinically useful; if thrombocytosis

was used as a prompt symptom, they may have been referred for investigations and therefore diagnosed sooner. Patients diagnosed in the first three months were excluded from the analysis based on the final nine months.

4.4.7 One year cancer incidence by primary cancer site

The types of cancer diagnosed were compared for the two cohorts of patients. The primary site of diagnosis was determined from medcodes in the CPRD data or cancer registry data; where patients were diagnosed with more than one type of cancer, the first recorded site was taken as the primary site. The methods used to determine the primary cancer site categories are described above. The following 14 categories were included for all patients: bladder, renal, leukaemia, lymphoma, myeloma, oesophagus, pancreas, stomach, brain, colorectal, lung, oral, melanoma, and a miscellaneous category. For male patients a further two categories were used: testicular and prostate. For female patients a further four categories were added: breast, cervical, uterine, and ovarian. In total, therefore, there were 16 cancer site categories for men and 18 for women. Although men can be diagnosed with breast cancer, none in the cohort were.

The one year incidence of each cancer type was calculated for men and women in each of the two cohorts and compared. This was reported as a percentage, with 95% confidence intervals. The one year incidence of each type of cancer excluded patients who were diagnosed with other types of cancer. In addition, the number of cancers of each type was expressed as a percentage of the overall number of cancers diagnosed in each of the two cohorts (for example, the proportion of cancers that were lung, out of all cancers diagnosed). This was displayed in a pie chart for men with thrombocytosis and with a normal platelet count, and for women with thrombocytosis and with a normal platelet count. The types of cancer diagnosed in male and female patients with thrombocytosis were compared to the cancers diagnosed in the general population. For this comparison, national UK cancer incidence figures were used from the Cancer Research UK website (Cancer Research UK, 2012).

4.4.8 Change in platelet count over time

This post-hoc analysis[1] was developed to examine the risk of cancer in patients depending on how their platelet count changed over time. Patients' records were searched to

[1]Whilst I was conducting the analysis for this thesis, I gave numerous conference presentations to multidisciplinary audiences including academics, clinicians, and other professionals. During question and answer sessions, I was asked several times by clinicians what the risk of cancer was in patients whose platelet count increased or changed over time. Many mentioned that in a clinical setting very little is ever done in response to a single test result especially when, like platelet count, it has such a low specificity, and that clinicians are likely to 'wait and see' or repeat the test within a few months to see if there was any change, and what the direction of change is.

find any second platelet count or blood test result recorded within six months of their index date. For some patients, no second count was available. Those who had a second platelet count available were grouped according to how their platelet count had changed within this time frame. Thus, four categories were created:

- Increased platelets: the patient's second platelet count had increased compared to their first count and was still in excess of 400×10^9/L, or had remained exactly the same.
- Decreased platelets: the patient's second platelet count had decreased compared to their first count, but was still in excess of 400×10^9/L.
- Normalised platelets: the patient's second platelet count showed that their platelets had returned to the normal range of 100 - 400×10^9/L.
- No second platelet count available.

The risk of cancer (as one year cancer incidence, percentage with 95% confidence intervals) was compared for men and women with thrombocytosis in each of these four groups.

4.4.9 Thrombocytosis accompanied by symptoms

The main analysis from this study reflects the risk of a patient being diagnosed with any type of cancer within one year of a blood test result showing thrombocytosis and no further clinical information. In a clinical setting, GPs have to decide which type of cancer is likely to know which investigative tests and procedures would be most appropriate and most likely to reveal an underlying malignancy. Clinicians will have information about their patient's medical history, current symptoms, and their own clinical judgement in addition to blood test results to aid in the decision making process. This analysis was carried out to quantify the risk of cancer in patients with thrombocytosis and one other relevant symptom. The most common symptoms of the two most commonly diagnosed cancers seen to be associated with thrombocytosis were chosen; cough for lung cancer, and change in bowel habit and rectal bleeding for colorectal cancer (two symptoms were tested for colorectal as they were equally common). In addition, two 'generic' cancer symptoms, loss of appetite and weight loss, were examined. In order to ensure that only symptoms near in time to the patients' index dates were included in this analysis, only those that were recorded in the three months before or one month after the index date were used.

For this analysis, a 'symptom library' of relevant medcodes developed by the research group in which this PhD was completed was used to identify medcodes that indicated the symptom of interest (see Appendix D).

4.4.9.1 Lung cancer and cough

Patients were included in this analysis if they had thrombocytosis and a cough reported within the four month peri-thrombocytosis period; the three months before and the one month after their index date. A number of medcodes were used to search for patients reporting a cough. The one year incidence of lung cancer in these patients was estimated.

4.4.9.2 Colorectal cancer, rectal bleeding, and change in bowel habit

Patients were included in this analysis if they had thrombocytosis and a record of rectal bleeding, or thrombocytosis and a change in bowel habit. Medcodes included under change in bowel habit were those pertaining to diarrhoea, constipation, or 'change in bowel habit'. The one year incidence of colorectal cancer in these patients was estimated.

4.4.9.3 Any cancer and loss of appetite

Patients were included in this analysis if they had thrombocytosis and a medcode indicating loss of appetite within the four month peri-thrombocytosis period. The incidence of any type of cancer was estimated for these patients.

4.4.9.4 Any cancer and weight loss

Patients were included in this analysis if they had thrombocytosis and a medcode indicating weight loss within the same four month time period. The incidence of any type of cancer was estimated for these patients.

4.4.10 Symptom profiles and NICE guidance

The UK national guidance for diagnosing cancer in primary care (Suspected cancer: recognition and referral, NG12, NICE (2015)) was updated during the course of this PhD. In the 2015 update to the guidance, thrombocytosis was included as a marker of some, but not all, types of cancer primarily in response to evidence that was reviewed in Chapter 3 of this thesis. The following analyses were carried out for the two most commonly diagnosed cancers in the cohort; lung and colorectal.

4.4.11 Lung cancer and NICE guidance

This analysis aimed to estimate what, if any, benefit there could be from the addition of thrombocytosis as an alarm sign of lung cancer to the 2015 NICE guidance update. The symptoms reported by patients with thrombocytosis and lung cancer in the year before their diagnosis were identified and compared to see how many more patients met the NICE criteria for investigation before and after the addition of thrombocytosis. Patient's electronic records were searched for any of the symptoms listed by the NICE guidance for lung cancer; this includes cough, fatigue, shortness of breath, chest pain, weight loss, and loss of appetite. As the lung cancer guidance varies depending on whether or not the patient has ever smoked, the analysis was stratified in this way. The proportion of lung cancer patients who had NICE-qualifying symptoms other than thrombocytosis in the year prior to diagnosis was calculated. If all patients with thrombocytosis and lung cancer had other symptoms that would have prompted investigation then there is little, or no, value in adding thrombocytosis. Thrombocytosis has the most potential clinical use if the results show a proportion of lung cancer patients who have thrombocytosis but no other investigation-triggering symptoms prior to their diagnosis.

4.4.12 Colorectal cancer and NICE guidance

The 2015 NICE update did not include thrombocytosis as a marker of colorectal cancer. This analysis was carried out to see what additional benefit there could be from including thrombocytosis as a marker of cancer to the NICE guidance for diagnosing colorectal cancer. A similar method was employed as for the earlier described lung cancer analysis. The electronic records of patients with thrombocytosis and colorectal cancer were searched for any symptoms listed in the 2015 NICE guidance as warranting investigation in the year prior to their diagnosis; weight loss, abdominal pain, rectal bleeding, iron-deficient anaemia, change in bowel habit, or a rectal or abdominal mass. The symptoms reported by patients with thrombocytosis and lung cancer prior to their cancer diagnosis were examined. The proportion with symptoms that matched NICE guidance for referral, and the proportion who did not have symptoms that matched the guidance (other than thrombocytosis) were compared.

4.4.13 Two year cancer incidence

The risk of cancer in the second year after index date, and in the overall two year period since index date, was investigated as a sensitivity analysis. Previously defined variables were used to identify patients who were diagnosed with cancer 1-24 months of

index date and 13-24 months of index date and the cancer incidence was estimated for patients in each cohort in these timeframes (with 95% confidence intervals). The types of cancer diagnosed in the second year were compared for patients with thrombocytosis and with a normal platelet count.

4.5 The association between thrombocytosis and cancer: sample characteristics

The following analyses are based on $50,000$ patients who had a full blood count, subdivided in to two cohorts; those with thrombocytosis ($N = 40,000$) and those with a normal platelet count ($N = 10,000$).

4.5.1 Application of exclusion criteria

10,770 patients were excluded for the following reasons (see Figure 4.2):

1. Pre-existing cancer
   8,369 thrombocytosis patients and $1,581$ normal platelet count patients with a cancer diagnosis dated prior to their index date were excluded because it could not be determined whether any subsequent records were in reference to a new cancer, a metastasised cancer, or related to a pre-existing cancer.

2. Platelet count exceeded set limits
   Patients were excluded if their platelet counts were too high or too low for their respective groups. 92 patients in the thrombocytosis cohort were excluded because their index date platelet count exceeded $1,000 \times 10^9$/L. In the case of four of these patients, it appeared likely that extra zeros had been mistakenly included in the test result (for example, 360,000 rather than 360) and so patients had incorrectly been classified as having thrombocytosis when their platelet count was actually in the normal range (a platelet count in the hundreds of thousands is biologically implausible). The remaining 88 were excluded because their platelet counts of over $1,000 \times 10^9$/L indicated either an incorrect test result, or a haematological malignancy; either of which would have confounded the results.

   In the normal platelet count cohort, 374 patients were excluded because their index platelet count exceeded 400×10^9/L. For patients with a normal platelet count, an index date was assigned which was the date of their platelet count nearest in time to the index date of their matched case. Because this index date was sometimes later than the first normal platelet count, these 374 (3.7%) of these had thrombocytosis at their index date and so were excluded.

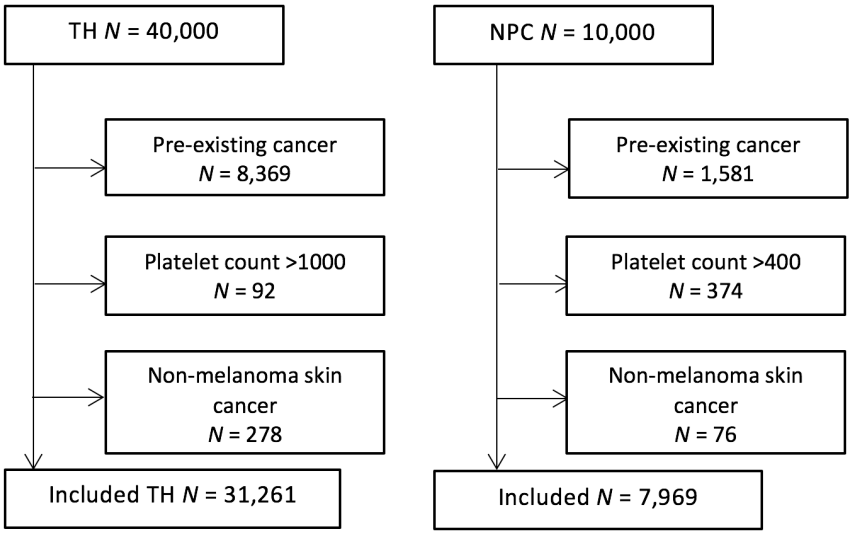


Figure 4.2: The number of patients excluded from the main cohort due to ineligible platelet count, pre-existing cancer, or being diagnosed with non-melanoma skin cancer.

3. Non-melanoma skin cancer
   278 of the thrombocytosis cohort and 76 of the normal platelet cohort were excluded as they were diagnosed with non-melanoma skin cancer.

4.5.2 Age profile of the cohorts

After application of the exclusion criteria, the thrombocytosis cohort included 31,261 patients and the normal platelet cohort included 7,969 patients (see Figure 4.2). The median age at the index date was 68.0 (IQR 57.1-78.1) years in the thrombocytosis cohort. The median age in the normal platelet cohort was 68.3 (IQR 58.1-78.5) years (see Table 4.1); although patients were matched on age, the selection of index dates for normal platelet patients resulted in some minor differences in age at index date. The ages of the two sub-cohorts are presented as box plots in Figure 4.3. When the age of the two sub-cohorts was grouped into 10 year age bands, there was an even distribution of ages across the groups (see Table 4.2).

4.5.3 Comparing cohort age profile with the general population

The age profile of the two study cohorts was compared to the age profile of adults aged 40 years and over in the general population in England. Table 4.3 shows the number and percentage of the population aged 40 years and over in each ten year age brackets for the thrombocytosis cohort, the normal platelet cohort, and the general population.

Table 4.1: Characteristics of patients with thrombocytosis and normal platelet counts in an English primary care setting.

	Thrombocytosis $N = 31,261$	**Normal platelet count** $N = 7,969$
Age at index date		
Age in years, median (IQR)	68.0 (57.1 - 78.1)	68.3 (58.1-78.5)
Sex distribution		
Men, *n* (%)	9,435 (30.2)	2,599 (32.6)
Women, *n* (%)	21,826 (69.8)	5,370 (67.4)
Platelet count		
Count $\times 10^9$/L, median (IQR)	441 (416-488)	255 (218-299)
Smoking status		
Ever smoked, *n* (%)	17,934 (57.4)	4,212 (52.7)
Never smoked, *n* (%)	12,668 (40.5)	3,688 (46.3)
No data, *n* (%)	659 (2.1)	69 (1.0)

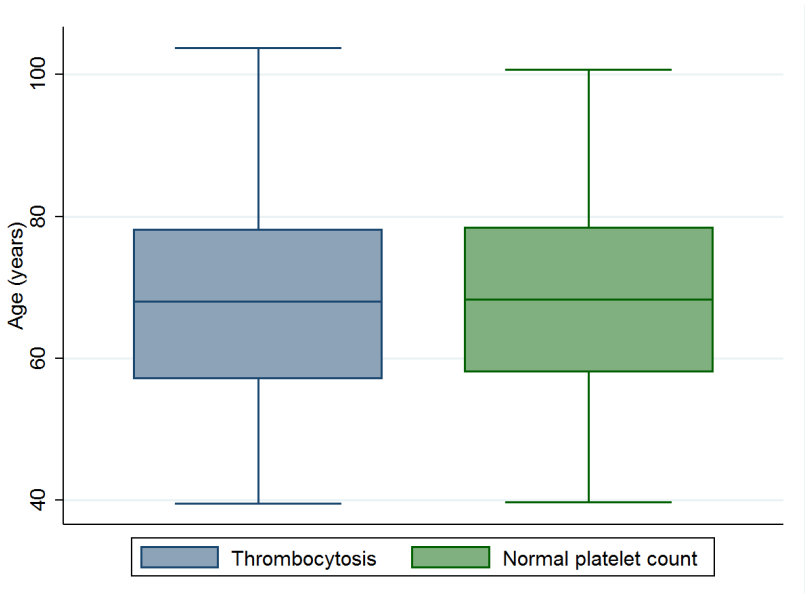


Figure 4.3: Box plot showing age distribution of all patients in the thrombocytosis cohort and the normal platelet count cohort.

Table 4.2: Number and percentage of patients in each age group, for thrombocytosis and normal platelet count cohorts.

Age group	**Thrombocytosis** $N = 31,261$		**Normal platelet count** $N = 7,969$	
	Men n (%)	Women n (%)	Men n (%)	Women n (%)
40-49	831 (8.8)	2,744 (12.6)	174 (6.7)	612 (11.4)
50-59	2,195 (23.3)	4,056 (18.6)	560 (21.6)	1,001 (18.6)
60-69	2,719 (28.8)	4,578 (21.0)	764 (29.4)	1,175 (21.9)
70-79	2,303 (24.4)	5,441 (24.9)	653 (25.1)	1,319 (24.6)
≥ 80	1,387 (14.7)	5,007 (22.9)	448 (17.2)	1,263 (23.5)

Table 4.3: Age distribution for patients in the thrombocytosis cohort, the normal platelet cohort, and in the English general population in 2008. Age is divided into ten year age brackets and the number and percentage of men and women in each age group is compared.

Age group	**Thrombocytosis** ($N = 31,261$)		**Normal platelet count** ($N = 7,969$)		**General population** (1,000s, $N = 25,101.4$)	
	Men n (%)	Women n (%)	Men n (%)	Women n (%)	Men n (%)	Women n (%)
40-49	831 (8.8)	2,744 (12.6)	174 (6.7)	612 (11.4)	3,763.1 (31.6)	3,825.8 (29.0)
50-59	2,195 (23.3)	4,056 (18.6)	560 (21.6)	1,001 (18.6)	3,052.7 (25.6)	3,130.4 (23.8)
60-69	2,719 (28.8)	4,578 (21.0)	764 (29.4)	1,175 (21.9)	2,586.5 (21.7)	2,737.4 (20.8)
70-79	2,303 (24.4)	5,441 (24.9)	653 (25.1)	1,319 (24.6)	1,670.0 (14.0)	1,980.3 (15.0)
≥ 80	1,387 (14.7)	5,007 (22.9)	448 (17.2)	1,263 (23.5)	850.2 (7.1)	1,505.0 (11.4)
Total	**9,435**	**21,826**	**2,599**	**5,370**		

The age distribution differs between the study cohorts and the general population. For men in the study cohorts, the proportion of patients in each age group increases through the 40s, 50s, and 60s, peaking in the 60s, and declines thereafter. For women, the proportion in each age group also steadily increases, but peaks in the 70s and decreases only slightly in the aged $\geq$ 80 years group. In the general population, the age distribution peaks in the 40s and steadily declines into the $\geq$ 80 years group. This reflects the current age distribution in the UK general population where the number of people decreases with age, whereas the patients included in this study are a group selected based on having had a blood test - the age profile in this cohort is different to that in the general population. The inclusion of a greater proportion of older patients in the cohort, compared to the general population, may introduce a degree of bias to the results as cancer incidence increases with age. However, the main analysis in this study compares the risk of cancer between patients with raised and those with normal platelet counts, not with the general population. Between the two cohorts, the age profile was very similar, as patients were matched on age, so this potential source of bias was eliminated. It is not possible to generalise the results presented in this chapter to the general population, a concept which is explored in more depth in the final chapter of this thesis, although they can be generalised to other patients selected for blood testing in general practice. The age profile of patients selected for a blood test cannot be determined with the data used in this study as the normal platelet count patients are age matched to the thrombocytosis cohort.

4.5.4 Gender profile of the cohort

A greater proportion of the cohort were female; 21,826 (69.8%) of the thrombocytosis cohort were female, and 5,370 (67.4%) of the normal platelet cohort. There is some evidence to suggest that women consult primary care more often than men (see Chapter 2), which could explain the higher proportion of women in this randomly selected sample rather than gender differences in the relationship between thrombocytosis and cancer. Differences in consulting rates between men and women are greatest in mid-life and equalise in later life. Around a third of patients in this sample are in mid-life (see Table 4.2) so it is possible that factors other than differences in consulting rate are contributing to the higher proportion of women in the sample. Thrombocytosis appears to be more common in women than in men. Evidence summarised in Chapter 2 suggests that cancer incidence is higher in men than in women; this may affect the difference in cancer risks reported between men and women within each cohort. However, the difference will apply equally to men and women in the thrombocytosis and normal platelet count cohorts and it is the difference between these two cohorts, rather than

Figure 4.4: Box plot of platelet count at index date in the thrombocytosis and normal platelet count cohorts.

between the two genders, which is under investigation in this thesis.

4.5.5 Platelet counts in the cohort

As expected, the median platelet count was higher in the thrombocytosis cohort than in the normal platelet cohort. The median platelet count in the thrombocytosis cohort was 441 $\times$ 10^9/L (IQR 416-448). In the normal platelet cohort, the median platelet count was 255 $\times$ 10^9/L (IQR 218-299) (see Table 4.1). The boxplots in Figure 4.4 show the distribution of platelet count in each of the cohorts.

In the thrombocytosis cohort, patients were selected based on their raised platelet count, defined as any value over 400 $\times$ 10^9/L. The box plot in Figure 4.4 shows the majority of the thrombocytosis cohort's platelet count being in the 400 - 500 $\times$ 10^9/L range whereas the normal platelet count cohort shows a largely normal distribution grouped around the average platelet count in an adult human (220 - 265 $\times$ 10^9/L) (Balduini & Noris, 2014).

4.5.6 Time interval between thrombocytosis and normal platelet index dates

The index date for patients with thrombocytosis was the date of their first raised platelet count. The index date for the normal platelet count patients was based on the date of their blood test nearest in time to the index date of their matched comparators

from the thrombocytosis cohort. The time interval between the index dates for those with thrombocytosis and their matched normal platelet comparators was examined; bias could have been introduced to the study if the index dates were later in either group, as those patients would be older and therefore more likely to be diagnosed with cancer. The median number of days between the index dates for thrombocytosis patients and their matched normal platelet counterparts was 145 (IQR -110 to 207; from 110 days later in the thrombocytosis patients to 207 days later in the normal platelet count patients); the index date was on average later in patients with a normal platelet count. This time difference of around five months is not likely to have introduced an important degree of bias to the study.

4.5.7 Smoking status

Smoking data were available for 98% of the thrombocytosis cohort. The majority of patients ($n = 17,934$, 57.4%) were either current or ex-smokers (ever smoked). 12,668 (40.5%) had never smoked. Smoking data were not available for 656 (2.1%). Smoking data were available for 99% of patients in the normal platelet count cohort. The majority of these patients had ever smoked ($n = 4,212$, 52.7%) and 3,688 (46.3%) had never smoked. Smoking status was not available for 69 (1%). In both cohorts, the majority of patients were current or ex-smokers, but a slightly greater proportion of the thrombocytosis cohort had ever smoked (57.4% vs 52.7%). A sub-analysis in this chapter stratifies the results by smoking status and smoking is examined as an interaction term in the logistic regression model.

4.5.8 Symptoms prior to diagnosis

The symptoms reported by each group in the month prior to their index date blood test are reported in Table 4.4. The ten most common symptoms in patients in each cohort were listed; eight of these were present in both cohorts. In total, therefore, 12 symptoms are reported in Table 4.4. All are recognised cancer symptoms, but they were reported by relatively small, and similar, proportions. Additionally, all are low risk symptoms which would not generally be expected to prompt urgent cancer investigation NICE (2015). It is unlikely that a great proportion of the blood tests carried out in the thrombocytosis group were triggered by a specific suspicion of cancer, as the symptoms were so similar between the groups.

4.5.9 Geographical region

The geographical region of the practices at which the patients were registered is tabulated in Table 4.5. This shows that some regions of England are underrepresented

Table 4.4: Symptoms reported by patients in the thrombocytosis and normal platelet count cohorts in the 28 days prior to their index date (date of blood test showing first raised platelet count, or equivalent in those with normal platelet counts).

	Thrombocytosis $N = 31,261$		**Normal platelet count** $N = 7,969$	
Symptom	n with symptom	%	n with symptom	%
Cough	1,221	3.9	151	1.9
Fatigue	942	3	205	2.6
Chest infection	935	3	63	0.8
Abdominal pain	826	2.6	133	1.7
Diarrhoea	716	2.3	63	0.8
Joint pain	704	2.3	149	1.9
Back pain	500	1.6	95	1.2
Chest pain	465	1.5	105	1.3
Shortness of breath	360	1.2	47	0.6
UTI	288	0.9	19	0.2
Dizziness	210	0.7	64	0.8
Palpitations	136	0.4	65	0.8

compared to others; the North East, Yorkshire and the Humber, and East Midlands have smaller proportions than the other regions. As the thrombocytosis and normal platelet count patients are practice matched, this difference will not have an effect on the differences in incidence between the two cohorts. However, it does mean that results from this study may be less readily applicable in these under-represented regions.

4.6 The association between thrombocytosis and cancer: results

4.7 One year cancer incidence

The one year cancer incidence was estimated in patients with thrombocytosis and a normal platelet count to address the objective of comparing the incidence between these two cohorts to determine the absolute increase in risk associated with thrombocytosis.

4.7.1 Thrombocytosis cohort

There were 3,050 qualifying cancer diagnoses in the thrombocytosis cohort (excluding non-melanoma skin cancers). The majority ($n = 2,453$, 80.4%) were diagnosed within one year of their index date. 1,098 (44.8%) of those with cancer were male and 1,355 (55.2%) were female. The one year cancer incidence for males in the thrombocytosis

Table 4.5: Number and percentage of patients registered at practices in each geographical region.

Region	n (%)
North East	1,427 (2.9)
North West	8,005 (16.0)
Yorkshire and the Humber	2,097 (4.2)
East Midlands	1,718 (3.4)
West Midlands	6,242 (12.5)
East England	6,176 (12.4)
South West	6,979 (14.0)
South Central	6,103 (12.2)
London	5,310 (10.6)
South East Coast	5,943 (11.9)

cohort was 11.6% (95% CI 11.0-12.3). For females, the one year incidence was 6.2% (95% CI 5.9-6.5). These results are shown in Table 4.6.

4.7.2 Normal platelet count cohort

The one year incidence of cancer was also estimated for patients with a normal platelet count. There were 332 qualifying diagnoses in this cohort. The majority (n = 225, 67.8%) were diagnosed within one year of their index date. 106 (47.1%)of these were in males and 119 (52.9%) were in females. The one year cancer incidence for males with a normal platelet count was 4.1% (3.4-4.9). For females, the one year incidence was 2.2% (1.8-2.6).

4.7.3 Comparison of one year cancer incidence between cohorts

These results show that the one year incidence of cancer is higher in patients with thrombocytosis, compared to those with a normal platelet count. The results here and for the rest of this chapter are presented separately for men and women, due to the difference in the magnitude of the association between these two sub-groups. In all patients with thrombocytosis the risk in male patients was almost double that in female patients. In patients with a normal platelet count the difference in risk between the two sexes is also roughly doubled. There is a similar magnitude of difference in risk between men and women with thrombocytosis and men and women with normal platelet counts. This suggests the higher cancer incidence seen in men is not mediated by differences in the relationship between platelets and cancer in men and women, but by other factors such as variance in consulting behaviour (women consult more often than men) or underlying cancer incidence (which is higher in men than in women), or more benign

causes of thrombocytosis in women than in men. If there was a relationship between gender and the platelet-cancer interaction, then we could expect to see a difference in the magnitude of risk between men and women with thrombocytosis, and men and women with a normal platelet count. This is explored further in Chapter 6.

Table 4.6: The one year cancer incidence for patients with thrombocytosis and patients with a normal platelet count. N: number, CI: confidence interval.

Time from index date (months)	**Thrombocytosis**				**Normal platelet count**			
	Men		**Women**		**Men**		**Women**	
	N	n cancers diagnosed, incidence % (95% CI)	N	n cancers diagnosed, incidence % (95% CI)	N	n cancers diagnosed, incidence % (95% CI)	N	n cancers diagnosed, incidence % (95% CI)
1-12	9,435	1,098 11.6 (11.0-12.3)	21,826	1,355 6.2 (5.9-6.5)	2,599	106 4.1 (3.4-4.9)	5,370	119 2.2 (1.8-2.6)

4.8 Epidemiological measures

The relationship between thrombocytosis and cancer was investigated using logistic regression to examine cancer occurrence (outcome) in relation to thrombocytosis as a binary predictor. In a crude (unadjusted) model, patients with thrombocytosis were more likely to be diagnosed with cancer than those with normal platelet counts (OR 2.9, 95% CI 2.6-3.4, $p < 0.0001$). Logistic regression was used to examine the interaction between patients' smoking status and platelet count with respect to cancer incidence. There was little evidence of an interaction between smoking and thrombocytosis ($p = 0.94$).

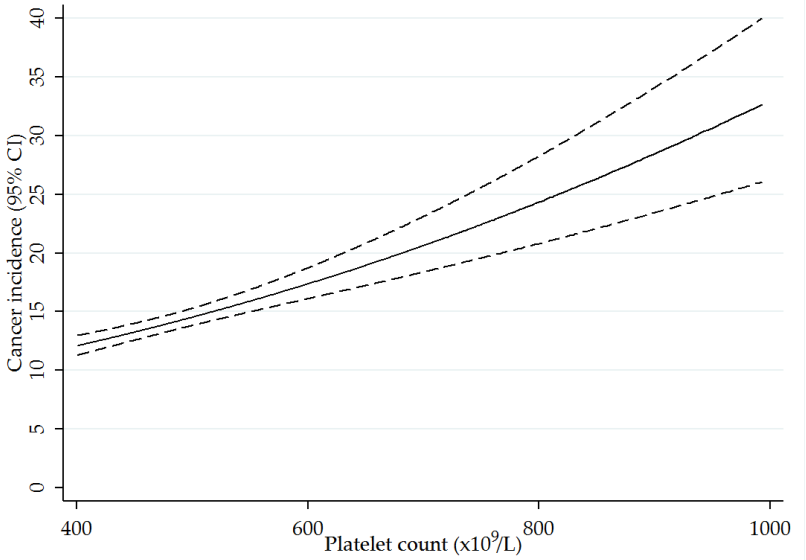
Cancer occurrence was also examined in relation to platelet count as a continuous variable (see Figure 4.5). The one year cancer incidence for the entire cohort of patients, including those with thrombocytosis and normal platelet count, is shown in Figure 4.5 for men (a) and women (b). Both show a steady increase in the risk of cancer (and corresponding confidence intervals) with increasing platelet count. The graph for male patients (a) shows an increasing risk of cancer with increasing platelet count. In contrast, the graph for female patients (b) shows a convex profile in the increase in risk of cancer with increasing platelet count, suggesting that the effect is limited above platelet counts of $1,000 \times 10^9$/L.

4.9 Diagnostic interval for one year cancer incidence

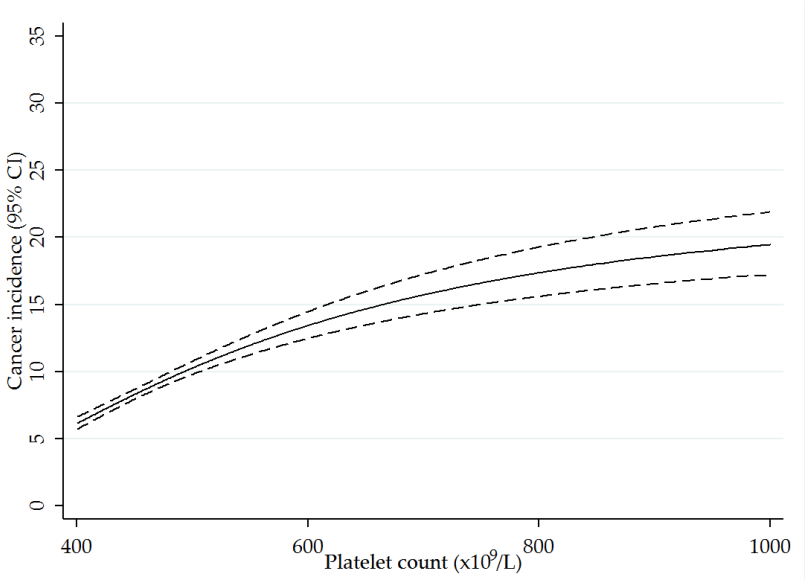
The diagnostic interval was calculated for each cohort; this is the number of days between the patient's index date and their recorded cancer diagnosis date, summarised using the median and interquartile range. The diagnostic interval was also examined graphically, to compare the distribution between the two cohorts. The results that follow are based on the one year cancer incidence. For this analysis the diagnostic interval is defined as the time between a blood test result and the patients' cancer being diagnosed.

4.9.1 Thrombocytosis cohort

The median number of days from index date to diagnosis in this cohort was 33 (IQR 2-106) and ranged from 1 to 353 days. The diagnostic interval in this cohort of patients is shown in Figure 4.6 (a). The histogram shows that the majority of cancers in this cohort were diagnosed within the first 100 days after diagnosis.



(a)



(b)

Figure 4.5: (a) fractional polynomial logistic regression model with platelet count as a continuous predictor variable for men aged over 40 years. (b) fractional polynomial logistic regression model with platelet count as a continuous predictor variable for women aged over 40 years.

4.9.2 Normal platelet count cohort

The median number of days from index date to diagnosis in this cohort was 50 (IQR 21-129) and ranged from 1 to 365 days. The diagnostic interval in this group of patients is shown in Figure 4.6 (b). This histogram also shows the majority of cancer diagnoses in this cohort are recorded within 100 days of the blood test date.

4.9.3 Comparison of diagnostic intervals between cohorts

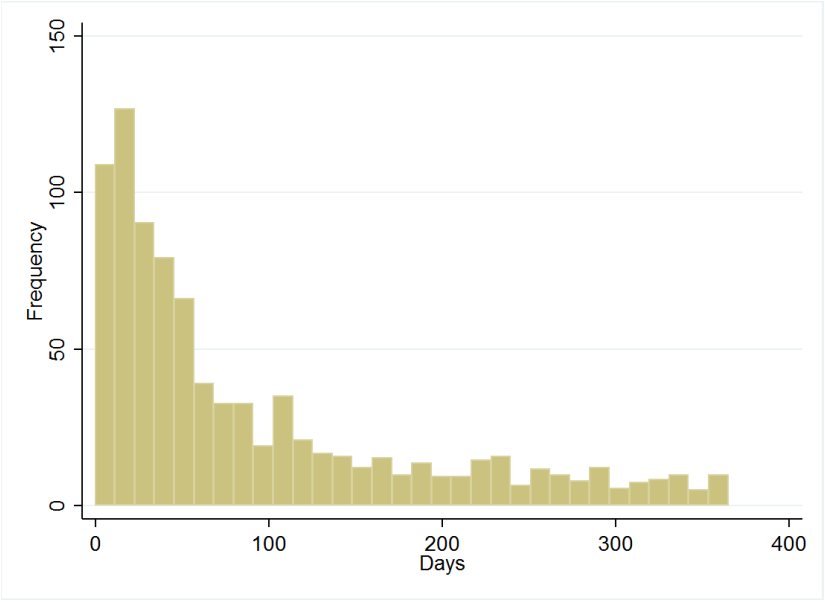
When comparing the histograms of the diagnostic interval in days between index date and date of cancer diagnosis for patients with thrombocytosis and normal platelets, both show most diagnoses being recorded within the first 100 days after the blood test result. The median number of days to diagnosis was greater by 17 days in the normal platelet count cohort. One potential limitation of this study is that, in patients with thrombocytosis, blood tests may have been ordered by general practitioners who already suspected malignancy in the patients for other reasons. The similar pattern of interval prior to diagnosis in each cohort suggest that cancer diagnoses were made slightly earlier in the thrombocytosis cohort compared to the normal platelet count cohort. A sub-analysis presented in the next section of this chapter compares the proportion of new diagnoses made 1-3 and 4-12 months after index date on the assumption that cancers in the latter time period were unlikely to be have been suspected at the time of the blood test.

4.10 Medium term risk of cancer

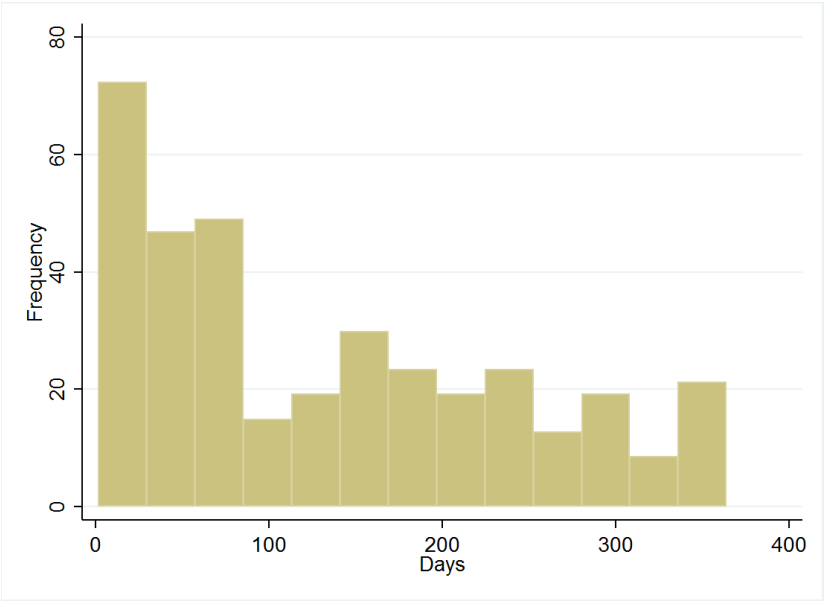
This sub-analysis split the one year incidence of cancer in to two time periods; the first three months and the last nine months after index date. Of the $2,453$ new cancer diagnoses made in the thrombocytosis cohort, $1,610$ (65.6%) were made in the 1-3 month time period and 843 (34.4%) were made within the 4-12 month time period. Thus, around one third of cancers in patients with thrombocytosis are diagnosed more than three months after a blood test result showing raised platelets.

4.11 One year cancer incidence by age group

The relationship between age and cancer incidence is estimated here with an analysis stratified by age group. The risk of most types of cancer increases with age. The data were examined to determine whether the risk of cancer in patients with thrombocytosis increases with age in line with or beyond the increase that would be expected with age, as measured against patients with normal platelet counts.



(a)



(b)

Figure 4.6: Histograms to show the number of days between patient's index date and their cancer diagnosis for (a) thrombocytosis cohort and (b) normal platelet count cohort.

Table 4.7: Table to show the number and percentage of patients in each age group in the thrombocytosis and normal platelet count cohorts, and the number and percentage of these with a new cancer diagnosis within a year of index date age brackets and the number and percentage of men and women in each age group is compared.

Age group (years)	**Thrombocytosis** $N = 31,261$		**Normal platelet count** $N = 7,969$	
	n in group	n diagnosed, % (95% CI)	n in group	n diagnosed, % (95% CI)
40-49	3,575	78, 2.2 (1.7-2.7)	786	10, 1.3 (0.9-2.1)
50-59	6,251	300, 4.8 (4.3-5.3)	1,561	25, 1.6 (1.0-2.2)
60-69	7,297	642, 8.8 (8.1-9.4)	1,939	52, 2.7 (2.0-3.4)
70-79	7,744	826, 10.7 (10.0-11.4)	1,972	76, 3.9 (3.0-4.7)
≥ 80	6,394	607, 9.5 (8.8-10.2)	1,711	62, 3.6 (2.7-4.5)

4.11.1 Thrombocytosis cohort

The risk of malignancy increased with age in the thrombocytosis cohort. The one year cancer incidence increased from 2.2% in patients aged 40-49 years and peaked at 10.7% in those aged 70-79 years (see Table 4.7). The scatter plot in Figure 4.7 shows the one year cancer incidence by age group for male and female patients with thrombocytosis separately. In those aged 40-49 years and 50-59 years there was no difference in the risk between males and females. However, with increasing age, the difference between males and females became more pronounced. Men were at significantly higher risk than women from age 60 years onwards. The line of best fit for men with thrombocytosis shows a continued increase in later years compared to female patients for whom the risk appears to decline slightly with age over about 80 years. UK cancer incidence figures discussed in Chapter 2 show a decrease in new diagnoses in this age group.

4.11.2 Normal platelet count cohort

For patients with a normal platelet count, the cancer risk also increases with age, similarly to national statistics. The risk increases from 1.3% in those aged 40-49 and peaks at 3.9% in those aged 70-79 (see Table 4.7). Although the risk increases with each decade of age, the actual increase from one group to the next is small. The scatter plot in Figure 4.8 shows the one year cancer incidence for male and female patients with a normal platelet count separately. While the plot shows an upwards trend, the line of best fit suggests that cancer incidence continues to rise in men, but does not rise further above a certain age in women.

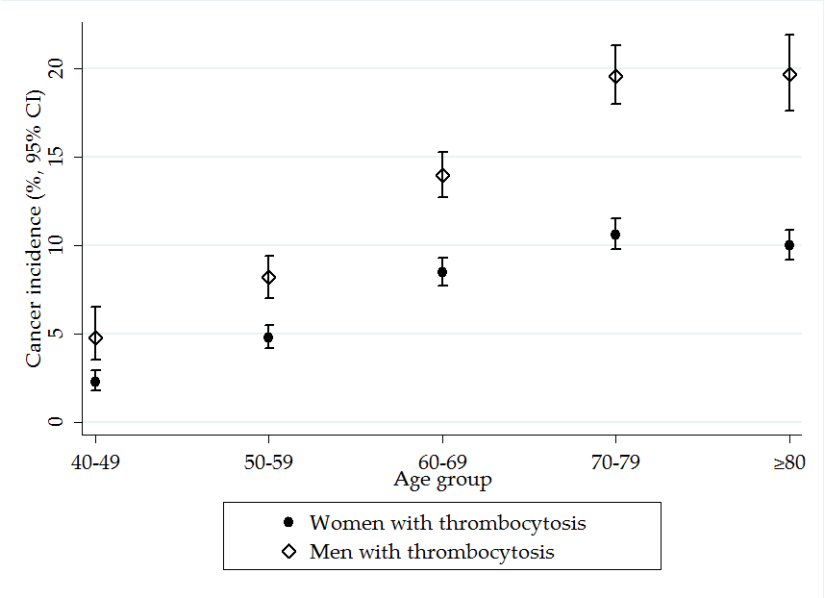


Figure 4.7: Scatter plot to show the risk of cancer diagnosis for men and women with thrombocytosis in five age groups.

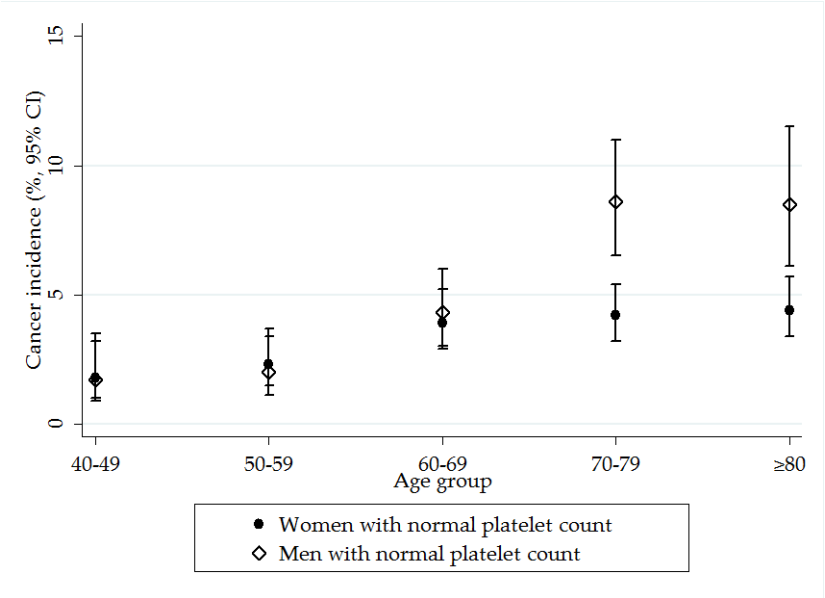
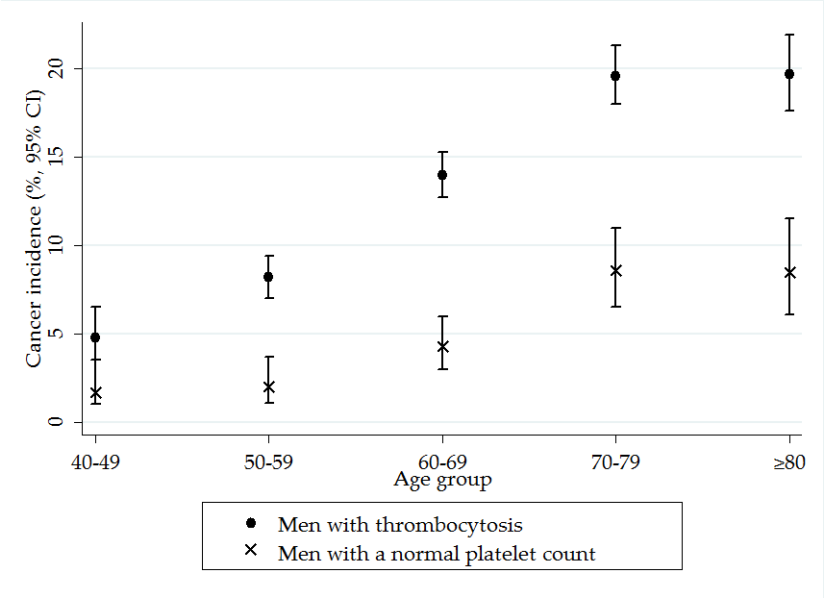
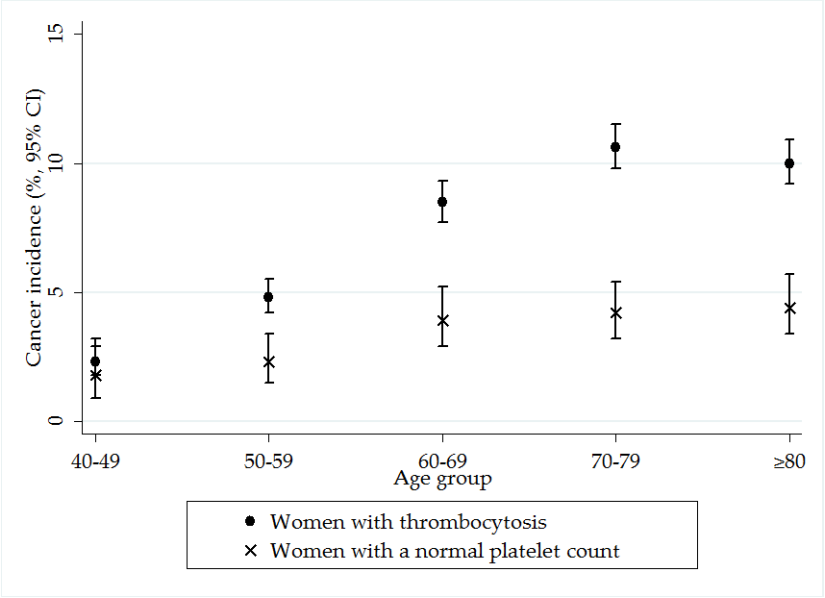


Figure 4.8: Scatter plot to show the risk of cancer diagnosis for men and women with a normal platelet count in five age groups.



(a)



(b)

Figure 4.9: The percentage of men (a) and women (b) with thrombocytosis and normal platelet counts in different age groups diagnosed with cancer within one year of a platelet count record, with 95% confidence intervals.

4.11.3 Comparison of risk with age between cohorts

The risk of cancer increased with age in both cohorts but those with thrombocytosis were at consistently greater risk than those with normal platelet counts across all ages, except for those in their 40s. The greatest difference in risk between thrombocytosis and normal platelet was in those aged 60-69 years. The confidence intervals for each group suggest that there is a difference in risk between the two groups in those aged 50 and over. These results are presented graphically in Figure 4.8; this shows the cancer risk increasing with age in both thrombocytosis and normal platelet cohorts, but the steeper gradient in patients with thrombocytosis indicates that the risk increases with age at a greater rate than in those with a normal platelet count. These differences are likely to be due to differences in platelet count and incidence with age, not differences in the biology of the interaction between cancer and platelets between patients of different ages.

4.11.4 Cancer incidence stratified by smoking status

Patients in each of the two cohorts were grouped based on whether they had ever smoked (current or past, or never smoked) and the incidence of cancer was compared between the two groups. For men with thrombocytosis, the cancer incidence among those who had ever smoked was 11.8 % (95% CI 11.1-12.6), and the incidence in those who had never smoked was 9.9% (95% CI 8.7-11.2).

For women with thrombocytosis, the cancer incidence among those who had ever smoked was 6.4% (95% CI 5.9-6.9), and the incidence in those who had never smoked was 5.7% (95% CI 5.3-6.2).

4.12 One year cancer incidence by primary cancer site

The following analysis will address the objective of determining whether some types of cancer are more likely to be diagnosed in patients with thrombocytosis compared to patients with a normal platelet count.

The type of cancer (primary cancer site) diagnosed in patients with thrombocytosis and with normal platelet counts is described below. All cancer diagnoses in the cohort were categorised into 20 groups based on site of diagnosis - these are described in the methods section 4.4.7.

Table 4.8: Cancers diagnosed in patients with thrombocytosis (TH) and a normal platelet count (NPC) within one year of index date. Each column shows the number of each type of cancer, the percentage of all cancers in this group, and the incidence of that type of cancer in that group, for men and women. Each incidence figure excludes patients diagnosed with any of the other types of cancer. n = number of cancer diagnoses recorded for that group. % = percentage of all cancers diagnosed in that group.

	1^{st} year after index date						1^{st} year after index date					
	Men with TH			**Men with NPC**			**Women with TH**			**Women with NPC**		
Cancer site	n	%	Incidence % (95% CI)	n	%	Incidence % (95% CI)	n	%	Incidence % (95% CI)	n	%	Incidence % (95% CI)
Bladder	45	4.1	0.6 (0.4-0.7)	11	10.4	0.4 (0.2-0.8)	36	2.7	0.2 (0.1-0.2)	3	2.5	0.1 (0.0-0.2)
Brain	8	0.73	0.3 (0.2-0.5)	2	1.9	0.2 (0.1-0.5)	10	0.7	0.2 (0.1-0.2)	3	2.5	0.1 (0.0-0.2)
Breast	-	-	-	-	-	-	77	5.7	0.4 (0.3-0.5)	23	19.3	0.5 (0.3-0.7)
Cervix	-	-	-	-	-	-	10	0.7	0.1 (0.0-0.1)	4	3.4	0.1 (0.0-0.2)
Colorectal	226	20.6	2.7 (2.4-3.1)	10	9.4	0.4 (0.2-0.7)	317	23.4	1.6 (1.4-1.7)	22	18.5	0.4 (0.3-0.6)
Kidney	30	2.7	0.4 (0.2-0.5)	1	0.9	0.0 (0.0-0.2)	37	2.7	0.2 (0.1-0.3)	2	1.7	0.0 (0.0-0.1)
Leukaemia	20	1.8	0.3 (0.2-0.3)	4	3.8	0.2 (0.1-0.2)	29	2.1	0.1 (0.1-0.2)	2	1.7	0.0 (0.0-0.1)
Lung	273	24.9	3.2 (2.9-3.6)	11	10.4	0.4 (0.2-0.7)	220	16.3	1.1 (0.9-1.2)	13	10.9	0.3 (0.1-0.4)
Lymphoma	31	2.8	0.4 (0.3-0.5)	3	2.8	0.1 (0.0-0.4)	34	2.5	0.2 (0.1-0.2)	3	2.5	0.1 (0.0-0.2)
Melanoma	7	0.6	0.1 (0.0-0.2)	2	1.9	0.1 (0.0-0.3)	16	1.2	0.1 (0.0-0.1)	1	0.8	0.0 (0.0-0.1)
Myeloma	11	1	0.1 (0.1-0.2)	0	0	0	11	0.8	0.1 (0.0-0.1)	1	0.8	0.0 (0.0-0.1)
Oesophagus	43	3.9	0.5 (0.4-0.7)	5	4.7	0.2 (0.1-0.5)	22	1.6	0.1 (0.1-0.2)	1	0.8	0.0 (0.0-0.1)
Oral	14	1.3	0.2 (0.1-0.2)	1	0.9	0.0 (0.0-0.2)	4	0.3	0.0 (0.0-0.1)	1	0.8	0.0 (0.0-0.1)

Cancer site	1^{st} year after index date						1^{st} year after index date					
	Men with TH			**Men with NPC**			**Women with TH**			**Women with NPC**		
	n	%	Incidence % (95% CI)	n	%	Incidence % (95% CI)	n	%	Incidence % (95% CI)	n	%	Incidence % (95% CI)
Other	177	16.1	2.0 (1.7-2.3)	19	17.9	0.7 (0.4-1.2)	290	21.4	1.4 (1.2-1.5)	17	14.3	0.3 (0.2-0.5)
Ovary	-	-	-	-	-	-	105	7.8	1.3 (1.1-1.4)	5	4.2	0.8 (0.7-0.9)
Pancreas	39	3.6	0.5 (0.3-0.7)	6	5.7	0.2 (0.1-0.5)	47	3.5	0.2 (0.2-0.3)	4	3.4	0.1 (0.0-0.2)
Prostate	120	10.9	1.4 (1.2-1.7)	28	26.4	1.1 (0.7-1.6)	-	-	-	-	-	-
Stomach	45	4.1	0.7 (0.5-0.9)	3	2.8	0.1 (0.0-0.4)	45	3.3	0.2 (0.2-0.3)	3	2.5	0.1 (0.0-0.2)
Testis	2	18	0.0	0	0	0	-	-	-	-	-	-
Uterus	-	-	-	-	-	-	40	3.0	0.2 (0.2-0.3)	9	7.6	0.2 (0.1-0.2)
All cancers	1,098	100	11.6 (11.0-12.3)	106	100	2.7 (2.4-3.1)	1,355	100	6.2 (5.9-6.5)	119	100	1.8 (1.6-2.0)

4.12.1 Thrombocytosis cohort

For men with thrombocytosis, there were 1,098 cancers diagnosed in the first year after their index date. The incidence of each type of cancer diagnosed is shown by primary cancer site in Table 4.8. The three most commonly diagnosed cancers in men in the general population are lung, colorectal, and prostate cancer - this was also the case for men with thrombocytosis. The one year incidence of lung cancer was 3.2% (95% CI 2.9-3.6). For colorectal cancer, it was 2.7% (95% CI 2.4-3.1) and for prostate cancer it was 1.4% (95% CI 1.2-1.7). The proportion of each type of cancer diagnosed out of all cancers is shown in the pie chart in Figure 4.10. The most commonly diagnosed cancers were lung ($n = 273$, 24.9% of cancers diagnosed); colorectal ($n = 226$, 20.6%); and prostate ($n = 120$, 10.9%). There were fewer cases in the next most commonly diagnosed cancers: bladder ($n = 45$, 4.1%) and stomach ($n = 45$, 4.1%).

For female patients with thrombocytosis, there were 1,355 cancers diagnosed in the first year after index date. The one year incidence of each type of cancer diagnosed is shown by primary cancer site in Table 4.8. For lung cancer this was 1.1% (95% CI 0.9-1.2). For colorectal cancer, it was 1.6% (95% CI 1.4-1.7) and for breast cancer it was 0.4% (95% CI 0.3-0.5). The proportion of each type of cancer diagnosed out of all cancers in women with thrombocytosis is shown in the pie chart in Figure 4.11. In this sub-group, the most common diagnosed cancers were colorectal ($n = 317$, n=23.4%); lung ($n = 220$, 16.2%); and ovarian ($n = 105$, 7.8%).

4.12.2 Normal platelet count cohort

For men with a normal platelet count, there were 106 cancers diagnosed in the first year after their index date. The incidence of each type of cancer diagnosed is shown by primary cancer site in Table 4.9. For men with a normal platelet count, the one year incidence of the most commonly diagnosed cancers was 0.4% (95% CI 0.2-0.8) for lung cancer. For colorectal cancer, it was 0.4% (95% CI 0.2-0.7) and for prostate cancer it was 1.1% (95% CI 0.7-1.6). When each type of cancer was reported as the proportion of all cancers diagnosed, the most commonly diagnosed cancers were prostate ($n = 28$, 26.4 of all diagnosed cancers in this group); lung ($n = 11$, 10.4%); bladder ($n = 11$, 10.4%); and colorectal ($n = 10$, 9.4%) (see Figure 4.12).

For female patients with a normal platelet count, there were 119 cancers diagnosed in the first year after index date. The incidence of each type of cancer diagnosed is shown by primary cancer site in Table 4.9. The one year incidence of lung cancer was 0.3% (95% CI 0.1-0.4). For colorectal cancer, it was 0.4% (95% CI 0.3-0.6) and for breast cancer it was 0.5% (95% CI 0.3-0.7). When each type of cancer was reported as the proportion of all cancers diagnosed, the most commonly diagnosed cancers were

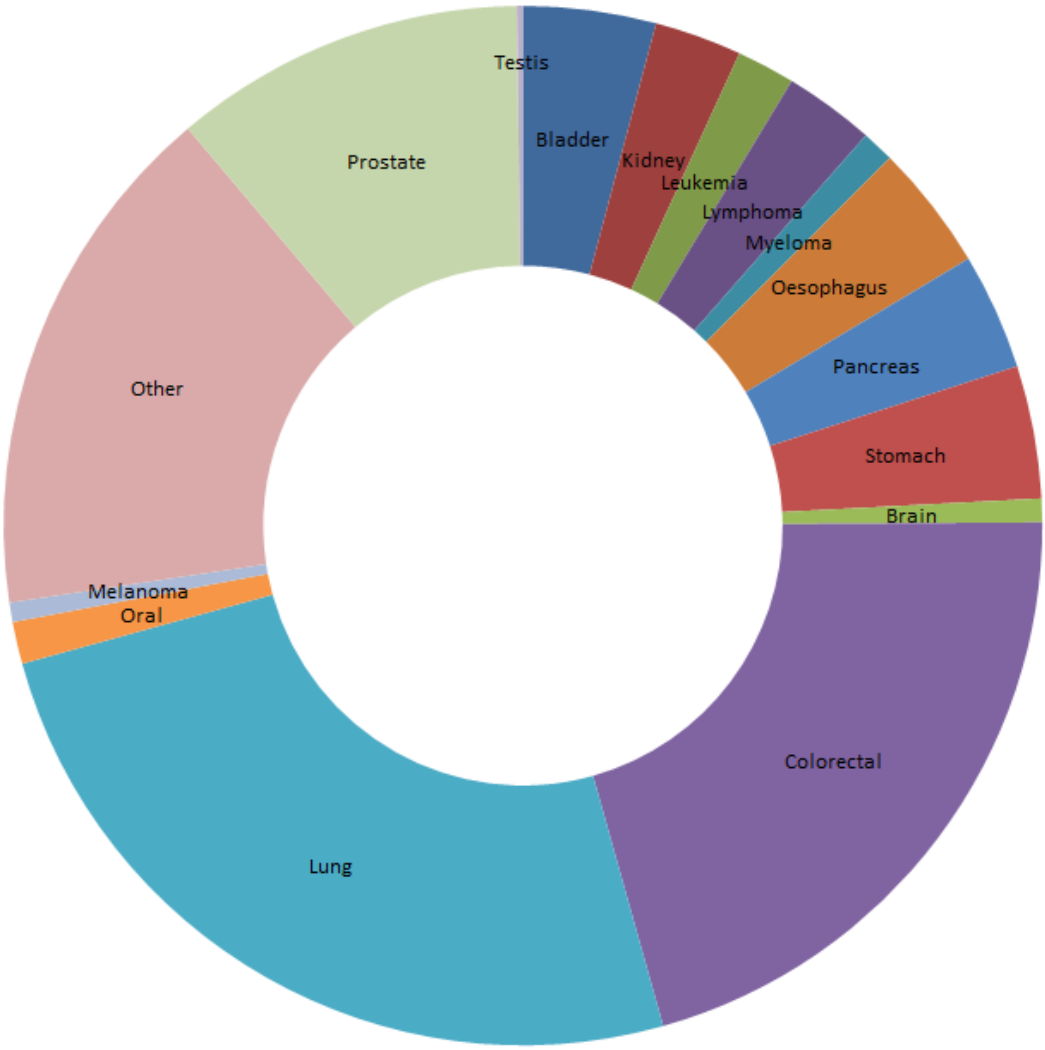


Figure 4.10: Pie chart to show the proportion of different types of cancer diagnosed in men with thrombocytosis within one year of index date (date of first raised platelet count). Corresponding values available in Table 4.8.

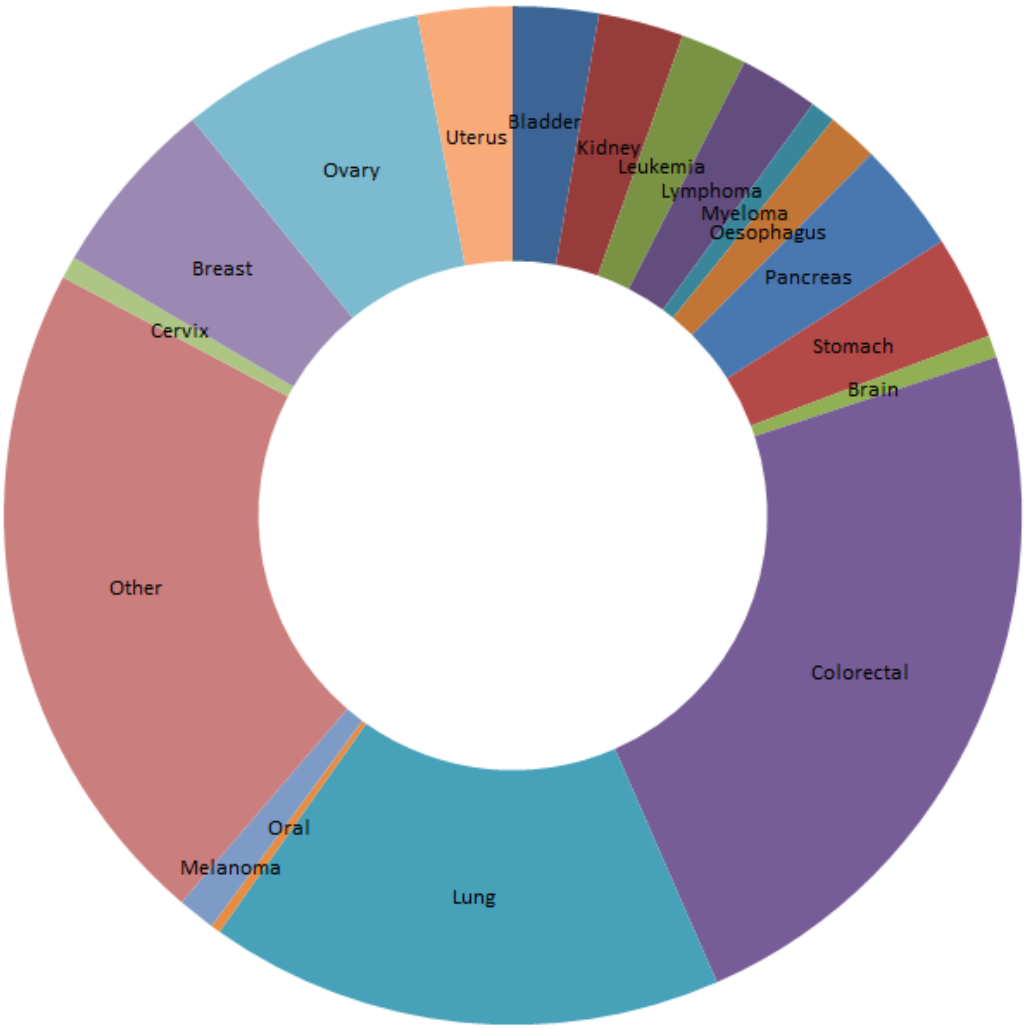


Figure 4.11: Pie chart to show the proportion of different types of cancer diagnosed in women with thrombocytosis within one year of index date (date of first raised platelet count). Corresponding values available in Table 4.8.

breast ($n = 23$, 19.3%); colorectal ($n = 22$, 18.5); lung ($n = 13$, 10.9%); uterine ($n = 9$, 7.6%); ovarian ($n = 5$, 4.2%). The percentage of cancers diagnosed at each site in female patients is presented in the pie chart (see Figure 4.13).

Table 4.9: Cancers diagnosed in patients with thrombocytosis (TH) and a normal platelet count (NPC) within the second year after index date. Each column shows the number of each type of cancer, the percentage of all cancers in this group, and the incidence of that type of cancer in that group, for men and women. Each incidence figure excludes patients diagnosed with any of the other types of cancer, and patients diagnosed in the first year. % =percentage of all cancers diagnosed in that group.

	2^{nd} **year after index date**						2^{nd} **year after index date**					
	Men with TH			**Men with NPC**			**Women with TH**			**Women with NPC**		
Cancer site	n	%	Incidence % (95% CI)	n	%	Incidence % (95% CI)	n	%	Incidence % (95% CI)	n	%	Incidence % (95% CI)
Bladder	13	5.8	0.2 (0.1-0.3)	3	8.6	0.1 (0.0-0.4)	11	3	0.1 (0.0-0.1)	1	1.4	0.0 (0.0-0.1)
Brain	0	0	0	1	2.9	0.2 (0.0-0.4)	6	1.6	0.1 (0.1-0.2)	0	0	0
Breast	-	-	-	-	-	-	57	15.3	0.3 (0.2-0.4)	17	23.6	0.3 (0.2-0.5)
Cervix	-	-	-	-	-	-	4	1.1	0.0 (0.0-0.1)	3	4.1	0.1 (0.0-0.1)
Colorectal	30	13.4	0.4 (0.2-0.5)	0	0	0	54	14.5	0.3 (0.2-0.3)	7	9.7	0.1 (0.0-0.3)
Kidney	8	3.6	0.1 (0.0-0.2)	1	2.9	0.0 (0.0-0.2)	7	1.9	0.0 (0.0-0.1)	1	1.4	0.0 (0.0-0.1)
Leukaemia	11	4.9	0.1 (0.0-0.1)	1	2.9	0.0 (0.0-0.2)	21	5.6	0.2 (0.1-0.2)	4	5.6	0.1 (0.0-0.1)
Lung	45	20.1	0.6 (0.4-0.7)	5	14.3	0.2 (0.0-0.5)	35	9.4	0.2 (0.1-0.2)	8	11.1	0.2 (0.0-0.3)
Lymphoma	6	2.7	0.1 (0.0-0.2)	1	2.9	0.0 (0.0-0.2)	7	1.9	0.0 (0.0-0.1)	1	1.4	0.0 (0.0-0.1)
Melanoma	3	1.3	0.0 (0.0-0.1)	0	0	0	14	3.8	0.1 (0.0-0.1)	3	4.2	0.1 (0.0-0.2)
Myeloma	3	1.3	0.0 (0.0-0.1)	0	0	0	4	1.1	0.0 (0.0-0.1)	1	1.4	0.0 (0.0-0.1)
Oesophagus	7	3.1	0.1 (0.0-0.2)	2	5.7	0.1 (0.0-0.3)	7	1.9	0.0 (0.0-0.1)	2	2.8	0.0 (0.0-0.1)
Oral	2	0.9	0.0 (0.0-0.1)	3	8.6	0.0 (0.0-0.3)	7	1.9	0.0 (0.0-0.1)	3	4.2	0.0 (0.0-0.1)

Cancer site	2^{nd} year after index date						2^{nd} year after index date					
	Men with TH			**Men with NPC**			**Women with TH**			**Women with NPC**		
	n	%	Incidence % (95% CI)	n	%	Incidence % (95% CI)	n	%	Incidence % (95% CI)	n	%	Incidence % (95% CI)
Other	40	17.9	0.5 (0.3-0.6)	10	28.6	0.4 (0.2-0.7)	93	25	0.4 (0.4-0.5)	12	16.7	0.2 (0.1-0.4)
Ovary	-	-	-	-	-	-	14	3.8	0.4 (0.3-0.6)	1	1.3	0.3 (0.2-0.5)
Pancreas	5	2.2	0.1 (0.0-0.1)	1	2.9	0.0 (0.0-0.2)	11	3	0.1 (0.0-0.1)	5	6.9	0.1 (0.0-0.2)
Prostate	45	20.1	0.6 (0.4-0.7)	5	14.3	0.2 (0.1-0.5)	-	-	-	-	-	-
Stomach	5	2.2	0.1 (0.1-0.2)	1	2.9	0.2 (0.1-0.5)	8	2.1	0.3 (0.2-0.4)	0	0	0.1 (0.0-0.3)
Testis	0	0	0	0	0	0	-	-	-	-	-	-
Uterus	-	-	-	-	-	-	9	2.4	0.3 (0.2-0.4)	3	4.1	0.1 (0.0-0.1)
All cancers	224	100	4.1 (3.4-4.9)	35	100	1.4 (1.0-1.9)	373	100	2.2 (1.8-2.6)	72	100	1.4 (1.1-1.7)

4.12.3 Comparison of primary cancer site between cohorts

In terms of cancer incidence, the difference in risk between patients with thrombocytosis and a normal platelet count was most pronounced for lung and colorectal cancer. The one year incidence of lung cancer in men with thrombocytosis was 3.2% (95% CI 2.9-3.6); significantly higher than the value of 0.4% (95% CI 0.2-0.7) in men with a normal platelet count. The added risk posed by thrombocytosis, or the absolute increase in risk, is 2.8%. Similarly, the absolute increase in risk of colorectal cancer in patients with thrombocytosis compared to those with a normal platelet count was 2.3%; an increase from 0.4% (95% CI 0.2-0.7) in patients with a normal platelet count to 2.7% (95% CI 2.4-3.1) in patients with thrombocytosis. However, for prostate cancer, there was little evidence of a difference in incidence between patients with thrombocytosis and patients with a normal platelet count (thrombocytosis: 1.4% (95% CI 1.2-1.7); normal platelet count 1.1% (95% CI 0.7-1.6).

In males, lung and colorectal cancer were much more commonly diagnosed in patients with thrombocytosis than in patients with a normal platelet count (lung: 25% of incidence cases vs 10%; colorectal: 21% vs 9%). Conversely, prostate cancer was much less frequently diagnosed in patients with thrombocytosis compared to those with a normal platelet count (11% of incidence cases in men with thrombocytosis, 26% in men with a normal platelet count).

In female patients, there was a smaller increase in the risk of lung and colorectal in those with thrombocytosis compared to those with a normal platelet count. The risk of lung cancer increased from 0.3% (95% CI 0.1-0.4) in patients with a normal platelet count to 1.1% (95% CI 0.9-1.2) in patients with thrombocytosis; an absolute increase of 0.8%. For colorectal cancer, there was a 1.1% increase from 0.4% (95% CI 0.3-0.6) in patients with a normal platelet count to 1.6% (95% CI 1.4-1.7) in patients with thrombocytosis. The risk of breast cancer was no higher in patients with thrombocytosis; 0.4% (95% CI 0.3-0.5) in patients with thrombocytosis and 0.5% (95% CI 0.3-0.7) in patients with a normal platelet count.

For female patients, colorectal and lung cancer had a higher incidence and were more commonly diagnosed in females with thrombocytosis than in those with a normal platelet count. Lung cancer accounted for 16% of incident cases in women with thrombocytosis vs 11% in women with a normal platelet count. Colorectal cancer accounted for 23% of cases in women with thrombocytosis and 19% in women with a normal platelet count. Although breast cancer incidence did not differ widely between patients with thrombocytosis and a normal platelet count, breast cancer accounted for a much smaller proportion of all incidence cases in women with thrombocytosis; 6% vs 19% of all incident cases. This could be misinterpreted as thrombocytosis protecting

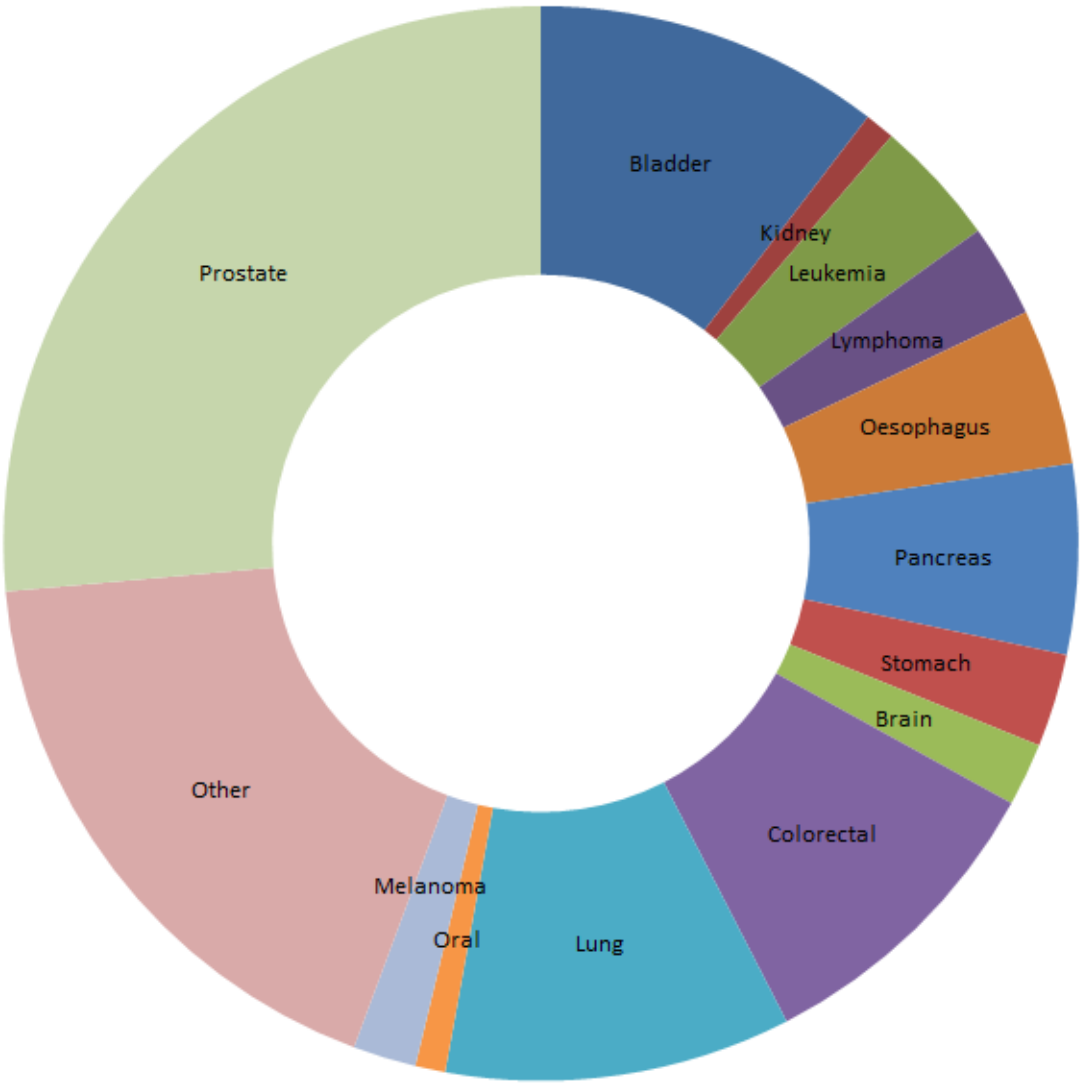


Figure 4.12: Pie chart to show the proportion of different types of cancer diagnosed in men with a normal platelet count within one year of index date (date of first raised platelet count). Corresponding values available in Table 4.9.

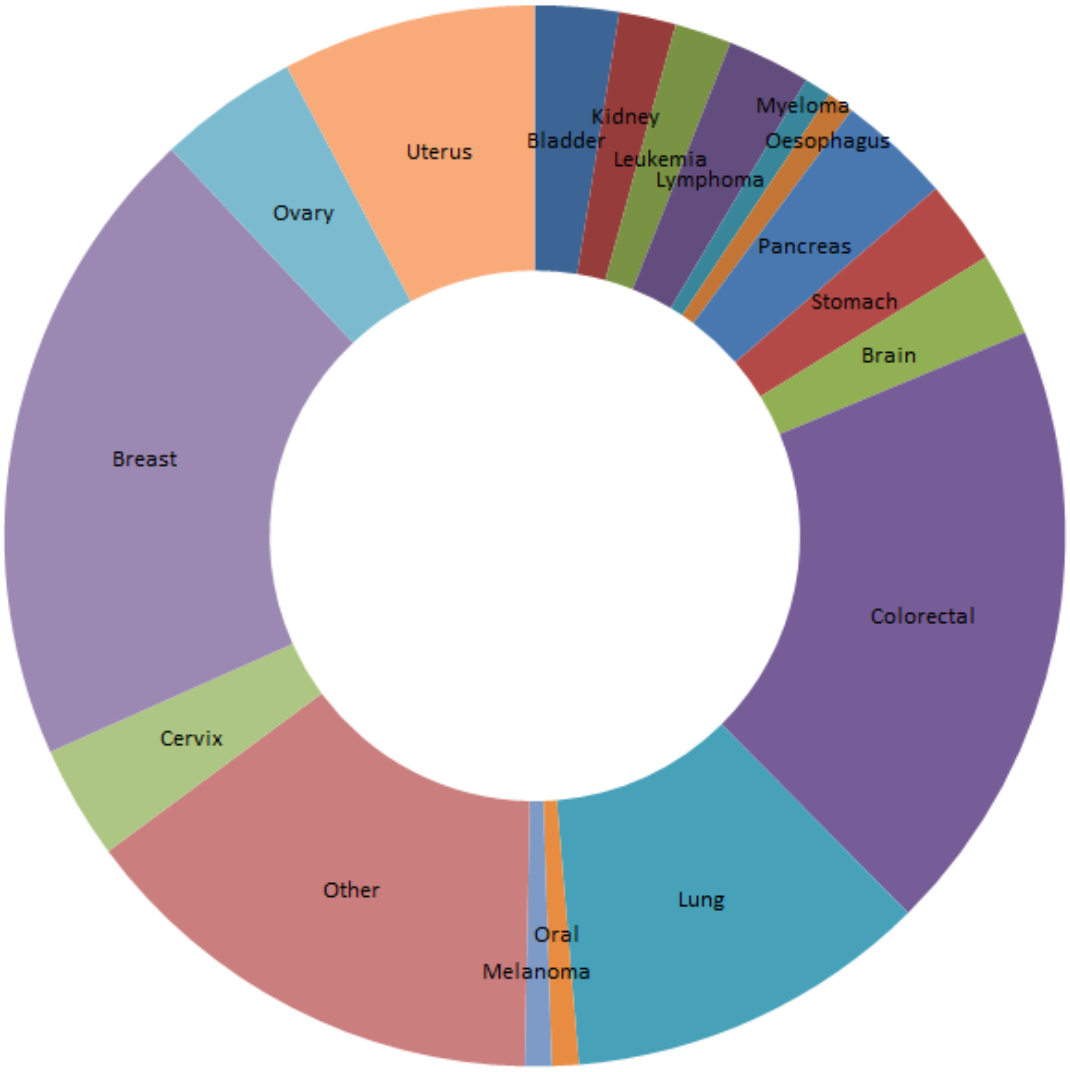


Figure 4.13: Pie chart to show the proportion of different types of cancer diagnosed in women with a normal platelet count within one year of index date (date of first raised platelet count). Corresponding values available in Table 4.9.

against breast cancer; this is not the case. Breast cancer is less commonly diagnosed in patients with thrombocytosis compared to those with a normal platelet count, and this analysis reports the percentage of each type of cancer diagnosed out of all diagnoses in each cohort.

The differences described here suggest that thrombocytosis is more strongly associated with some types of cancer than others. Not only are patients with thrombocytosis more likely to be diagnosed with cancer, but they are more likely to be diagnosed with lung or colorectal cancer than a patient with a normal platelet count. Breast and prostate cancer are less commonly diagnosed in patients with thrombocytosis compared to patients with a normal platelet count. The evidence reviewed in Chapter 2 found that the lungs are an important site for platelet production; this could be part of the explanation for the stronger relationship between thrombocytosis and lung cancer compared to other cancer sites. Colorectal cancer is often associated with bleeding; a condition which usually induces the production of platelets. For other types of cancer, incidence was higher in patients with thrombocytosis than in patients with a normal platelet count - specifically for stomach and kidney cancer and for lymphoma. For other cancer types, the difference was less pronounced. However, for these types of cancer there were few incident cases in either cohort; a larger sample size may have revealed larger differences in incidence.

4.12.4 Comparison of cancer site between the thrombocytosis cohort and the general population

The cancers diagnosed in the thrombocytosis cohort were compared to national cancer incidence data showing the most common types of cancer in the general population. The pie chart in Figure 4.14 shows the proportion of each type of cancer diagnosed in men with thrombocytosis in the outer ring, and the incidence of the most common cancer types in the general population in the inner ring (image from CRUK).

Lung and colorectal cancer were most commonly diagnosed in the thrombocytosis group and were diagnosed more commonly in these patients than in the general population. Prostate cancer was the most commonly diagnosed cancer in the general population, accounting for a quarter of all cancers in men over 40 years of age. Bowel and lung cancer are the second most commonly diagnosed, accounting for 14% of diagnoses each.

In females with thrombocytosis, colorectal was the most commonly diagnosed cancer (23% of diagnoses) followed by lung (16%) and breast (6%) (see Figure 4.15). In the general population, breast is the most commonly diagnosed cancer in women (30% of all cases). Colorectal and lung are the second most commonly diagnosed (11% and

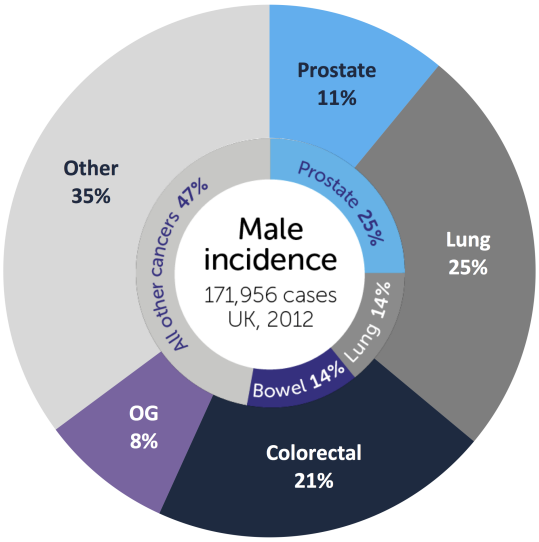


Figure 4.14: Pie chart to show commonly diagnosed cancer types in men. Outer ring shows incidence of cancers diagnosed at different sites in men with thrombocytosis. Inner ring shows incidence of cancer types in the general population. OG: oesophago-gastric (image and values from Cancer Research UK).

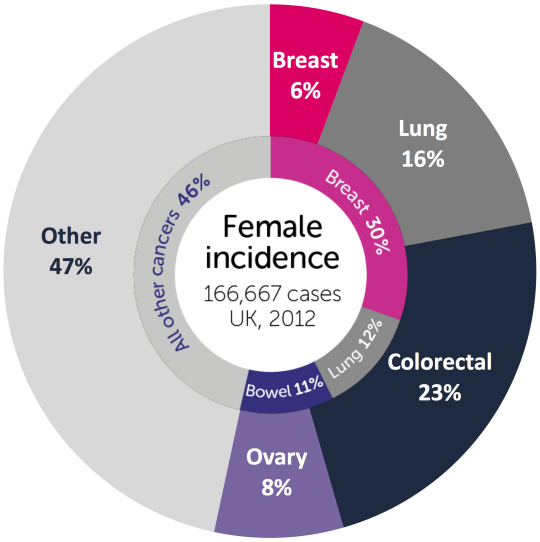


Figure 4.15: Pie chart to show commonly diagnosed cancer types in women. Outer ring shows incidence of cancers diagnosed at different sites in men with thrombocytosis. Inner ring shows incidence of cancer types in the general population (image and values from Cancer Research UK).

12% respectively). Colorectal cancer in women with thrombocytosis accounts for more than double the proportion of total diagnoses compared to the general population (23% vs 11%). Conversely, the proportion of breast cancer diagnoses in women with thrombocytosis is less than a quarter of that in the general population (6% vs 31%).

4.13 The relationship between change in platelet count over time and cancer

The analyses presented thus far reflect the patient's risk of being diagnosed with cancer within one year of having a single blood test result showing thrombocytosis. In a clinical setting, patients will often have multiple blood tests measuring platelet count over time. Therefore, the objective of this analysis was to investigate how patient's risk of cancer changes depending on how their platelet count changes over time. Only the thrombocytosis cohort were included in this analysis. Sub-groups were defined based on the value of the patients' second platelet count, if taken within six months of their index date. The results are based on one year cancer incidence. The sub-groups were:

1. Increased platelets (second blood test shows an increase in platelet count (or the same reading));
2. Decreased platelets (second blood test shows a decrease in platelet count, but still exceeding normal values);
3. Normalised platelets (second blood test shows a decrease in platelet count to normal levels);
4. No second platelet count available.

The number of patients in each of these categories and the corresponding one-year incidence of cancer in each group is shown in (Table 4.10). The risk of cancer was greatest in males whose next platelet count showed an increase in platelets or the same platelet count: the cancer incidence was 18.1% (95% CI 15.9-20.5). The cancer incidence in those whose next platelet count decreased, but still exceeded 400×10^9/L was 15.5% (95% CI 13.3-17.8). In patients whose platelet count returned to normal within six months, the cancer incidence was 7.1% (95% CI 6.2-8.1). Finally, the cancer incidence was 8.5% (95% CI 7.8-9.3) in those for whom no second platelet count was available.

For female patients, the same four groups were examined. Those whose second platelet count was still abnormally high were at the greatest risk of cancer, with an

Table 4.10: The one year cancer incidence for men and women with thrombocytosis categorised by their change in platelet count over time. Increasing platelets: second platelet count within six months shows an increase in count or the same count. Decreasing platelets: second platelet count has decreased compared to the first, but is still within the abnormal range. Normalising platelets: second platelet count shows that levels have returned to the normal range. N: sample size.

	Men		**Women**	
	N	Incidence % (95% CI)	N	Incidence % (95% CI)
Increasing platelets	1,121	18.1 (15.9-20.5)	2,637	10.1 (9.0-11.3)
Decreasing platelets	1,061	15.5 (13.3-17.8)	2,522	7.3 (6.3-8.4)
Normalising platelets	3,065	7.1 (6.2-8.1)	6,314	4.0 (3.5-4.5)
No second platelet count	5,767	8.5 (7.8-9.3)	13,615	4.5 (4.2-4.9)

increasing or steady platelet count presenting the greatest risk (cancer incidence 10.1%, 95% CI 9.0-11.3). A lower second reading, still within the abnormal range, resulted in a one–year cancer incidence of 7.3% (95% CI 6.3-8.4). Finally, for those whose second platelet count showed a return to normal value, the cancer incidence was 4.0% (95% CI 3.5-4.5). For females with no second platelet count within six months, the value was 4.5% (95% CI 4.2-4.9).

4.14 Thrombocytosis accompanied by symptoms

Most cancers are diagnosed through primary care consultations in which patients present to their GP with symptoms. The primary analysis in this thesis investigated the risk of any type of cancer being diagnosed within one year of a single raised platelet count irrespective of other features, whereas in a clinical setting, GPs will use the patient's symptoms, past medical history, risk factors, and their own clinical judgement to consider which type of cancer, if any, is likely in the patient. The following analyses addressed the objective of investigating the risk of cancer in patients who report symptoms in addition to thrombocytosis. The most common symptoms for the two most common cancers in patients with thrombocytosis were chosen: cough as a sign of lung cancer (Hamilton *et al.*, 2005a), and rectal bleeding and change in bowel habit as symptoms of colorectal cancer (Hamilton *et al.*, 2005b) Two common vague and 'generic' symptoms of many types of cancer, weight loss and loss of appetite, were also studied. Only symptoms recorded within the month prior to or three months after the

index date were included in this sub–analysis, to maximise the chance that both the thrombocytosis and the additional symptom were related to the same clinical episode.

4.14.1 Lung cancer and cough

A cough is the most common symptom of lung cancer in adults; around 65% of patients subsequently diagnosed with lung cancer consult their GP with a cough in the year prior to their diagnosis (Hamilton *et al.*, 2005a) 2,915 patients with thrombocytosis also reported a cough. 125 of these were diagnosed with lung cancer within one year. The one year incidence of lung cancer in patients with thrombocytosis and a cough was 3.9% (95% CI 3.2–4.7). The one year incidence of lung cancer in male patients with thrombocytosis alone was 3.2% (95% CI 2.9-3.6). These values show that thrombocytosis plus a cough is not much more predictive of lung cancer than thrombocytosis alone.

The PPV of cough as a single symptom of lung cancer is 0.4% (95% CI 0.3-0.5) and when the patient has attended twice with a cough, the value is 0.6% (95% CI 0.4-0.8) (Hamilton *et al.*, 2005a). Thrombocytosis in addition to a cough therefore has the potential to identify patients with lung cancer with a much greater predictive value.

4.14.2 Colorectal cancer, rectal bleeding, and change in bowel habit

A change in bowel habit (usually constipation or diarrhoea) is the most common symptom of colorectal cancer; 63% of colorectal cancer patients consult their GP regarding a change in bowel habit in the year prior to their diagnosis (Hamilton *et al.*, 2005b). The second most common symptom, rectal bleeding, is reported by 42% of colorectal cancer patients. 2530 patients with thrombocytosis also reported a change in bowel habit at around the same time as their first raised platelet count; 153 of these were subsequently diagnosed with colorectal cancer.

The one year incidence of colorectal cancer in patients with thrombocytosis and a change in bowel habit is 5.7% (95% CI 4.8–6.7). 354 patients with thrombocytosis also had a record of rectal bleeding; 32 of these were diagnosed with colorectal cancer within a year. The one year incidence of colorectal cancer in patients with thrombocytosis and rectal bleeding was 8.4% (95% CI 5.8-11.9). The fact that rectal bleeding is a less commonly reported symptom and therefore provides a smaller sample size for this calculation is reflected in the wider confidence intervals for the incidence. For comparison, the one year incidence of colorectal cancer in male patients with thrombocytosis only was 2.7 (2.4-3.1); additional alarm symptoms results in increased incidence.

The risk of colorectal cancer increased when patients presented with a change in bowel habit or rectal bleeding in addition to thrombocytosis. The PPV of diarrhoea

alone is 0.9% (95% CI 0.7-1.1) and for constipation it is 0.4% (95% CI 0.3-0.5) (Hamilton *et al.*, 2005b). Therefore, the presence of thrombocytosis could be an important addition in predicting colorectal cancer in patients with a change in bowel habit (which encompasses diarrhoea and constipation); the risk of cancer for patients with both was 5.7% (95% CI 4.8–6.7).

The PPV of rectal bleeding as a single symptom is 2.4% (95% CI 1.9-3.2) (Hamilton *et al.*, 2005b). This is already an alarm symptom of colorectal cancer and may prompt suspicions of cancer in practice without the addition of thrombocytosis.

4.14.3 Loss of appetite and weight loss

The increased incidence with known alarm symptoms shown above may be of limited clinical use as the presence of the alarm symptom may alert GPs to the possibility of cancer without needing blood test results as an additional prompt. Alternatively, the raised platelet count in addition to vague, non-specific symptoms may be of greater clinical use. Loss of appetite has been reported as a marker of several types of cancer NICE (2015). 257 thrombocytosis patients also reported a loss of appetite, and 64 of these were subsequently diagnosed with cancer. The one year incidence of any type of cancer in patients with thrombocytosis and a loss of appetite is 22.6% (95% CI 17.6–28.2). Weight loss was also reported by 698 thrombocytosis patients, with 182 diagnosed with cancer. The one year incidence of cancer in patients with thrombocytosis and weight loss is 24.1% (95% CI 20.9–27.4). The one year incidence of any type of cancer for male patients with thrombocytosis alone was 11.6 (11.0-12.3). Therefore, there is a sizeable increase in cancer incidence in patients who are experiencing weight loss or loss of appetite, and have thrombocytosis; around double the number of incident cases.

4.15 Symptom profiles and NICE guidance

The following section addresses the objective of estimating the potential impact of the recognition of thrombocytosis as a marker of cancer in UK suspected cancer guidance by examining the proportion of patients who have thrombocytosis but no other cancer symptoms or markers. For patients who have symptoms that, according to UK NICE guidance, are sufficiently indicative of cancer to warrant further investigation, the addition of thrombocytosis as a risk marker has little use; other symptoms will prompt investigation. However, for patients with thrombocytosis and no other additional symptoms, the recognition of thrombocytosis as a risk marker has the potential to initiate investigation and therefore diagnosis earlier. The aim of the following analysis was to estimate the proportion of patients with cancer and thrombocytosis who fall into the

latter group; those who would not have been referred for suspected cancer investigation based on symptom profiles that do not include thrombocytosis. The following section addresses the objective of estimating the potential impact of the recognition of thrombocytosis as a marker of cancer in UK suspected cancer guidance by examining the proportion of patients who have thrombocytosis but no other cancer symptoms or markers. For patients who have symptoms that, according to UK NICE guidance, are sufficiently indicative of cancer to warrant further investigation, the addition of thrombocytosis as a risk marker has little use; other symptoms will prompt investigation. However, for patients with thrombocytosis and no other additional symptoms, the recognition of thrombocytosis as a risk marker has the potential to initiate investigation and therefore diagnosis earlier. The aim of the following analysis was to estimate the proportion of patients with cancer and thrombocytosis who fall into the latter group; those who would not have been referred for suspected cancer investigation based on symptom profiles that do not include thrombocytosis.

4.15.1 Lung cancer

The UK national guidance for suspected cancer recommends cancer investigation in patients who do not smoke with two or more of the following symptoms: cough, fatigue, shortness of breath, chest pain, weight loss, or appetite loss (NICE, 2015). If the patient has ever smoked, only one of these symptoms is required to initiate investigation (usually a chest X-ray).

573 patients in the thrombocytosis cohort were diagnosed with lung cancer. Smoking data were available for 546 of these. Of the 490 patients who had ever smoked, 172 (35%) had no recorded lung cancer symptoms in the year prior to their diagnosis, and of 56 who had never smoked, 23 (41%) reported no symptoms in the year prior to diagnosis. For all patients with thrombocytosis and no further recorded symptoms who were subsequently diagnosed with lung cancer, the median number of days between their first blood test showing thrombocytosis and their cancer diagnosis was 50 (IQR 18-126). This suggests that the addition of thrombocytosis to the 2015 NICE guidance update should result in around a third of lung cancers being diagnosed earlier, perhaps before respiratory symptoms have developed.

4.15.2 Colorectal cancer

There are no recommendations in the 2015 NICE guidance for action in response to thrombocytosis relating to colorectal cancer. 627 patients with thrombocytosis were diagnosed with colorectal cancer, and 206 (33%) had no symptoms in the year before diagnosis warranting urgent investigation for cancer. The median number of days

between thrombocytosis and diagnosis date for asymptomatic patients with colorectal cancer was 67 (IQR 27–174). These results suggest that in around a third of colorectal cancer patients, thrombocytosis could have triggered investigation in a patient who previously did not meet recommendations for investigation; the median number of days indicates that this investigation could have occurred around two months earlier.

4.16 Two year cancer incidence

4.16.1 Thrombocytosis cohort

There were 3,050 qualifying cancer diagnoses in the patients with thrombocytosis in the two years after their index date; 597 in addition to the 2,453 diagnosed in the first year. Of these additional 597, 224 men and 373 women were diagnosed. In total over the two-year period, there were 1,322 new diagnoses in men and 1,728 diagnoses in women. The two year cancer incidence in the men with thrombocytosis was almost double that in women: 14.0% (95% CI 13.3-14.7) for men and 7.9% (95% CI 7.6-8.3) for women (see Table 4.11). The additional risk acquired in the second year after index date is shown in the fourth row of Table 4.11. This row shows the incidence of cancer within the second year after index date (rather than within the entire two year period). Comparing the cancer incidence in 1-12 months after index date and 13-24 months after index date, it is clear that the vast majority of new cancers in this cohort are diagnosed within the first year after index date. In the second year after index date (13-24 months), the cancer incidence decreased to 2.7% (95% CI 2.4-3.1) in men and 1.8% (95% CI 1.6-2.0) in women; this excludes patients who were diagnosed in the first year, almost a return to baseline levels.

4.16.2 Normal platelet count cohort

There were 332 qualifying (meeting the inclusion criteria) cancer diagnoses in the patients with a normal platelet count in the two years after their index date; 107 in addition to the 225 diagnosed in the first year. Of these additional 107, 35 were in men and 72 in women. In total over the two year period, there were 141 new diagnoses in men and 191 diagnoses in women. The two year cancer incidence in the normal platelet count cohort was 5.4% (95% CI 4.6-6.4) for males and 3.6% (95% CI 3.1-4.1) for females (see Table 4.11). The additional risk acquired in the second year after index date is shown in the fourth row of Table 4.11. In the second year after index date, the cancer incidence was 1.4% (95% CI 1.0-1.9) in men and 1.4% (95% CI 1.1-1.7) in women; this excludes patients diagnosed in the first year.

4.16.3 Comparison of first year cancer incidence with second year cancer incidence

When the cancer incidence in the first year after index date is compared to that in the second year (comparing new diagnoses 1-12 months after index date to new cases in the 13-24 months after index date), the risk of cancer remains only slightly elevated in men with thrombocytosis compared to those with a normal platelet count. In women with thrombocytosis, the cancer risk in the second year after index date is no different to that in patients with a normal platelet count. This suggests that over a year after a raised platelet count, the risk of cancer largely returns to baseline levels.

4.16.4 Types of cancer diagnosed in the second year

The types of cancer diagnosed in patients with thrombocytosis and with a normal platelet count were compared for the second year after index date (13-24 months).

Table 4.11: The percentage of patients with thrombocytosis and normal platelet counts diagnosed with cancer, within 1-12, 1-24, and 13-24 months of platelet count index date. N: number, CI: confidence interval.

Time from index date (months)	**Thrombocytosis**				**Normal platelet count**			
	Men		**Women**		**Men**		**Women**	
	N	n cancers diagnosed, incidence % (95% CI)	N	n cancers diagnosed, incidence % (95% CI)	N	n cancers diagnosed, incidence % (95% CI)	N	n cancers diagnosed, incidence % (95% CI)
1-12	9,435	1,098 11.6 (11.0-12.3)	21,826	1,355 6.2 (5.9-6.5)	2,599	106 4.1 (3.4-4.9)	5,370	119 2.2 (1.8-2.6)
1-24	9,435	1,322 14.0 (13.3-14.7)	21,826	1,728 7.9 (7.6-8.3)	2,599	141 5.4 (4.6-6.4)	5,370	191 3.6 (3.1-4.1)
13-24	8,337**	224 2.7 (2.4-3.1)	20,471**	373 1.8 (1.6-2.0)	2,493**	35 1.4 (1.0-1.9)	5,251**	72 1.4 (1.1-1.7)

** Excluding patients with cancer diagnoses recorded within months 1-12.

4.17 Chapter summary

This large scale cohort study is the first from primary care to compare the overall risk of cancer between patients with thrombocytosis and those with normal platelet counts. It addresses the objectives set out in the introduction to investigate the relationship between thrombocytosis and cancer. This section of the chapter reviews the strengths and limitations of this study, considers the results in line with existing literature, and outlines the clinical implications of these findings.

4.17.1 Objectives addressed in this chapter

- **To examine the incidence of cancer in two cohorts of patients; those with thrombocytosis and those with a normal platelet count, to determine the risk of cancer in each cohort.**
- **To compare the cancer incidence between these two cohorts to determine the absolute increase in risk associated with thrombocytosis.**
- **To examine how the relationship between thrombocytosis and cancer differs across subgroups defined by age, sex, and smoking status.**

This study found that the one year incidence of cancer in males with thrombocytosis is 11.6% (95% CI 11.0-12.3), and in males with a normal platelet count it is 4.1% (95% CI 3.4-4.9). In females with thrombocytosis the one year cancer incidence is 6.2% (95% CI 5.9-6.5) and in female patients with a normal platelet count, the equivalent incidence is 2.2% (95% CI 1.8-2.6). This represents an absolute increase in risk of 7.5% for raised platelets over normal platelets for men, and an absolute increase of 4.0% for women, although the overall risk is the primary reported value as these are more clinically useful.

The risk of cancer increased with age in both cohorts of patients; unsurprisingly. However, the incidence of cancer increased at a greater rate in patients with thrombocytosis than in patients with a normal platelet count, meaning that more cancers are diagnosed in patients with thrombocytosis than in patients with a normal platelet count in older age groups. Scatter plots of the cancer incidence by age group in each cohort show the risk of cancer increasing steadily in men, but appearing to level off in women. This could reflect fewer cancers being diagnosed in older women; it is also possible that this reflects less recording of cancers in older women. Cancer incidence data described in Chapter 2 appear to show a decrease in cancers diagnosed in older patients. However, anecdotally this could be due to fewer cancers in older patients being recorded.

A stratified analysis estimated the risk of cancer in patients with thrombocytosis who had ever smoked, or had never smoked. This found that although the overall cancer incidence was higher in patients who had ever smoked than in those who had never smoked, there was some overlap between the confidence intervals; although this does not mean that there is no statistically significant difference between the two groups, the difference in incidence is not clinically significant.

- **To determine whether some types of cancer are more likely to be diagnosed than others in patients with thrombocytosis compared to patients with a normal platelet count.**

The incidence was found to be higher for some types of cancer than others in patients with thrombocytosis; it was particularly high for lung and colorectal cancer, and low for prostate and breast cancer compared to patients with a normal platelet count, and the general population. It could be argued that lung and colorectal cancer appeared to have higher incidence in patients with thrombocytosis simply because they are most commonly in the general population. However, comparing the incidence of these two types of cancer between patients with thrombocytosis and with a normal platelet count suggests that this is not the case. If thrombocytosis was not more strongly associated with lung and colorectal cancer, than similar incidence would be observed for all cancer sites among patients with thrombocytosis and a normal platelet count. However, this is not the case. Lung cancer incidence was 3.2% in men with thrombocytosis (2.9-3.6) and 0.4% (0.2-0.7) in men with a normal platelet count; a absolute increase of 2.8%. Similarly, colorectal cancer incidence was 2.7% (2.4-3.1) in men with thrombocytosis and 0.4% (0.2-0.7) in men with a normal platelet count; an increase of 2.3%. For women, a similar pattern was observed. Lung and colorectal cancer incidence in those with thrombocytosis was 1.1% (0.9-1.2) and 1.6% (1.4-1.7) respectively. For women with a normal platelet count, incidence of lung and colorectal cancer was 0.3% (0.1-0.4) and 0.4 (0.3-0.6).

For prostate and breast cancer, the other two most commonly diagnosed cancers in the general population, there appeared to be no difference in incidence between patients with thrombocytosis and with a normal platelet count. Breast cancer incidence was 0.5% (0.3-0.7) in women with thrombocytosis, and 0.4% (0.3-0.5) in women with a normal platelet count. For prostate cancer, the incidence in men with thrombocytosis was 1.4% (1.2-1.7) and the incidence in men with a normal platelet count was 1.1% (0.7-1.6). Other hormone-dependent cancers had a similar incidence observed in each of the cohorts; cervical cancer incidence was 0.1% (0.0-0.1) in women with thrombocytosis and 0.1% (0.0-0.2) in women with a normal platelet count. Ovarian cancer incidence was 1.3% (1.1-1.4) in women with thrombocytosis and 0.8% (0.7-0.9) in women with

a normal platelet count. Uterine cancer incidence was 0.2% (0.2-0.3) in women with thrombocytosis and 0.2% (0.1-0.2) in women with a normal platelet count. There were only two cases of testicular cancer; both in the thrombocytosis cohort.

- **To investigate how patient's risk of cancer changes depending on how their platelet count changes over time.**

An increase in platelet count within a six month period was associated with increased cancer incidence in patients with thrombocytosis; this fits with earlier analyses which show increasing cancer risk with increasing baseline platelet count. However, this result is open to bias; any other symptoms are likely to worsen or appear within a six month period, so it is more likely that second blood tests within six months were ordered by clinicians who suspected cancer. A more thorough investigation in to this hypothesis could include an analysis of other clinical features in patients over the six month time period, or examine patterns of symptom presentation in patients who are later diagnosed with cancer.

- **To investigate the risk of cancer in patients who report symptoms in addition to thrombocytosis.**

Patients presenting with a cough (the most common symptom of lung cancer) in addition to thrombocytosis had a similar incidence of cancer than patients presenting with thrombocytosis alone; but a greater risk than a cough as a single symptom. For patients with a persistent cough, the presence of thrombocytosis could be valuable prompt in encouraging clinicians to suspect cancer.

For colorectal cancer, patients presenting with thrombocytosis and a change in bowel habit had a much greater risk of cancer than patients with thrombocytosis or a change in bowel habit as single symptoms. Again, the greater predictive value of these two symptoms combined could be clinically useful. Although there was an increase in the risk of colorectal cancer for patients with thrombocytosis and rectal bleeding compared to thrombocytosis alone, rectal bleeding is already considered an alarm symptom of colorectal cancer and thus the clinical utility of this symptom combination is more limited than the previous two examples.

Thrombocytosis plus weight loss (24.1%, 95% 20.9-27.4) and a loss of appetite (22.6%, 95% CI 17.6-28.2) was more strongly predictive of cancer than thrombocytosis alone (11.6% (11.0-12.3) for men and 6.2% (5.9-6.5) for women); the incidence of any type of cancer was roughly doubled with the addition of either of these symptoms. The range of possible cancers is wide so clinicians would have to use other signs and symptoms, the patient's medical history, and their own clinical judgment to determine which investigative services would be most appropriate for these patients.

- **To investigate the stage at which cancers are diagnosed in patients with thrombocytosis and with a normal platelet count.**

It was not possible to meet this objective due to incomplete data in the cancer registry data file.

- **To estimate the potential impact of the recognition of thrombocytosis as a marker of cancer in UK suspected cancer guidance by examining the proportion of patients who have thrombocytosis but no other cancer symptoms or markers.**

Around a third of lung and colorectal cancer patients with thrombocytosis had no symptoms that would have qualified them for investigation of suspected cancer as per NICE guidance in the year before their diagnosis. This is an exciting finding which has the potential to expedite a good proportion of cancer diagnoses annually.

4.17.2 Strengths and limitations

A key strength of this study is its size and setting. It uses primary care data, where the initial suspicion of possible cancer is generally made, and uses data from two recognised strong data sources, the CPRD and the cancer registry. The cancer registry is considered the gold standard of cancer recording in England. Previous studies have found cancer recording in the CPRD to be highly valid; this evidence was summarised in Chapter 2.

The present analysis only includes data from England, not Wales, Scotland, or Northern Ireland. Furthermore, within England, some regions were under-represented (particularly the North East of England), with a greater proportion of practices from the South East and North West of England. However, as patients in each of the two cohorts were practice matched, this will not have introduced an important bias to the differences in cancer incidence found between the two. This could, however, bias the results if there were strong differences in cancer diagnosis between different regions of England, or different regions of the UK. Figures released in 2016 show generally poor performance across the UK in cancer diagnosis, with no strong geographical patterns in diagnosis. Even if a pattern existed, it is very unlikely to have an impact on the relationship between platelets and cancer, although in an area with very high early diagnosis of cancer, thrombocytosis could be of lesser value in expediting diagnosis.

These results rely on the accuracy of data in the CPRD. Due to the electronic transmission of blood test results from the laboratory to patient records, errors in the recorded values are unlikely. However, as each patient's record begins from the time

they join a CPRD practice, it is possible that they had thrombocytosis earlier than the used index date, whilst registered at a non-CPRD practice. Any under-recording of platelet count would be similar for both cohorts.

It was not possible to meet the objective of investigating the stage at which cancers are diagnosed in patients with thrombocytosis and with a normal platelet count; this is because the registry staging data were incomplete. It is only possible to state that at least one quarter of patients with thrombocytosis were diagnosed at an early stage and therefore most likely to benefit from earlier diagnosis in terms of survival. This proportion may be higher.

The reasons for full blood counts being ordered for patients in each of the cohorts are not known. Blood tests could have been ordered in the thrombocytosis cohort because their clinician already suspected cancer in the patient. Although it is not possible to determine the reason for testing, the signs and symptoms reported by the patients in the lead up to their blood test can be compared between the two cohorts. This can give an indication of any differences in symptomatic presentation at baseline. Some of the symptoms reported by each cohort were recognised cancer symptoms, but they were reported by relatively small, and similar, proportions of each sub-cohort, and they are low risk symptoms which would not be expected to prompt urgent cancer investigation (NICE, 2015). This suggests that the blood tests carried out in the thrombocytosis cohort were no more likely to be prompted by suspicions of cancer than those in the normal platelet count cohort.

In primary care, blood tests are ordered for a wide range of reasons; sometimes in response to symptoms, sometimes for routine reasons, and rarely to look for thrombocytosis specifically. Approximately a quarter of the UK adult primary care population have a full blood count taken in any one year (Hamilton *et al.*, 2008) Patients selected for a full blood count would be expected to have more ill health than those who are not; therefore comparing cancer incidence between patients with thrombocytosis and those with a normal platelet count, rather than those with no testing at all, is a key strength. This should reduce (or eliminate) any bias from having been selected for blood testing. Therefore, these results do not report the value of measuring the platelet count in predicting cancer; rather, they report the value of an elevated platelet count result. This is a subtle but real distinction.

These limitations, and their implications for the interpretation of the results, are discussed further in Chapter 6; the overall discussion and conclusion chapter.

The next chapter presents the validation study of CPRD data, using linked cancer registry data. A sensitivity analysis is presented that repeats some work from the present chapter, using only cancer registry recorded diagnoses. The is used to estimate what, if any, impact there is of including CPRD-recorded cancer diagnoses that have

not been verified by cancer registry records.

Chapter 5

CPRD validation study

5.1 Chapter summary

In this chapter, a sensitivity analysis is presented which examines the effect of the inclusion of cancers recorded in the CPRD, but not the cancer registry, on the results. In addition, the validity of the CPRD data sample used in Chapter 4 is examined by comparing this source with English cancer registry data. This validation exercise uses the same cohort as in Chapter 4 and involves comparing the recorded diagnoses in each source, the dates these diagnoses were made, the type of cancer diagnosed, and differences in patient characteristics between those cancers that are recorded in both sources or either source. The sensitivity analysis involves repeating the primary analysis from Chapter 4, the one year cancer incidence in patients with thrombocytosis and a normal platelet count, including only cancer registry-recorded diagnoses. This sensitivity analysis is placed here, rather than in Chapter 4, because it ties in with the 'validation' aims of this chapter and it relies upon data comparisons made in the first part of this chapter. The primary aim of this chapter was to examine the validity of the data sample used in Chapter 4, and to carry out a sensitivity analysis to determine whether there are any effects of including non-registry validated data in the main analysis in that chapter. Secondly, this chapter presents an opportunity to build on the work of Boggon *et al.* (2013) Dregan *et al.* (2012) who also carried out validation studies on CPRD data samples. In those two studies, similar to the work presented in this chapter, selected samples of patients were used; in the case of Boggon *et al.* (2013), the cohort were selected for inclusion if they had diabetes, and Dregan *et al.* (2012) selected patients with certain risk symptoms including haematuria, haemoptysis, dysphagia, or rectal bleeding, or lung, colorectal, gastro-oesophago, or urinary cancer. The selected samples of patients used in those and in the present study are a limitation which should be considered before generalising results about the validity of CPRD data to the data

source as a whole.

5.2 Chapter background

Electronic medical records are increasingly being used in medical research as they offer an opportunity to access large amounts of routinely collected data from a broad range of patients, often collected in the setting in which any research outputs aim to be of clinical use. Electronic medical records are not without limitation; two of the greatest limitations are that the data are primarily collected for reasons other than for research, and that there can be heterogeneity in recording styles between different sources. Ideally data collection for research purposes should be homogenous and complete, but electronic medical records are often compiled by medical professionals for whom achieving research quality data is not a priority. Data recording errors can also occur.

Validation studies are one way to assess the quality of the recorded data. This involves comparing recording in the data source of interest to recording in a 'gold standard' data source. The validity of diagnostic coding in the CPRD has been systematically reviewed by Khan *et al.* (2010) (described in Chapter 2), and more recently the validity has been explored specifically in relation to cancer diagnoses by Dregan *et al.* (2012) and Boggon *et al.* (2013). These studies found a high level of concordance between the CPRD and the 'gold standard' cancer registry in terms of cancer diagnosis records. Dregan *et al.* (2012) found that $5,216$ of $5,429$ (96%) CPRD cancer diagnoses were confirmed by the cancer registry (Dregan *et al.*, 2012). Boggon *et al.* (2013) found that $4,830$ of $5,797$ (83%) CPRD cancer diagnoses were confirmed by the cancer registry (Boggon *et al.*, 2013) - and of the 967 diagnoses that were not verified by cancer registry records, 528 (54.6%) were confirmed by hospital records or practice notes.

5.3 Implications of validation of CPRD diagnoses

In Chapter 4, counting patients as cases if they had a cancer record in either source could have resulted in an over estimate of the true association if some of these were false positives. In the present chapter, examining the concordance between the two data sources will enable the assessment of the extent of this overestimation, if it exists. In an attempt to determine the impact of including patients diagnosed from either source on the study results, a sensitivity analysis is included in this chapter in which the primary analysis from Chapter 4 is repeated, including only patients with a cancer registry-recorded cancer. Patients who have a cancer record in the CPRD but not the cancer registry are excluded. These results are compared to results from the first

'diagnosis in either source' results to determine what effect, if any, the inclusion of CPRD diagnoses that are not validated by the cancer registry has on the results.

The previous chapter in this thesis described a cohort study in which the relationship between thrombocytosis and cancer was investigated using CPRD data. Linked cancer registry data were available for some patients in the cohort who were diagnosed with cancer; the data linkage is described in Chapter 2. Previous validation studies indicate that not all incident cases of cancer are captured by the CPRD, so patients were included in the Chapter 4 analysis as having cancer if they had a recorded diagnosis in either the CPRD, or the cancer registry. Another possible approach was to only include patients whose CPRD record was confirmed by a cancer diagnosis recorded in the cancer registry; this is considered the 'gold standard' of cancer recording in England. There is a balance between including only cases with a record in the cancer registry and risking the exclusion of some CPRD-recorded true positives that were not recorded in the cancer registry, and including cases recorded in either source and risking the inclusion of some false positives which could occur in either source. In this chapter, the validity of the cancers recorded in the CPRD data is assessed using data from the English cancer registry.

5.4 Research question and objectives

In this chapter, the primary aim was: **To assess the validity of cancer recording in the Clinical Practice Research Datalink sample used in this thesis, using cancer recording in the English cancer registry as the 'gold standard'.**
The objectives developed to answer this research question were:

- To compare cancer recording in the CPRD and in the cancer registry to determine the level of concordance between the two sources.
- To compare the age and sex of patients recorded in both, or either, source.
- For cancers recorded in both sources, to compare the date of recording between the two.
- To estimate predictors of concordance between the two data sources.
- To examine the extent to which the inclusion of unverified CPRD-recorded cancer diagnoses causes overestimates in incidence figures from CPRD data, by repeating the primary analysis from Chapter 4 including only cancer registry recorded diagnoses.

5.5 Methods

5.5.1 Data sources and included patients

This chapter uses data from the cohort of 50,000 patients described in Chapter 4. The analysis in Chapter 4 only included patients diagnosed with cancer after a certain date (the date of the patient's first recorded thrombocytosis). Any patients with a cancer recorded prior to this index date were excluded. In the present chapter, the validation study includes any cancers diagnosed in the cohort in the entire study period for which cancer registry data are available: 2000-2010. All practices contributing data to the study were in England.

Data from the CPRD data file and the cancer registry data file were combined into one file for this study. In addition to the patient characteristics, the file detailed the date and site of the cancer diagnosed according to each of the two sources. The recorded date of diagnosis was available according to each of the two data sources. The date of diagnosis is recorded in the cancer registry in the day-month-year format; this was converted to the number of days since 1st January 1900 to enable comparison with the CPRD data, as described in Chapter 4. A patient was considered to have been diagnosed with cancer according to the CPRD if they had any one of the 2,134 cancer-related medcodes (see Appendix C) which are grouped into 20 common cancer sites. Where more than one record of a cancer exists, the first record in time was taken as the primary site and date. A patient was considered to have a cancer registry recorded diagnosis if a record existed for their patient ID number in the raw cancer registry data file.

5.5.2 Patient subgroups

The patient subgroups examined in this analysis are:

1. Cancer diagnoses recorded in both sources; patients have a record in the CPRD and in the cancer registry.

2. CPRD only; patients have a record of a cancer diagnosis in the CPRD file, but no record in the cancer registry.

3. Cancer registry only; patients have a record of a cancer diagnosis in the cancer registry file, but no record in the CPRD.

5.6 Statistical methods and analysis

5.6.1 Sample exclusions

Using the previously described 'days since 1900' approach, the patient's age on the date of diagnosis was determined by dividing the difference in days between their birth date and diagnosis date, both as days since 1st January 1900, by 365.25 (the extra 0.25 reflects leap years). Any patients aged under 40 years at the time of diagnosis were excluded from the analysis.

Cancer registry data were available for patients diagnosed from 2000-2010. However, for CPRD data, a more extensive medical history is available encompassing all patient records from when they registered at their practice, including cancer diagnoses from any point in time. To ensure an even comparison, only cancers diagnosed between 2000 and 2010 in each data source were included. Therefore, any patients with a diagnosis recorded prior to 2000 or after 2010 were excluded.

5.6.2 Outcome measures

To determine the level of concordance between the two sources, new variables were created which identified each patient as having a cancer record in the CPRD or the cancer registry. A list of CPRD-recorded diagnoses were merged into a file with a list of cancer registry-identified diagnoses. These variables were used in combination to determine which patients belonged in each of the subgroups described in section 5.5.2. Where records matched between the two files, the diagnosis was considered recorded in both sources. If a CPRD-originating ID number had no cancer registry match, or a cancer registry-originating ID number had no CPRD match, the patient's diagnosis was considered recorded in only one or the other source.

The sex of patients in each source was compared by tabulating the number in each group for patients diagnosed in both (or either) sources. Age at diagnoses was summarised using the median and interquartile range for each of the three source groups, and by frequency in ten year age groups (40-49, 50-59, 60-69, 70-79, and ≥ 80).

5.6.3 Distribution of diagnoses over time

The number and percentage of diagnoses made each year from 2000 to 2010 in each source was determined and compared. Trends in incidence over the study period were examined.

5.6.4 Diagnoses recorded in both sources

The characteristics of patients with a cancer diagnosis that was recorded in both the CPRD and the cancer registry were described, including the proportion of males and females in this group, and the age distribution, both as median and IQR and in ten year age groups. The date of diagnosis was compared between the two sources by subtracting the date of diagnosis in the CPRD (recorded as days since 1900) from the date of diagnosis in the cancer registry (also recorded as days since 1900). The median difference and IQR in days between recording in the two sources was described. A positive value indicated that the CPRD record was made first and a negative value indicated that the cancer registry record was made first. Finally, the type or primary site of cancer diagnosed was explored. The proportion of diagnoses recorded in both sources with the same, and different, primary sites was described. For those diagnoses with different primary sites, the most commonly recorded sites were examined.

5.6.5 Validation measures

Reporting guidelines for validation studies of routinely collected health data were recently published by Benchimol *et al.* (2010). These guidelines recommend reporting at least four estimates of diagnostic accuracy, including sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), when comparing a large scale health dataset (such as the CPRD) to another reference standard or 'gold standard' data source (in this case, the cancer registry). These measures were taken for cancer diagnosis recording in the CPRD, compared to the cancer registry.

5.6.6 Types of cancer recorded

Where patients with a cancer record in the CPRD also had a corresponding diagnosis recorded in the cancer registry, the primary site of diagnosis in the CPRD was compared to that in the cancer registry. The 20 cancer sites are described in Section 4.3.14.

5.6.7 Diagnoses recorded in the CPRD only, or the cancer registry only

The measures described above were repeated for those diagnosed in the CPRD only or the cancer registry only. This included the proportion of males and females, the median age, and the proportion in each ten year age group.

5.6.8 Determining predictors of concordance

Logistic regression was used to determine predictors of concordance between the two data sources. A binary variable, *concordance*, was created where 1 = sources agree on diagnosis and 0 = sources do not agree on diagnosis. Age in ten year age groups, gender, and year of diagnosis were included as predictor variables.

Each of these three variables were investigated individually in univariable logistic regression models. All three were included in a multivariable model. P values were reported to reflect those variables which retained significance as predictors of concordance in the multivariable model.

5.6.9 Sensitivity analysis

The methods used in Chapter 4 to determine cancer incidence were repeated here. However, in this analysis, patients were only counted as cases if they had a cancer registry-recorded diagnosis. Patients with a CPRD cancer diagnosis not validated by the cancer registry were included in the analysis as not having cancer. Cancer registry cases with no corresponding CPRD record were included as cases.

5.7 Results

The cohort included $50,000$ patients. $4,554$ were excluded as they had a cancer diagnosis recorded prior to 1st January 2000, and a further 282 were excluded as they had a recorded cancer diagnosis after 31st December 2010. 20 patients were excluded as they were aged less than 40 years at the time of diagnosis. Therefore after exclusions there were $45,144$ patients in the cohort (Figure 5.1). $8,889$ of these (19.7%) had a recorded cancer diagnosis in either the CPRD or the cancer registry. $7,785$ diagnoses were recorded in the cancer registry, and $7,028$ were recorded in the CPRD. The number and percentage of diagnoses in each source is presented in Figure 5.2.

5.7.1 Distribution of diagnoses over time

The decade in which cancers were recorded in each of the two data sources was compared (see Table 5.1). The number of recorded diagnoses increases throughout the first four years of the study period, from 549 to 786 per year in the cancer registry and 367 to 727 per year in the CPRD. It is likely that these increases reflected increased recording of diagnoses during this time rather than a true increase in incidence. The number of cancers diagnosed yearly from 2004 to 2008 remains fairly steady. A decrease in

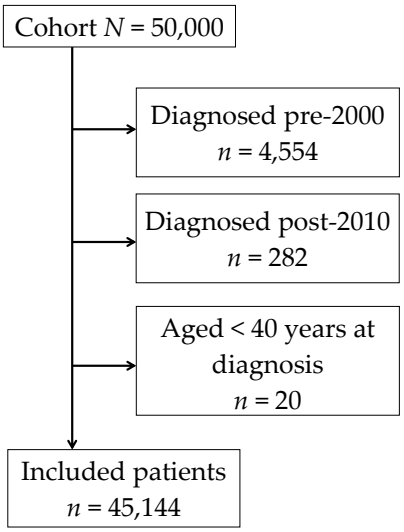


Figure 5.1: Patients excluded from the cohort due to having a recorded diagnosis prior to 2000, after 2010, or being less than 40 years of age at the time of diagnosis (when diagnosed from 2000-2010).

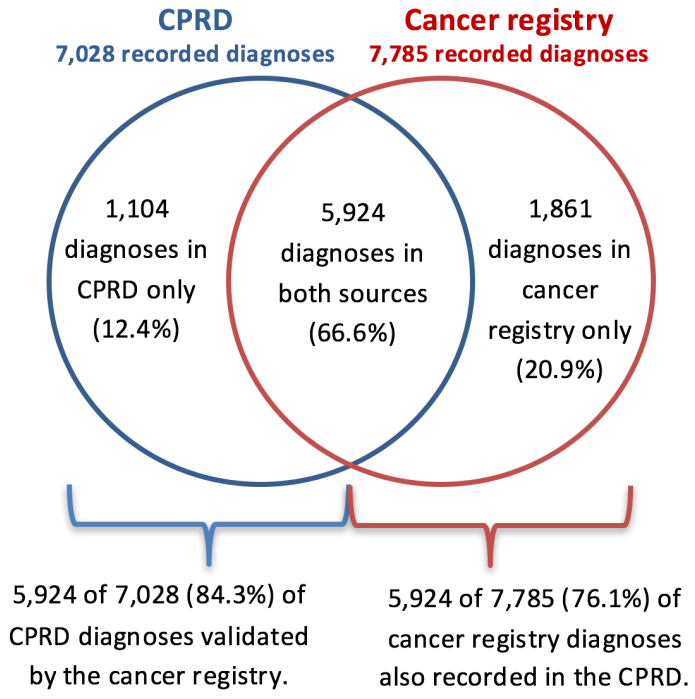


Figure 5.2: Venn diagram to show the number of cancers recorded in the CPRD, and in the cancer registry. The left circle represents CPRD records and the right represents cancer registry records. Overlap between the circles indicate matching records.

Table 5.1: The number of recorded cancer diagnoses in the cancer registry and the CPRD by year.

Year of diagnosis	n in cancer registry	%	n in CPRD	%
2000	549	7.1	367	5.2
2001	687	8.8	599	8.5
2002	736	9.5	654	9.3
2003	786	10.1	727	10.4
2004	789	10.1	697	9.9
2005	838	10.8	776	11.1
2006	846	10.9	755	10.8
2007	804	10.3	716	10.2
2008	738	10.1	707	10.1
2009	638	8.2	573	8.2
2010	329	4.2	448	6.4
Total	**7,785**	**100**	**7,019**	**100**

incidence in 2009 and 2010 is likely to reflect delays in the process of recording new diagnoses, not true population decreases in cancer incidence in this time.

5.7.2 Characteristics of patients diagnosed with cancer

The age and sex profile of patients diagnosed with cancer are shown in Table 5.2. This shows the median (IQR) age at diagnosis, age at diagnosis in age groups, and number and percentage of men and women in each of several groups: cancers recorded in both the CPRD and the cancer registry, cancers recorded only in the CPRD (with no corresponding record in the cancer registry), and cancers recorded only in the cancer registry (no corresponding record in the CPRD). The following section describes and explores cancers that were recorded in both sources or just one or the other. Records made in one source only are investigated later in this chapter.

Table 5.2: Characteristics of patients with a recorded cancer diagnosis in the CPRD and the cancer registry.

	Either source $N = 8,889$	**Both sources** $N = 5,924$	**All CPRD** $N = 7,028$	**All cancer registry** $N = 7,785$	**CPRD only** $N = 1,104$	**Cancer registry only** $N = 1,861$
Age at diagnosis date, years						
Median (IQR)	72.4 (63.6 - 79.8)	71.5 (63.1 - 78.7)	71.8 (63.2 - 79.0)	72.3 (63.3-79.6)	72.8 (63.6-80.6)	75.0 (66.0-82.6)
Age group, years						
40-49, *n* (%)	244 (2.7)	197 (3.3)	183 (2.6)	213 (2.7)	51 (4.6)	70 (3.8)
50-59, *n* (%)	1,009 (11.4)	823 (13.9)	809 (11.5)	897 (11.5)	142 (12.9)	216 (11.6)
60-69, *n* (%)	2,092 (23.5)	1,648 (27.8)	1,745 (24.8)	1,862 (23.9)	253 (22.9)	360 (19.3)
70-79, *n* (%)	2,946 (33.1)	2,013 (34.0)	2,357 (33.5)	2,592 (33.3)	354 (32.1)	593 (31.9)
80 and older, *n* (%)	2,598 (29.2)	1,243 (21.0)	1,934 (27.5)	2,221 (28.5)	304 (27.5)	622 (33.4)
Sex distribution						
Men, *n* (%)	3,800 (42.8)	2,673 (45.1)	3,071 (43.7)	3,402 (43.7)	398 (35.9)	729 (39.2)
Women, *n* (%)	5,089 (57.2) 3,251 (54.9)	3,957 (56.3)	4,383 (56.3)	706 (64.1)	1,132 (60.8)	

5.7.3 Diagnoses appearing in both sources

5.7.3.1 Characteristics of diagnoses with concordance

5,924 patients had a cancer recorded in both the CPRD and the cancer registry. Of these, 2,673 (45.1%) were male and 3,251 (54.9%) were female. The median age in this group was 71.5 years (IQR 63.1-78.7). Age group distribution was as follows: 197 (3.3%) aged 40-49 years; 823 (13.9%) aged 50-59 years; 1,648 (27.8%) aged 60-69 years; 2,013 (34.0%) aged 70-79 years; 1,243 (21.0%) aged 80 years and over.

5.7.3.2 Date of diagnosis

Where patients had a recorded cancer diagnosis in both sources, the date of diagnosis was compared between the two datasets. Diagnoses in the CPRD were recorded a median of seven days later in time than in the cancer registry (IQR -25 to 7; 25 days later in the CPRD to 7 days later in the cancer registry).

The age and sex of patients were compared depending on whether their diagnosis was recorded first in the CPRD or the cancer registry. There was little difference between the two in age at diagnosis; for patients with their first record in the CPRD, the median age was 73.1 (IQR 64.3-80.8) and for patients with their first record in the cancer registry, the median age was 71.3 (IQR 62.7-78.5). A slightly greater proportion of diagnoses in women were recorded in the CPRD prior to the cancer registry; 3,122 out of 5,092 women (61.3%) were diagnosed first in the CPRD. 2,146 out of 3,789 men (56.5%) had their diagnosis recorded earlier in the CPRD.

5.7.3.3 Types of cancer diagnosed

The primary type of cancer was compared for patients who had a diagnosis recorded in both sources. Of the 5,924 cancers recorded in both datasets, 5,339 (90.1%) had the same primary site identified. 585 (9.9%) of the records had different primary sites recorded in each of the two datasets. The majority of CPRD recorded cancers with a different primary site recorded in the cancer registry were coded as either skin (non-melanoma) cancer (191/585, 32.7%) or 'miscellaneous' (230/585, 39.3%). Non-melanoma skin cancers were excluded from the analysis in Chapter 4.

5.7.4 Diagnoses recorded in the CPRD only

1,104 recorded diagnoses in the CPRD had no corresponding record in the cancer registry. Of these 1,104 patients, 398 (36.0%) were male and 706 (64.0%) were female. The median age at diagnosis was 72.8 years (IQR 63.6-80.6). The distribution of age by 10 year age group was as follows: 51 (4.6%) aged 40-49 years; 142 (12.9%) aged

50-59 years; 253 (22.9%) aged 60-69 years; 354 (32.1%) aged 70-79 years; 304 (27.5%) aged 80 years and older.

5.7.5 Diagnoses recorded in the cancer registry only

1,861 recorded diagnoses in the cancer registry had no corresponding records in the CPRD. Of these 1,861 patients, 729 (39.2%) were male and 1,132 (60.8%) were female. The median age at diagnosis was 75.0 years (IQR 66.0-82.6). The distribution of age by 10 year age group was as follows: 70 (3.8%) aged 40-49 years; 216 (11.6%) aged 50-59 years; 360 (19.3%) aged 60-69 years; 593 (31.9%) aged 70-79 years; 622 (33.4%) aged 80 years and older.

5.7.6 Comparison of patients with diagnoses recorded in only one source

There was no difference in the proportion of men and women with records solely in one source or the other. Patients with a diagnosis recorded in the cancer registry only were older than those whose diagnosis were recorded in the CPRD only; the median age was 72.8 years for CPRD only records and 75.0 years for cancer registry only records. The cancer registry only records also had a larger proportion of patients aged ≥ 80 years of age; 33.4% vs 27.5%.

5.7.7 Validation measures

There was concordance between the two data sources on the majority of recorded diagnoses: 5,924 (66.6%) of patients had a cancer record in both the CPRD and the cancer registry. 1,861 (20.9%) were recorded only in the cancer registry, with no record in the CPRD. 1,104 (12.4%) were recorded in the CPRD, and did not have a record in the cancer registry.

A 2 by 2 table outlining the number of cancer diagnoses in each of the two sources is shown in Table 5.3. This cross tabulates the number of patients with a recorded cancer diagnosis, and the number with no recorded diagnosis, in the CPRD and the cancer registry. Sensitivity reflects the proportion of patients with the disease in question (according to the reference standard), who are correctly identified as such by the dataset being validated (the CPRD, in this case). Using the values in Table 5.3, this was calculated as 5,924 identified as having cancer by the CPRD, out of 7,785 with a cancer registry record. The sensitivity of a CPRD recorded diagnosis was 76.1% (95% CI 75.1 - 77.0). This means that 76.1% of patients with a cancer diagnosis recorded in the cancer registry are identified as having cancer in the CPRD.

Table 5.3: Patients with a recorded cancer diagnosis and no recorded cancer diagnosis in the CPRD and the cancer registry

		Cancer registry		
		Cancer record	No cancer record	**Total**
CPRD	Cancer record	5,924	1,104	7,028
	No cancer record	1,861	36,255	38,116
	Total	7,785	37,359	45,144

Specificity measures the proportion of patients who do not have the disease in question (according to the reference standard), who are correctly identified as being free of disease by the dataset (the CPRD). Using the values in Table 5.3, this was calculated as 36,255 identified as not having cancer in the CPRD, out of 37,359 with no cancer registry record. The specificity of a CPRD recorded diagnosis was 97.0% (95% CI 96.9-97.2); 97.0% of patients who do not have a cancer registry recorded diagnosis also do not have a record in the CPRD. Critically, in this measure, the absence of a cancer record in the CPRD does not mean that the patient is specifically coded as being free of cancer. Therefore, the specificity in this case measures how likely a patient with no cancer diagnosis record in the CPRD is to be free of disease.

The PPV estimates the proportion of cancer diagnosis records in the CPRD that are confirmed by a corresponding cancer registry record. The PPV of a CPRD recorded diagnosis was 84.3% (95% CI 83.4 - 85.1). Similarly, the NPV reflects the proportion of patients who do not have a cancer recorded in the CPRD, who also do not have a cancer recorded in the cancer registry. The NPV was 95.1% (95% CI 94.9 - 95.3).

5.7.8 Predictors of concordance between the cancer registry and the CPRD

Logistic regression was used to investigate potential predictors of concordance between the two data sources. Predictor variables included sex (binary), age group (40-49 years; 50-59 years; 60-69 years; 70-79 years; 80 years and over), and year of first record of diagnosis (2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010).

In univariable (unadjusted) models, being aged 50-59 years ($p = 0.042$) or 60-69 years ($p = 0.001$) significantly predicted concordance, whereas being aged 70-79 years ($p = 0.068$) or aged 80 years and over ($p = 0.471$) did not (the 40-49 years age group was used as the reference group). Sex also predicted concordance ($p < 0.001$) with cancers in

male patients more likely to match between the CPRD and the cancer registry. Finally, diagnoses being recorded from 2005 onwards significantly predicted concordance in a univariable (unadjusted) model. With 2000 as the reference group: 2001 $p = 0.414$; 2002 $p = 0.598$; 2003 $p = 0.890$; 2004 $p = 0.052$; 2005 $p = 0.036$; 2006 $p = 0.011$; 2007 $p = 0.010$; 2008 $p = 0.034$; 2009 $p = 0.005$; 2010 $p = 0.025$.

The multivariable model included age (categorical, 10 year age bands), sex, and year of diagnosis as predictor variables. Only the 60-69 age group retained significance as a predictor of concordance ($p = 0.006$) within the age groups. Sex remained a significant predictor of concordance ($p < 0.001$). Years of diagnosis 2006 ($p = 0.029$), 2007 ($p = 0.039$), 2009 ($p = 0.014$), and 2010 ($p = 0.007$) also retained significance as predictors of concordance.

5.8 Sensitivity analysis using cancer registry data

This section addresses the final objective of this chapter: to examine the extent to which the inclusion of unverified CPRD-recorded cancer diagnoses causes overestimates in incidence figures from CPRD data, by repeating the primary analysis from Chapter 4 including only cancer registry recorded diagnoses.

5.8.1 Primary analysis: recap of one year cancer incidence

After exclusions (defined in the previous chapter), there were 31,261 patients with thrombocytosis (9,435 men and 21,826 women) and 7,969 patients with a normal platelet count (2,599 men and 5,370 women). Table 5.4 shows the number of men and women with thrombocytosis, and the number of each who had a record in either the CPRD or the cancer registry showing a diagnosis of cancer within one year of their index date (the date of their first blood test showing thrombocytosis, or equivalent).

5.8.1.1 Thrombocytosis cohort

There were 2,453 cancers recorded in either the CPRD or the cancer registry in the thrombocytosis cohort; this represents a cancer incidence of 11.6% (95% CI 11.0-12.3) in men and 6.2% (95% CI 5.9-6.5) in women (Table 5.4).

5.8.1.2 Normal platelet count cohort

There were 225 diagnoses of cancer recorded in either the CPRD or the cancer registry in the normal platelet count cohort; this represents a cancer incidence of 4.1% (95% CI 3.4-4.9) in men and 2.2% (95% CI 1.8-2.6) in women (Table 5.4).

Table 5.4: Comparison of the number of incident cancer cases in each cohort when data from both the CPRD and the cancer registry, or only the cancer registry, are included.

		Thrombocytosis		
		Men		**Women**
Included diagnoses	N	n cancers diagnosed Incidence % (95% CI)	N	n cancers diagnosed Incidence % (95% CI)
Either source	9,435	1,098 11.6 (11.0-12.3)	21,826	1,355 6.2 (5.9-6.5)
Cancer registry	9,333	1,021 10.9 (10.3-11.6)	21,668	1,265 5.8 (5.5-6.2)

		Normal platelet count		
		Men		**Women**
Included diagnoses	N	n cancers diagnosed Incidence % (95% CI)	N	n cancers diagnosed Incidence % (95% CI)
Either source	2,599	106 4.1 (3.4-4.9)	5,370	119 2.2 (1.8-2.6)
Cancer registry	2,580	93 3.6 (2.9-4.3)	5,345	109 2.0 (1.7-2.4)

5.8.2 Sensitivity analysis: one year cancer incidence

The number of cancer registry recorded diagnoses was determined for each group. These and the respective incidences are presented in Table 5.4.

There were 2,488 qualifying records of cancer diagnoses in the cancer registry. The majority of these (2,286, 91.9%) were in patients with thrombocytosis; 1,021 in men with thrombocytosis and 1,265 in women with thrombocytosis. In patients with thrombocytosis, the one year incidence of cancer in men with thrombocytosis, when only cancer registry-recorded diagnoses are included, was 10.9% (95% CI 10.3-11.6) for men and 5.8% (95% CI 5.5-6.2) for women. In patients with a normal platelet count, there were 93 records of diagnoses in men and 109 in women. The one year cancer incidence in this subgroup was 3.6% (95% CI 2.9-4.3) for men and 2.0% (95% CI 1.7-2.4) for women.

5.8.3 Comparing one year incidence when all recorded diagnoses or only cancer registry recorded diagnoses are included

Including only cancer registry-recorded diagnoses in the one year incidence analysis resulted in 190 fewer diagnoses; 90 fewer in male patients and 100 fewer in female patients. The one year incidence decreased in all groups by less than 1%; a decrease of 0.7% for men with thrombocytosis, 0.4% decrease for women with thrombocytosis, 0.5% for men with a normal platelet count, and 0.2% for women with a normal platelet count.

This plus the overlapping confidence intervals between measures for each group suggests that the one year incidence of cancer does not greatly change in any of the subgroups as a result of including only cancer-registry recorded diagnoses in the analysis; and that there is little overestimation in incidence figures as a result of including CPRD-recorded diagnoses that were not confirmed by the cancer registry.

5.9 Chapter discussion

The validation study presented in the first part of this chapter has encouraging results; 5,924 of 7,028 cancer diagnoses recorded in the CPRD were confirmed by the cancer registry (a PPV of 84.3%). This result is supported by Boggon *et al.* (2013) who found that 4,830 of 5,797 CPRD recorded cancers were confirmed by the cancer registry (83.3%). In that study, there were more cases of colorectal, lung, urinary tract, and pancreatic cancer recorded in the cancer registry than the CPRD; lung and pancreatic cancer have particularly poor prognosis. Patients with these diagnoses may be more likely to die in hospital soon after their diagnosis, and details of this may not be fed back to their primary care record. That study also found that cancers diagnosed through routes other than histology (such as myeloma and leukaemia) were under-recorded in the cancer registry; some cancer registries are typically over-reliant on histology. The majority (528 out of 967, 54.6%) of CPRD cancer records that were not validated by a cancer registry record in Boggon *et al.* (2013) were validated by other means; hospital records or practice records. CPRD cancers that are not confirmed by the cancer registry cannot be assumed to be false records.

Dregan *et al.* (2012) found a higher level of concurrence between the CPRD and the cancer registry; 92% of CPRD cancers were confirmed by the registry data. However, that study included only colorectal, gastro-oesophago, respiratory, and urinary tract cancers. These four sites were chosen as the validation study builds on an earlier study from the same authors that investigated the incidence of cancer in patients with haematuria, haemoptysis, dysphagia, and rectal bleeding. Comparing the present study with Dregan *et al.* (2012)'s work, it appears that records of some types of cancer in the CPRD are more valid than others.

The present study found that sex, year of diagnosis, and age group at diagnosis were significant predictors of concordance in recording between the CPRD and the cancer registry. Boggon *et al.* (2013) also examined predictors of concordance, and found that age was predictive (with less CPRD recording with greater age). Sex did not predict concordance in that study, and year of diagnosis was not examined.

It is possible that some patients have codes in their CPRD records which indicate a cancer diagnosis, when they are not a true case. These errors can be caused by processes

or errors in the healthcare system, or data recording issues. Healthcare system issues include diagnoses being mistakenly recorded in the CPRD; Dregan *et al.* (2012) found that 77% of incorrect CPRD cancer records had an alarm symptom recorded; this could have resulted in mistakenly diagnosed cancer, or a 'suspected' cancer record being picked up in research studies as a diagnosis. Disagreement between the two sources on the primary site of diagnosis could also be attributed to erroneous recording if the first wrongly diagnosed site was recorded in the CPRD, and the correct site later recorded; the first cancer in the patient's records would not match the later, correct, cancer registry record. The two data sources collect data in different ways at different points in the healthcare system; the CPRD data are updated monthly with multiple consultations and entries per diagnosis, whereas the cancer registry aims to make one single record for each cancer diagnosis in each patient. Data recording issues include the possibility that a cancer registry cancer is recorded just after the end date of the patient's registration with a CPRD practice; and their diagnosis may not be fed back to their old practice. The CPRD and the cancer registry each use different coding dictionaries which could result in inconsistencies, and mistakes in the patient ID number used to link the two data sources may mean that a CPRD patient appears to have no record in the cancer registry, or vice versa.

The sensitivity analysis presented at the end of this chapter found that only including cancer registry diagnoses in the primary analysis (identifying the cancer incidence in patients with thrombocytosis and with a normal platelet count) did not greatly alter the results; the estimated cancer incidence for men with thrombocytosis was 11.6% (11.0-12.3) using CPRD and cancer registry diagnoses and 10.9% (10.3-11.6) using cancer registry diagnoses only. Similarly, for women with thrombocytosis the incidence was 6.2% (5.9-6.5) including CPRD and cancer registry diagnoses, and 5.8% (5.5-6.2) using cancer registry diagnoses only.

5.10 Strength and limitations

Although this study used a large sample of 50,000 patients from the CPRD, 40,000 of the patients in the sample were selected based on their having had at least one platelet count showing thrombocytosis; therefore, the sample is not generally representative of all CPRD patients. Therefore, whilst this chapter does examine the validity of the CPRD data sample used in Chapter 4, the findings have limited applicability to the CPRD as a whole. Other CPRD validation studies have also included specific populations of patients defined based on particular conditions or symptoms; Dregan *et al.* (2012) included patients with haematuria, haemoptysis, dysphagia, rectal bleeding, or lung, colorectal, gastro-oesophagic, or urinary cancer, from English practices with

registry linked data available. Boggon *et al.* (2013) included patients from English practices with cancer registry and HES linked data with diabetes.

Patients with thrombocytosis may have greater health concerns which mean that they are more likely to come in to contact with health services. One criteria for selection of patients from the CPRD for this study was that cancer registry data linkage must be available for the patients. Not all CPRD-contributing practices participate in the cancer registry linkage scheme. There could be inherent differences between practices that do and do not participate in the linkage which affect the accuracy and validity of cancer recording. A further limitation of this study is that the exact date of diagnosis is not available; only the month and year. This could in part explain some of the differences in date of diagnosis seen between the two sources, although the impact of this is likely to be small.

Additionally, the present analysis only included data from English practices. It is not possible to apply the findings to Welsh, Scottish, or Northern Irish data.

5.11 Chapter Summary

The findings of this validation study are encouraging for future studies that use CPRD data; particularly those using cancer records. Overall, the quality of cancer diagnosis recording in the CPRD was reasonably high. This is supported by results from Boggon *et al.* (2013) and Dregan *et al.* (2012). The availability of linked cancer registry data is a great advantage of CPRD data, and studies requiring accurate incidence figures should make use of this linkage to ensure that all incident cases are captured.

Chapter 6

Discussion and conclusions

6.1 Summary of main findings

The evidence drawn together in this thesis strongly supports a link between thrombocytosis and undiagnosed cancer. It suggests that for at least a proportion of patients with an underlying malignancy, platelets begin to rise in excess of normal levels before other symptoms become evident. Unlike some physical symptoms of cancer, which can be subject to patient delay in presentation, which in turn can delay diagnosis (Macleod *et al.*, 2009), platelet count is measured in general practice as part of a full blood count and is often an incidental finding. A full blood count is arguably the most commonly ordered test in general practice; around a quarter of adults in the UK have one in a year (Hamilton *et al.*, 2008).

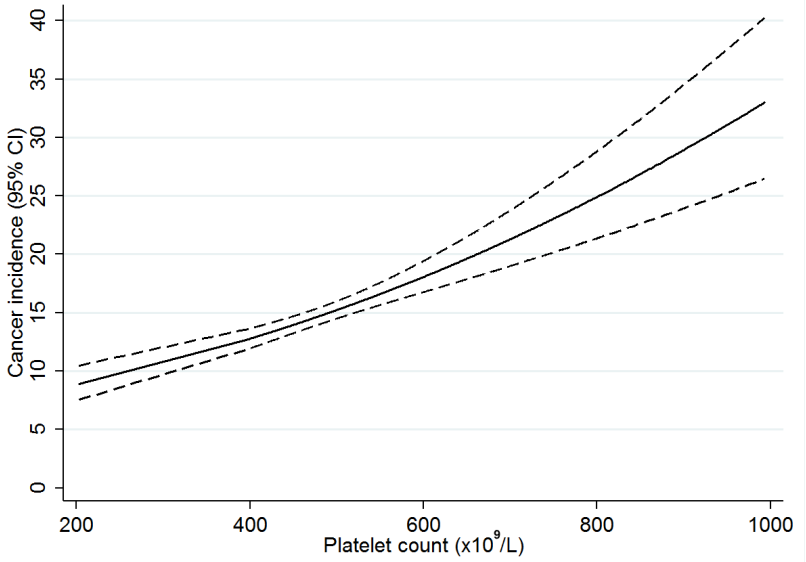
Both the systematic review in Chapter 3 and the cohort study presented in Chapter 4 investigated the relationship between thrombocytosis and cancer. The studies identified by the systematic review had only examined the relationship for some types of cancer, and found thrombocytosis to be a marker of lung, kidney, oesophago-gastric, and uterine cancer. It was not a marker of colorectal, pancreatic, ovarian, or breast cancer. The impact of sex, age, and change in platelet count over time had not been examined in any of the included studies. The cohort study in Chapter 4 addressed this gap and found that the risk of cancer increases steadily with increasing platelet count (see Figure 6.1). Thrombocytosis is more strongly indicative of cancer in men than in women, and the presence of thrombocytosis is more predictive of cancer in older rather than younger patients. A platelet count that remains elevated over time is also more strongly predictive of underlying malignancy.

The association between thrombocytosis and cancer was further examined by combining thrombocytosis with other symptoms. Thrombocytosis plus two 'vague' cancer symptoms, loss of appetite and weight loss, was more strongly indicative of underlying

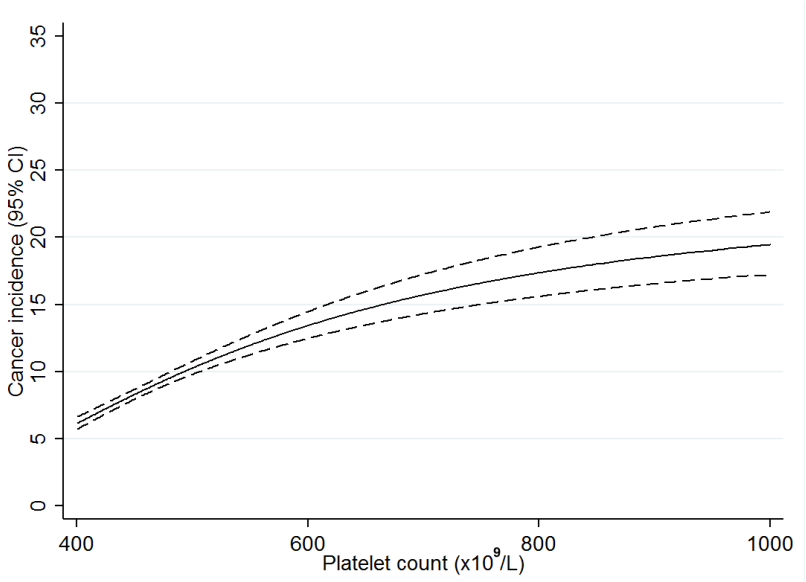
malignancy than thrombocytosis alone, as was thrombocytosis and cough or a change in bowel habit or rectal bleeding. It was revealed that at least a third of patients with thrombocytosis and lung or colorectal cancer had no other recorded clinical features of cancer in the year prior to diagnosis. The recognition of thrombocytosis as a risk marker of cancer and the inclusion of this clinical feature in national suspected cancer guidance could have a sizeable impact in practice, prompting earlier investigation and potentially earlier diagnosis before other symptoms become apparent. The validation study presented in Chapter 5 thoroughly examined the validity of cancer recording in the CPRD and differences in data recording between the CPRD and the cancer registry. This study found the quality of data recording in the CPRD to be reasonably high, and the sensitivity analysis indicated that the inclusion of non-cancer registry confirmed cases in the main analysis in Chapter 4 would not have greatly altered the results or interpretation.

6.2 Overarching findings and comparison with existing literature

The cohort study presented in this thesis is the first and only study to examine thoroughly all aspects of the relationship between thrombocytosis and cancer, and to investigate the potential clinical utility of this risk marker in practice. The systematic review (Chapter 5) found that female-only cancers (all of which are hormone-dependent) were not as strongly associated with thrombocytosis as non-hormone-dependent cancers. This was supported by the finding that prostate cancer (also hormone-dependent) was much less commonly diagnosed in men with thrombocytosis compared to men with a normal platelet count, or the general population. It is not clear whether these differences occur due to thrombocytosis being less implicated in hormone-dependent cancers (which are more commonly diagnosed in women), or whether the difference is an artefact of variation in consulting behaviour and cancer incidence between men and women in the general population. Women aged 40 years and over consult more often than men, and cancer incidence is lower in women than in men in this age group; these two factors could result in a 'dilution' of cancer incidence in women with thrombocytosis compared to men with thrombocytosis. Alternatively, this difference could be due to thrombocytosis in women having more benign causes. There is also evidence that women have higher baseline platelet counts than men (Bain, 1996; Biino *et al.*, 2013; Gader *et al.*, 1995; Graham *et al.*, 1987; Segal & Moliterno, 2006; Stevens & Alexander, 1977). Overall, thrombocytosis was a much stronger risk marker of non-hormone-dependent cancers, and a stronger marker in men than in women.



(a)



(b)

Figure 6.1: Fractional polynomial logistic regression model with platelet count as a continuous predictor variable. (a) increasing cancer incidence (and 95% confidence intervals) with increasing platelet count in males. (b) increasing cancer incidence (and 95% confidence intervals) with increasing platelet count in females.

The studies identified by the literature review in Chapter 3 examined the relationship between thrombocytosis and cancer in a retrospective manner; starting with patients diagnosed with a particular type of cancer and looking back to examine the proportion that had thrombocytosis in the year prior to their diagnosis. The cohort study in Chapter 4 examined the association prospectively; starting with patients with thrombocytosis and observing which patients went on to be diagnosed with cancer, compared to a group of patients with a normal platelet count.

Another key difference between the case-control studies in Chapter 3 and the cohort study in Chapter 4 is the population being studied. Although both included patients registered with and attending a CPRD-subscribed general practice, the case-control patients had to have consulted one or more times within the two years prior to their cancer diagnosis. In the cohort study, patients were eligible for inclusion only if they had a blood test. There was no blood test criteria for inclusion in the case-control studies. These case-control studies are therefore likely to include patients with less illness and less likely to be diagnosed with cancer than the cohort study. It is reasonable to expect differences between the two study types in their findings on the association between thrombocytosis and cancer, due to these differences in the study population.

Nine cancer sites were investigated in studies found by the systematic review, and 20 cancer sites in the cohort study. Evidence from three other publications (concerning lung (Pedersen & Milman, 1996), ovarian (Stone *et al.*, 2012), and any type of cancer (Davis & Mendez Ross, 1972)) was not included in the systematic review as these were hospital based, so the included patients were further along the diagnostic pathway. As a patient is more likely to be diagnosed with cancer from the point of referral compared to in a primary care consultation, these studies would have found a much higher incidence of cancer in patients with thrombocytosis, which would have introduced bias to the results. However, the evidence from these studies is included below to give a wider context.

The combined evidence for each cancer site is presented below.

6.2.1 Thrombocytosis and lung cancer

In a study of 100 patients with thrombocytosis seen in hospital for undisclosed reasons, 4 were diagnosed with lung cancer (4%) (Davis & Mendez Ross, 1972). In that study thrombocytosis was defined as a platelet count $> 500 \times 10^9$/L and as the study was carried out in a hospital setting, the incidence of cancer in the study population was likely to be higher than that in primary care. A study of patients with chest x-ray abnormalities warranting further investigation for lung cancer found thrombocytosis in 57% of patients with malignant disease, and 8% with benign disease (Pedersen &

Milman, 1996); this fits with findings from the case-control study identified in Chapter 3 in which 14% of lung cancer patients had thrombocytosis prior to diagnosis (Hamilton *et al.*, 2005a). Data from the latter study were used to calculate the likelihood ratio of thrombocytosis for lung cancer; this was 8.9 (95% CI 5.19-15.41).

The independently calculated PPV of thrombocytosis for lung cancer using data from that study was 1.63 (0.92-2.90). In the cohort study in Chapter 4, the incidence of lung cancer in male patients with thrombocytosis was 3.2% (2.9-3.6); this represents an absolute increase of 2.8% compared to the risk in patients with a normal platelet count (which was 0.4% (0.2-0.8)). For female patients there was also an increased risk associated with a raised platelet count; lung cancer incidence in those with thrombocytosis was 1.1% (0.9 - 1.2), an absolute increase of 0.8% compared to that in females with a normal platelet count (0.3%, 0.1 - 0.4). The risk of lung cancer increased if patients with thrombocytosis presented with a cough; in patients with thrombocytosis and a cough, the risk of a lung cancer diagnosis in the following year was 3.9% (3.2-4.7).

The work presented in this thesis concerning platelet count and lung cancer suggests that the platelet count rises progressively in the months before diagnosis in some patients with lung cancer. Crucially, around a third of patients with lung cancer had no clinical features of malignancy other than thrombocytosis in the year prior to their diagnosis. Staging data in the study by Pedersen and Milman showed that thrombocytosis was more common in patients with advanced disease (Pedersen & Milman, 1996). The proportion of lung cancer patients with thrombocytosis rose from 14% (Hamilton *et al.*, 2005a) in a primary care setting to 57% in a secondary care setting (Pedersen & Milman, 1996). Overall, this evidence strongly supports a diagnostically useful association between thrombocytosis and undiagnosed lung cancer; one which could be used to identify cancer at an earlier disease stage, before other symptoms become evident.

6.2.2 Thrombocytosis and colorectal cancer

The previously mentioned study by Davis & Mendez Ross (1972), of 100 patients with thrombocytosis seen in hospital for undisclosed reasons, found that 4 patients were diagnosed with colon cancer (4%). The case-control study identified by the systematic review in Chapter 3 did not find thrombocytosis to be a significant predictor of colorectal cancer in multivariable models. This was one of the smaller studies included in the review, with 349 cases and 1,744 controls. In Chapter 4, the incidence of colorectal cancer was markedly higher in men and women with thrombocytosis compared to those with a normal platelet count. For men with thrombocytosis, the incidence of colorectal cancer was 2.7% (2.4-3.1); this represents an absolute increase of 2.3% from the incidence in patients with a normal platelet count, which was 0.4% (0.2-0.7). For

women with thrombocytosis, the incidence of colorectal cancer was 1.6% (1.4-1.7); an absolute increase of 1.2% from the value of 0.4% in women with a normal platelet count (0.3-0.6). These incidence values were higher in patients with a record of rectal bleeding (8.4%, 95% CI 5.8-11.9) or a change in bowel habit (5.7%, 95% CI 4.8-6.7) although this may be of limited clinical use since these symptoms alone may be enough to prompt investigation for suspected cancer in practice. Certainly, NICE guidance recommends referral for suspected colorectal cancer for any patients with rectal bleeding aged 50 years and over, or for patients aged 60 years and over with a change in bowel habit (NICE, 2015).

There are several factors that could explain the difference in findings between the case-control study identified by the literature search in Chapter 3, and the cohort study presented in Chapter 4. The former found that thrombocytosis was not a predictor of underlying colorectal cancer, whereas the latter found an association between thrombocytosis and undiagnosed colorectal cancer. The two studies included large samples; 2,093 in the case-control study and 39,230 in the cohort study. There are also differences between the patients included in each sample. In the case-control study, patients were eligible for inclusion if they had a consultation recorded in the two years prior to their cancer diagnosis. This consultation did not necessarily include a full blood count. In that study, patients with no platelet count available were assumed to have a normal platelet count and categorised as such. In the cohort study, patients were selected for inclusion on the basis of their platelet count; so all included patients had a full blood count. They are therefore likely to have more illness than the case-control sample, who may not all have had blood tests. The base rate of cancer is likely to be different in each sample. This difference may affect the association seen (or not seen) between platelet count and cancer in each of the two studies.

6.2.3 Thrombocytosis and hormone-dependent cancers

The evidence surrounding thrombocytosis and various hormone-dependent cancers is mixed.

Ovarian cancer was the most commonly diagnosed cancer ($n = 7$) in the Davis & Mendez Ross (1972) study of 100 patients with thrombocytosis; there were also five cases of breast cancer in the cohort. 60% of patients in that study were female. Another study found that 31% of 619 patients with ovarian cancer had thrombocytosis at the point of diagnosis (Stone *et al.*, 2012), and others have found thrombocytosis to be a predictor of poor prognosis for ovarian cancer (Allensworth *et al.*, 2013). The ovarian cancer case-control study identified by the literature search in Chapter 3 did not find thrombocytosis to be a significant predictor of undiagnosed ovarian cancer. In

the cohort study in Chapter 4, the risk of ovarian cancer was higher in women with thrombocytosis (1.3%, 95% CI 1.1-1.4) than in women with a normal platelet count (0.4%, 95% CI 0.3-0.6).

In the cohort study in Chapter 4, the incidence of uterine cancer did not differ between women with thrombocytosis or with a normal platelet count (thrombocytosis: 0.2%, 95% CI 0.2-0.3, normal platelet count: 0.2%, 95% CI 0.2-0.4). For cervical cancer, there was also no difference in risk between women with thrombocytosis (0.1%, 95% CI 0.0-0.1) and with a normal platelet count (0.0%, 95% CI 0.0-0.1). Cervical cancer differs from other hormone-dependent cancers discussed in this section, in that the majority of cases are diagnosed in women aged under 45 years; the incidence is at its highest in women aged 25-30 (Cancer Research UK, 2015b). This does not match the age profile of the cohorts investigated in Chapter 4 of this thesis; the median age of patients in the cohort was 62 years of age (and the minimum age was 40 years). There may be too few incident cases in the cohort for any association with thrombocytosis to be evident (there were 10 cases in the thrombocytosis cohort); or it may be a true finding. Further investigation is needed.

14% of cervical cancers are diagnosed through the national screening programme (National Cancer Intelligence Network, 2010), and the majority of cervical cancers are diagnosed at stage I or stage II (Cancer Research UK, 2015b). Therefore, this cancer type may generally be diagnosed earlier, before symptoms have developed, and possibly before the platelet count starts to rise. Additionally, the vast majority of cervical cancers are caused by the human papilloma virus (Cancer Research UK, 2014c); this causal mechanism may be independent of any platelet-producing pathway.

There was also little difference in the incidence of breast cancer between women with thrombocytosis (0.4%, 95% CI 0.3-0.5) and with a normal platelet count (0.5%, 95% CI 0.3-0.7). Breast cancer also accounted for a much smaller proportion of all cancers diagnosed in women with thrombocytosis than in the general population (6% vs 30%). Breast cancer has an extensive screening programme in the UK. The majority of breast cancers are diagnosed in women in their 60s, 31% are identified through screening programmes, and 84% are either stage I or stage II at diagnosis (based on staging data available for 84% of incidence cases in 2013) (Cancer Research UK, 2015b). Therefore, many breast cancers are diagnosed before symptoms develop, possibly before the platelet count begins to rise. Breast cancer also has an alarm symptom, a breast lump, which has a PPV of 4.8% for women in their 40s and 48% for women in their 70s and over (Walker *et al.*, 2014) so thrombocytosis may not be a clinically useful addition, diagnostically.

Male hormone-dependent cancers include testicular and prostate cancer. There were only two cases of testicular cancer in the thrombocytosis cohort, and none in the normal

platelet count cohort. This is unsurprising as testicular cancer incidence is highest in men aged 25-40 years and declines with age; this does not match the age profile of the cohorts investigated in this study. The systematic review did not find any papers that had investigated the early markers of testicular cancer and included thrombocytosis, and no studies from other healthcare settings were found. Prostate cancer incidence was only marginally higher in patients with thrombocytosis (1.4%, 1.2-1.7) compared to those with a normal platelet count (1.1%, 0.7-1.6). No other studies were found that had investigated thrombocytosis and prostate cancer.

6.2.4 Thrombocytosis and other types of cancer

The evidence for thrombocytosis and other types of cancer is mixed. Previous studies found thrombocytosis to be a marker of kidney and oesophago-gastric cancer, in addition to the previously discussed cancers. The cohort study in Chapter 4 did not find a strong association with kidney cancer. Oesophageal and stomach cancer were investigated as separate cancer sites; combining the results for these two sites gives an incidence of 1.2% (95% CI 0.9-1.6) in male patients with thrombocytosis and 0.2% (95% CI 0.1-0.4) for men with a normal platelet count; an absolute increase of 1.0%. In the case-control study of the clinical features of oesophago-gastric cancer, thrombocytosis was a significant predictor of malignancy in multivariable models (OR 2.4 (95% CI 2.0-2.9). The incidence of other types of cancer investigated in the cohort study was low, and there is very little (or no) evidence from other previously published studies.

6.3 Limitations

6.3.1 Selection bias

One of the major limitations of the cohort study (Chapter 4) is that patients having a blood test are a selected group, who are more likely to be 'ill', and therefore more likely to be diagnosed with cancer than patients attending primary care without a blood test, or the general population. The cancer incidence would be artificially high in this selected group; this is why the findings presented here do not support the use of a platelet count as a diagnostic test for cancer in patients who otherwise would not be tested; this would require a very different type of study. In an attempt to address this bias, the cancer incidence in patients with thrombocytosis was compared to that in patients with a normal platelet count. This group is a more suitable comparator as they have also been selected for a blood testing, and therefore have a comparable baseline risk of cancer. The overall risk of cancer in patients with thrombocytosis is the primary objective of the cohort study (and not the absolute increase in risk from a normal

platelet count). This is the most clinically useful outcome; in practice, GPs will need to know the risk of cancer in the patient they are currently seeing, not how much more at risk they are compared to another patient. However, the absolute difference in the risk of cancer between patients with and without thrombocytosis enables the additional value of a raised platelet count, above a normal platelet count, to be demonstrated.

6.3.2 Reasons for testing

A further limitation of this work is that the reasons for full blood counts being ordered are not known in either cohort of patients. Patients with raised platelet count could be presenting with a clinical feature which has already prompted the GP to consider cancer, and the decision to order a blood test is responsive to that. As a proxy for this, the symptoms reported by patients in the month prior to their index date (those which could have prompted a blood test) were compared between the two cohorts. Some of the most commonly reported symptoms were indicative of cancer, including cough, fatigue, abdominal pain, diarrhoea, and shortness of breath. The proportion of patients reporting these symptoms was marginally higher in the thrombocytosis cohort than in the normal platelet count cohort. However, the proportion reporting a single symptom was less than 4% (Table 4.4). With the exception of cough (reported by 3.9% of thrombocytosis cohort and 1.9% of the normal platelet count cohort), there was a difference of less than 1% in the proportion reporting the symptom between the two cohorts. This suggests that there is little difference between the cohorts in cancer alarm symptoms that could have prompted a blood test in the month before their platelet count index date.

6.3.3 Sample size

The cohort study in Chapter 4 found that some cancers were more strongly associated with thrombocytosis than others. Whilst a true lack of association between thrombocytosis and some types of cancer could be responsible for the low incidence of some cancer types, it could also be due to too few cases of these types of cancer being identified in the sample. Power calculations carried out prior to the commencement of this PhD found that $40,000$ patients with thrombocytosis was sufficient to estimate an incidence of any type of cancer of 5% with a margin of error no greater than 0.22%, using the upper bound of the 95% confidence interval. $10,000$ patients was sufficient to estimate the same incidence of cancer with a margin of error no greater than 0.45% similarly. To accurately estimate the incidence of less common cancers may require a larger sample size, and may be of limited clinical use.

6.3.4 CPRD data quality

A further limitation of this work, and of six of the nine case-control studies identified in the systematic review, is that it relies on the accuracy and validity of CPRD data. Furthermore, only patients registered at CPRD-contributing practices that had subscribed to the cancer registry linkage scheme were included in the cohort. This excludes patients who are at non-linkage subscribed practices. There may be inherent differences between practices that do and do not subscribe to the cancer registry linkage scheme that could introduce bias to the study; practices with high cancer awareness and cancer investigation rates may be more likely to subscribe to the scheme, although there is no evidence that this is so. As earlier described, there is evidence to suggest that cancer recording is of reasonably high quality in the CPRD (Boggon *et al.*, 2013; Dregan *et al.*, 2012). The inclusion of linked cancer registry data in this study is a particular strength; as is the testing of the validity and diagnostic accuracy of the CPRD data using the linked registry data (Chapter 5). This study found high rates of cancer recording in the CPRD. The sensitivity of CPRD cancer recording was 76.1%, and the PPV of a CPRD cancer diagnosis was 84.3%. This included all types of cancer; the figure may be higher for some types of cancer, as suggested by Dregan *et al.* (2012). The sensitivity analysis which repeated the primary analysis from Chapter 4 found that the inclusion of non-cancer registry confirmed CPRD-recorded cancers did not greatly inflate the results.

The data sample in this validation study included cancer registry data from 2000-2010. The UK Department of Health's Improving Outcomes: a strategy for cancer policy, published in 2011, initiated a drive towards better recording of cancers in the cancer registry. It may be that more recently collected registry data are of higher quality or more complete. Results presented in Chapter 5 suggested that CPRD cancers recorded from 2005 onwards were more likely to be verified by cancer registry records. CPRD cancers recorded in male patients were more likely to be matched by cancer registry data than cancers in female patients, as were those recorded in younger patients. These two factors are likely to be linked; cancer is more likely to be diagnosed in women in older age groups.

6.3.5 Missing cancer staging data

The relative incompleteness of data in the 2000-2010 timeframe is illustrated by the lack of satisfactory staging data; staging was available for only half of cases. The lack of staging data and subsequent lack of staging analysis is a weakness of this work. Thrombocytosis as a marker of cancer is less useful in terms of expediting diagnosis to improve outcomes if it only identifies late stage cancers, although these are also worth

diagnosing, even if the prognosis is poor. The staging data that were available only suggest that at least a quarter of thrombocytosis cancers are early stage. This is a factor that needs to be investigated in future research.

6.3.6 Misclassification bias

Laboratory test results for the platelet count were used to define the two cohorts of patients in Chapter 4. These results are electronically transmitted to patient records, and so any misclassification bias resulting from this process is likely to be non-differential, and small. Three of the 50,000 patients in the cohort were misclassified due to errors in platelet count recording. In all three cases, a normal platelet count had three extra 0s mistakenly added so that the platelet count appeared to exceed the normal value. Patients were therefore incorrectly classed as having thrombocytosis. Misclassification bias in outcome reporting in CPRD data is more likely; patients may be incorrectly coded as having cancer if a cancer medcode was used mistakenly by the GP, or if a 'suspected' or 'absent' coded cancer was incorrectly taken as cancer being present. Such records are rarely, if ever, removed from the CPRD record. In a previous CPRD data validation study, 77% of incorrect CPRD cancer records had an alarm symptom recorded; this could by explained by a 'suspected' cancer record being picked up in research studies as a diagnosis (Dregan *et al.*, 2012). In the present study, the use of linked cancer registry data attempts to ameliorate this bias by providing a confirmation of cancer diagnoses.

6.3.7 Detection bias

The main exposure variable in this study (platelet count) relies on electronically coded data, which are unlikely to be affected by detection bias. However, some of the secondary analyses presented in this thesis rely on symptom data coded by GPs in consultations; specifically, concerning cough, change in bowel habit, rectal bleeding, weight loss, and loss of appetite. These analyses could be affected by detection bias. Whilst all symptoms should be recorded in practice using the numerical Read codes (Booth, 1994), there is evidence to suggest that GPs do not record all symptoms in this way. A CPRD study by Price *et al.* (2016) found that 38% of clinical features were recorded by GPs only in the free text section of the patients' medical records, and not coded using Read codes. In that study there was some differential loss of recorded symptom data to the free text for visible haematuria and jaundice. Patients with cancer were more likely to have these symptoms recorded using Read codes, whereas patients without cancer were more likely to have these symptoms recorded in the free text section of their CPRD records. These differential coding differences may apply to some patients

included in samples used in this thesis. However, given the similarities between the recorded symptoms in patients with thrombocytosis and a normal platelet count, any effect of this should be minor.

6.3.8 Generalisability of results

The generalisability of these results outside of England and the UK is limited. The systematic review found only UK-based studies, and the cohort study included data from English practices only. Despite extensive efforts to identify international researchers who may have relevant data or be conducting similar investigations, none were found. A recently initiated study in Sweden is collecting data on 200 consecutive cases of colorectal, lung, and urethral cancer from seven Swedish regions (personal communication: Hans Thulesius). Preliminary results are showing thrombocytosis in 14% of lung cancer patients; this matches exactly the estimate from the case-control study of clinical features of lung cancer identified by the systematic review (Hamilton *et al.*, 2005a). Despite the lack off data from outside of England, it is unlikely that the relationship between thrombocytosis and cancer would only be seen in patients registered at English general practices, although factors such as ethnicity and cultural differences in when patients choose to consult with symptoms may have an impact. The over-representation of general practices in the South East and North West of England limits the applicability of these findings to other regions of England. This could introduce bias if there were significant geographical patterns in cancer diagnosis in the UK; however, data published in 2016 (NHS England, 2016) show all regions of the UK performing poorly in cancer diagnosis, with no geographical trend.

6.3.9 Publication and reporting bias

Five of the nine studies included in the systematic review were not identified by the literature search as they did not report on thrombocytosis (although they had collected the data). This reflects the publication bias which is a key limitation of all systematic reviews. In four of these studies, thrombocytosis was not a predictor of cancer in the multivariable (adjusted) analysis so it was not reported. However, the identification and inclusion of these studies (and inclusion of negative unpublished results) is a strength of the systematic review presented in this thesis. It is probable that there are other unidentified studies that collected, but did not report, platelet count data that were not identified through contact with experts and networking. The exclusion of non-English language papers due to a lack of translating facilities is a further limitation of the systematic review. The search strategy itself included three groups of search terms relating to thrombocytosis, cancer, and primary care. The concept of primary care

differs from country to country, and not all healthcare systems are organised this way. Studies set in another region's equivalent to a primary care setting may not have been identified if this setting is described differently.

6.3.10 Platelet count reference range

The range for a normal platelet count of 150 - 400 $\times 10^9$/L, and threshold of 400 $\times 10^9$/L for thrombocytosis, was chosen in this study to encompass both definitions used in English laboratories (400 or 450 $\times 10^9$/L). However, a growing body of evidence showing statistically and clinically important differences in the normal platelet count range for different sexes (Bain, 1996; Biino *et al.*, 2013; Gader *et al.*, 1995; Graham *et al.*, 1987; Segal & Moliterno, 2006; Stevens & Alexander, 1977), ages (Biino *et al.*, 2013; Segal & Moliterno, 2006), and ethnic groups (Bain, 1996; Gader *et al.*, 1995; Segal & Moliterno, 2006) has prompted some researchers to propose new tailored reference ranges for platelet count (Biino *et al.*, 2013). The largest study to date, Biino *et al.* (2013), included 40,987 individuals from nine regions in Italy. The sample comprised 46.2% males, with a mean age of 50.7 years (SD 17.5, range 10 months to 105 years). That platelet count is consistently higher in women and declines with age is reflected in the new reference ranges proposed by this group. They are stratified by age and sex:

- Men aged 15-64 years, normal range 141 - 362 $\times 10^9$/L (mean 238, SD 57.9)
- Women aged 15-64 years, normal range 156 - 405 $\times 10^9$/L (mean 264, SD 65.3)
- Men aged 64 years and over, normal range 122 - 350 $\times 10^9$/L (mean 220, SD 59.5)
- Women aged 64 years and over, normal range 140 - 379 $\times 10^9$/L (mean 245, SD 61.2)

These findings suggest that a platelet count of under 362 $\times 10^9$/L should only be considered 'normal' for women, and that in men, the upper limit for a normal platelet count should lie closer to 350 $\times 10^9$/L than 400 $\times 10^9$/L. Using these revised thresholds to define patients with thrombocytosis in the present study (Chapter 4) may have produced quite different results; and accepting a lower threshold of platelets to prompt cancer investigation could increase the clinical utility of this test result.

6.4 Implications for practice

Cancer incidence figures of 11.6% in men and 6.2% in women with thrombocytosis are very high; well above the threshold value of 3% risk used for urgent investigation of

suspected cancer in recent UK guidance (NICE, 2015), and even further above the 1% risk at which patients would like to be investigated for suspected cancer (Banks *et al.*, 2014). National cancer incidence data show that the overall risk of cancer for an adult aged 40 years and over is around 1.5% (Cancer Research UK, 2016). Results presented in this thesis show that the risk of cancer in patients with thrombocytosis is 11.6% (95% CI 11.0-12.3) for male patients and 6.2% (95% CI 5.9-6.5) for female patients; well above the 'background rate' of cancer in the general population. In comparison, the PPV of a breast lump in women aged 50-59 years is 8.5% (95% CI 6.7-11.0) (Walker *et al.*, 2014), and the PPV of hypercalcaemia in men aged 40 years and over is 11.5% (95% CI 11.1-11.9) (Hamilton *et al.*, 2014).

Notably, patients with a recorded normal platelet count also appear at an increased risk of cancer compared to the general population. The present study found that the risk of cancer for male patients with a normal platelet count is 4.1% (95% CI 3.4-4.9), and 2.2% (95% CI 1.8-2.6) for female patients. This diagnostic artefact reflects the fact that patients who experience symptoms and are unwell enough to consult primary care are different to (and more likely to be ill than) the general population. A subgroup of these consulting patients will cause enough concern for their general practitioner to be sent for a blood test. In that respect, this select group are more likely to have a medical problem than a non-consulting patient. It is, therefore, the difference in cancer incidence between those with normal platelet counts and those with thrombocytosis that represents the additional risk of a cancer diagnosis in those with a raised platelet count.

The absolute increase in risk associated with thrombocytosis above that associated with a normal platelet count is 7.5% in men and 4.0% in women. Whilst it could be argued that this figure more accurately represents the usefulness of thrombocytosis in predicting cancer, it is actually the overall risk that prove the most useful in general practice. In a consultation, GPs will generally consider the risk of cancer in a symptomatic patient as an individual, taking their symptoms and test results into consideration, and through a process of diagnostic reasoning consider whether the patient's risk is great enough to warrant further investigation. Knowing the overall risk associated with thrombocytosis is therefore more clinically valuable than the absolute increase in risk.

The risk figures estimated in this work may seem small, but the predictive value of most cancer symptoms for cancer is low. To put these findings into context, the PPV of a breast lump for a diagnosis of breast cancer in a woman aged 50-59 years in primary care is 8.5% (95% CI 6.7-11.0%) (Walker *et al.*, 2014). The risk of lung cancer in patients with haemoptysis is 3.5% (95% CI 1.6-7.5) in those aged 40 years and over (NICE, 2015). This is similar to the risk of colorectal cancer in those with

rectal bleeding (8.1% (95% CI 6.0-11.0)) in over 50s (Astin *et al.*, 2011). The PPV of an elevated PSA level is 17.9% (Grubb *et al.*, 2008). The incidence of cancer in patients with thrombocytosis is similar to that in patients with hypercalcaemia, which has a one year incidence of 11.5% (95% CI 11.1-11.9) in men and 4.1% (95% CI 3.9-4.4) in women (Hamilton *et al.*, 2014).

From a policy perspective, these results may be of even more value. Thrombocytosis has not generally been viewed as a risk marker for cancer, at least until the 2015 update of the UK national guidance for suspected cancer included it as a marker of lung, oesophago-gastric, and uterine cancers (NICE, 2015). In the studies that underpinned that guidance, the proportion of cancer patients with thrombocytosis ranged from 4-14% (Hamilton *et al.*, 2005a; Shephard *et al.*, 2013; Stapley *et al.*, 2013; Walker *et al.*, 2013); reviewed in Chapter 3. In those studies, patients who had no platelet count available were merged with those who have a normal platelet count, so it is likely the true proportion of patients with cancer and thrombocytosis is higher. Results presented in this chapter show that at least a third of lung and colorectal cancer diagnoses could possibly be expedited by at least two months with the inclusion of thrombocytosis in national guidance. Based on this, even if only a conservative estimate of 5% of patients with cancer have thrombocytosis before diagnosis (16,500 out of 330,000 new cases annually in the UK), a third of them have the potential to have their diagnosis expedited by at least three months; equating to 5,500 earlier diagnoses annually.

6.5 Implications for research

The validation study and sensitivity analysis presented in Chapter 5 have strong implications for future research. Results from this study suggest that cancer recording in the CPRD is generally of a good standard with a high diagnostic value. Researchers should be aware that some types of cancer appear to be more reliably recorded than others; for example, the study by Dregan *et al.* (2012) found CPRD data to have a higher diagnostic value than the present study, but they had only included four types of cancer. Non-melanoma skin cancer (excluded in the present study), and cancers diagnosed through non-histological routes are particularly susceptible to under recording in the cancer registry. The present study found that 5,924 of 7,785 (76.1% sensitivity) cancer registry recorded diagnoses were picked up by the CPRD. This suggests that future research using CPRD data that requires accurate incidence figures should obtain cancer registry linked data to ensure all incident cases are included.

There are four main areas where future research could follow on from this work. Firstly, a large scale study is needed to examine the stage at diagnosis for patients with thrombocytosis and cancer. The present study failed to meet the objective of

Lung Cancer Assessment Tool for Smokers

Cough	Fatigue	Dyspnoea	Chest pain	Loss of weight	Loss of appetite	Thrombocytosis	Abnormal spirometry	Haemoptysis	
0.9	0.8	1.2	1.3	2.1	1.8	4.2	4.0	4.5	Risk as a single symptom
1.3	1.0	1.4	0.9	2.3	2.8	6.5	3.6	3.9	Cough
	1.2	1.4	1.3	2.0	2.3	2.4	>10	6.1	Fatigue
		1.5	2.2	3.1	5.5	2.4	>10	6.9	Dyspnoea
			1.4	4.4	7.6	>10	>10	4.1	Chest pain
				1.7	5.0	>10	>10	*	Loss of weight
					2.7	*	*	*	Loss of appetite
							*	12	Haemoptysis

Figure 6.2: Risk Assessment Tool for lung cancer in smokers. This tool plots several clinical features against one another and gives the risk of cancer in patients presenting with both symptoms.

investigating the stage at diagnosis due to missing data. Identifying the clinical features of patients with thrombocytosis and early stage disease would have the potential to make a significant contribution to earlier diagnosis.

Future research should focus on combining thrombocytosis with other clinical features to improve the predictive value. Various studies have shown that incorporating more than one clinical feature leads to higher predictive values; this is the basis of the Risk Assessment Tools produced by several studies by Hamilton *et al.*. An example of one of these Risk Assessment Tools is shown in Figure 6.2; thrombocytosis as a single finding carries a risk of 4.2, but this risk increases when combined with cough, chest pain, or weight loss. Other abnormal readings in blood tests have been found to be associated with increased risk of cancer; hypercalcaemia (Hamilton *et al.*, 2014; Stewart, 2005) and iron-deficient anaemia (Hamilton *et al.*, 2008) could be combined with thrombocytosis to give higher predictive values.

As discussed in the limitations section of this chapter, new reference ranges for platelet count have recently been proposed (Biino *et al.*, 2013). Using a lower threshold to define thrombocytosis could have an important impact on the cancer incidence observed in patients with thrombocytosis; results from the present study suggest that patients with platelet counts at the upper limit of normal could be at increased risk of cancer. Further research would be needed to investigate this hypothesis, examining cancer incidence using platelet count as a continuous variable from 300×10^9/L, or categorising in bands of 20 or 50. The results presented in Chapter 4 show that the association between platelets and cancer could begin at levels lower than 300×10^9/L;

Figure 6.1 shows that the cancer incidence in male patients was still at around 9% at platelet counts of 400×10^9/L. Those at the upper range of 'normal' could also be at increased risk of cancer. This is well worth investigating in future work.

Finally, a feasibility trial could investigate ways to integrate the findings from this thesis in to clinical practice in a way that would be clinically beneficial. The platelet count is rarely independently tested and thrombocytosis is commonly ignored in blood panels. A warning or notification to alert GPs to unexpectedly raised platelets in blood test results could prompt suspicions of cancer earlier. This would be particularly beneficial for patients who have no other symptoms that would warrant investigation for cancer.

6.6 Conclusions

The overall body of evidence presented in this thesis strongly suggests that cancer should be considered when a full blood count is received showing thrombocytosis, even if cancer was not being considered when the blood test was ordered. The risk of cancer is greater the higher the platelet count. Clinicians should be particularly mindful of cancer in patients who show a sustained high platelet count, are male, or are older. Greater suspicions are warranted if thrombocytosis is present with other vague, low risk cancer symptoms such as weight loss and loss of appetite. The range of possible cancers is wide; although lung and colorectal cancer are more likely than breast or prostate. Clinicians will have to seek additional indicators from examination and the patient's medical history to select the most appropriate route for investigation.

Whilst the findings presented in this thesis encourage clinicians to consider cancer, they do not go so far as to suggest that platelet count should be used as a diagnostic or screening test for cancer. These conclusions are based on results from data from a specific group of patients who were attending primary care and were selected for a full blood count. There is insufficient evidence so far to support the testing of platelet count as a means of identifying undiagnosed cancers, especially not in the asymptomatic general population. However, for a considerable proportion of patients with unexpectedly high platelets, this finding could indicate an underlying malignancy and has great potential to improve earlier diagnosis and survival.

6.7 Publications and presentations

6.7.1 Refereed papers

- **Bailey, S.**, Ukoumunne, O., Shephard, E., Hamilton, W. (2016) Clinical relevance of thrombocytosis in primary care: a prospective cohort study of cancer incidence using UK electronic medical records and cancer registry data. *British Journal of General Practice.* (In press).

- **Bailey, S.**, Ukoumunne, O., Shephard, E., Hamilton, W. (2016) How useful is thrombocytosis in predicting an underlying cancer in primary care?: a systematic review. *Family Practice.* pii: cmw100 [epub ahead of print].

6.7.2 Oral presentations

- **Bailey, S.**, Ukoumunne, O., Shephard, E., Hamilton, W. (2015) Thrombocytosis: an important risk marker of cancer in primary care. *Cancer and Primary Care Research International Network, Aarhus.*

- **Bailey, S.**, Ukoumunne, O., Shephard, E., Hamilton, W. (2015) Thrombocytosis: an important risk marker of cancer in primary care. *South West Society for Academic Primary Care, Birmingham.*

- **Bailey, S.**, Ukoumunne, O., Shephard, E., Hamilton, W. (2015) What evidence is there for an association between thrombocytosis and cancer in primary care? *A systematic review and meta-analysis. South West Society for Academic Primary Care, Birmingham.*

- **Bailey, S.**, Ukoumunne, O., Shephard, E., Hamilton, W. (2014) Thrombocytosis: an underused risk marker of cancer in primary care? *Cancer and Primary Care Research International Network, Winnipeg.*

6.7.3 Poster presentations

- **Bailey, S.**, Ukoumunne, O., Shephard, E., Hamilton, W. (2016) In symptomatic lung and colorectal cancer patients, could thrombocytosis allow the diagnosis to be expedited? *Cancer and Primary Care Research International Network, Boston.*

- **Bailey, S.**, Ukoumunne, O., Shephard, E., Hamilton, W. (2016) The validity of cancer recording in the Clinical Practice Research Datalink compared with UK cancer registries: a cohort study. *Cancer and Primary Care Research International Network, Boston.*

- **Bailey, S.**, Ukoumunne, O., Shephard, E., Hamilton, W. (2015) Thrombocytosis in primary care. *National Early Diagnosis Initiative, London.*

6.7.4 Prizes

- Best junior abstract, South West Society for Academic Primary Care, Cardiff (2016).
- Runner-up patient poster prize, National Early Diagnosis Initiative, London (2015).

Appendix A

Study protocol

This appendix contains a copy of the study protocol.

This can be found online at http://ore.exeter.ac.uk/repository/handle/10871/20964

A. Study protocol

How useful is thrombocytosis in predicting an underlying cancer in primary care?: systematic review protocol

Bailey SER[1], Ukoumunne OC[2], Shephard EA[1], Hamilton W[1].

1 Primary Care Diagnostics, University of Exeter Medical School, College House, St Luke's Campus, University of Exeter, Exeter, Devon, EX1 2LU.

2 PenCLAHRC, University of Exeter Medical School, South Cloisters, St Luke's Campus, University of Exeter, Exeter, Devon, EX1 2LU.

Corresponding author: Bailey SER s.e.r.bailey@exeter.ac.uk

Rationale and objective

Early diagnosis of cancer is imperative to reduce the cancer burden in the UK and improve cancer survival. Identifying early markers of cancer can help general practitioners to direct patients at the greatest risk of cancer to appropriate investigative services. A raised platelet count, or thrombocytosis, has been linked to malignancy[1 2] and identified as a marker of poor prognosis in secondary care[3-9], but there is little evidence around the importance of this marker in a primary care setting, within a diagnostic context. This review aims to identify and explore the body of evidence concerning the association between thrombocytosis and cancer in primary care. This protocol was produced using guidance from the PRISMA-P statement.[10]

Research questions

- Are adults aged over 40 with thrombocytosis at greater risk of cancer than those with normal platelet counts?
- Which cancer sites have been found to be associated with thrombocytosis in primary care, and which are not?

Inclusion criteria

- Any type of cancer or cancer site
- Studies that have investigated the association between cancer and platelets (using platelets as a diagnostic tool, not prognostic factor).
- Studies based in primary care.
- Includes adults $\geq$40y
- English language
- Any observational, cohort, case-control study or literature or systematic review (due to the nature of the research question, RCTs were excluded)
- The last 30 years

Exclusion criteria:

- Studies that have investigated thrombocytosis or platelet count as a prognostic tool or guide to therapy.

- Case studies and RCTs
    - Case studies are excluded as this review is interested in population-level or cohort based studies that can provide evidence on a large number of patients.
    - RCTs are excluded as the independent variable in this study cannot be randomised.

Search strategy

The search strategy will comprise the following elements to identify relevant papers:

- Searching of electronic bibliographies including:
    - EMBASE (OvidSP) 1996-2014 (week29)
    - Medline (OVIDpalt)
    - Web of Science
    - The Cochrane Library

- Scrutiny of reference lists of included studies
- Contact with experts in the field

Study selection criteria and procedures

References will be uploaded to reference management software (Endnote X5). Duplicates will be checked and removed. The abstracts and titles of references retrieved by the electronic searches will be screened for relevance by SB using the pre-defined inclusion and exclusion criteria. Hits will be double screened by ES. Any differences in opinion over paper to include or exclude will be resolved by discussion.

Full text copies of potentially relevant studies that appear to meet all inclusion criteria will be obtained. The retrieved articles will be assessed for inclusion by checking each article against the pre-specified inclusion and exclusion criteria. A flow chart will be produced which will outline the study selection process, and reasons for exclusion of full text papers will be detailed.

Quality assessment and data extraction strategy

Study quality will be assessed using QUADAS-2. Data will be extracted into custom made data extraction forms by SB and WH. Data concerning the number of patients with and without cancer who had pre-diagnosis thrombocytosis will be the main extracted values.

Evidence synthesis

The evidence will be presented and synthesised as appropriate using meta-analysis if appropriate, or a narrative synthesis if meta-analysis is not appropriate.

Search strategy - Embase

Thrombocytosis OR platelet* (as a key word) OR thrombocyte OR thrombocyte count OR thrombocyte volume (MeSH terms)

AND

neoplasm OR carcinoma (mesh terms) OR cancer, carcinoma or paraneoplastic as key words

AND*

primary medical care OR primary health care OR general practice OR family medicine (mesh terms)

Search strategy – Medline

Thrombocytosis OR platelet* (as key words) OR blood platelets OR thrombocytosis OR platelet count (MeSH terms)

AND

neoplasms OR carcinoma (mesh terms) OR cancer as a key word

AND*

primary health care OR family practice OR general practice (mesh terms) OR primary care, family medicine, primary medical care as key words

Search strategy – Web of Science & Cochrane Library

Thrombocytosis OR platelet* OR thrombocyte

AND

neoplasm OR carcinoma OR cancer

AND*

primary medical care OR primary health care OR general practice OR family medicine

*the search will be run with and without this final filter in the preliminary stages to see how the filter affects the results.

Endnote labels

1=include

2=exclude – not cancer

3=exclude – not investigating association with thrombocytosis

4=exclude – not primary care

5=exclude – not >40y

6=exclude – case study

7=exclude - RCT

Contributions

Study design: SB. Search strategy and execution: SB. Title and abstract screening: SB and EA. Data extraction: SB and WH. Data interpretation and analysis: all authors.

Funding

The Policy Research Unit in Cancer Awareness, Screening and Early Diagnosis receives funding for a research programme from the Department of Health Policy Research Programme. It is collaboration between researchers from seven institutions (Queen Mary University of London, UCL, King's College London, London School of Hygiene and Tropical Medicine, Hull York Medical School, Durham University and University of Exeter). OU is funded, and WH is part-funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) for the South West Peninsula at the Royal Devon and Exeter NHS Foundation Trust. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health in England.

1. Arslan C, Coskun H. Thrombocytosis in solid tumors: review of the literature. Turkish Journal of Haematology 2005;**22**(2):59-64.
2. Nash GF, Turner LF, Scully MF, et al. Platelets and cancer. The Lancet Oncology 2002;**3**(7):425-30.
3. Aoe K, Hiraki A, Ueoka H, et al. Thrombocytosis as a useful prognostic indicator in patients with lung cancer. Respiration; international review of thoracic diseases 2004;**71**(2):170-3.
4. Bensalah K, Leray E, Fergelot P, et al. Prognostic value of thrombocytosis in renal cell carcinoma. The Journal of Urology 2006;**175**(3):859-63.
5. Carr BI, Guerra V. Thrombocytosis and hepatocellular carcinoma. Digestive diseases and sciences 2013;**58**(6):1790-6.
6. Gogus C, Baltaci S, Filiz E, et al. Significance of thrombocytosis for determining prognosis in patients with localized renal cell carcinoma. Urology 2004;**63**(3):447-50.
7. Holgersson G, Sandelin M, Hoye E, et al. Swedish lung cancer radiation study group: the prognostic value of anaemia, thrombocytosis and leukocytosis at time of diagnosis in patients with non-small cell lung cancer. Med Oncol 2012;**29**(5):3176-82.
8. Ishizuka M, Nagata H, Takagi K, et al. Preoperative thrombocytosis is associated with survival after surgery for colorectal cancer. Journal of surgical oncology 2012;**106**(7):887-91.
9. Li AJ, Madden AC, Cass I, et al. The prognostic significance of thrombocytosis in epithelial ovarian carcinoma. Gynecologic oncology 2004;**92**(1):211-4.
10. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015;**349**:g7647.

Appendix B

Data extraction form

This appendix contains an example of a completed data extraction form.

B. Data extraction form

Thrombocytosis systematic review.
Data extraction form V1, July 2014

Document reference:	Hamilton et al 2005		
Endnote record n°:		Data extracted by:	Sarah Bailey
Country:	UK	Date:	18-08-2014
Aim of study:			

Study characteristics	
Study type	Population based case-control study
Study population (n, sex, age)	Men and women aged ≥40. All cases of lung cancer in Exeter over a four year period (n=247) plus 1235 age, sex and practice matched controls.
Location	Exeter, Devon
Cancers recorded	Lung only
Platelets/TH investigated?	Yes

What was the recruitment strategy in this study? Did they include sample size calculations?

All people diagnosed with lung cancer in Exeter, Devon over a four year period were recruited to this study.

Up to five controls were recruited for each case using computerised random numbers. These were age, sex and practice matched.

Exclusion criteria were: GP record unattainable; no records in two years prior to index date; subject had previous lung cancer; lived outside of Exeter at the time of diagnosis. Ineligible controls were replaced with a randomly selected back up.

There were sample size calculations – see paper for full details.

How were cancer cases identified? Which cases were excluded?

How were platelets measured?

Cases were IDd from the cancer registry at the RD&E. In addition, computerised searches were carried out at 21 general practices in the city.

Histological records were used to confirm cases, and where there were none of these, they were only accepted if the diagnosis was made by a specialist based on strong clinical evidence.

Platelet counts were taken from GP records. Only 132/247 (53%) of cases had a platelet count. Only 396/1235 controls (32%) had platelet counts.

Study results.

132/247 (53%) of cases had a platelet count. 34 (26%) of these had thrombocytosis.

396/1235 (32%) controls had platelet counts taken. 19 (5%) had thrombocytosis.

Likelihood ratio 8.9 (5.2-15)

Odds ratio (in multivariate model)9.3 (3.4-26) p<0.001

Key strengths and weaknesses of the study.

All cases occurring within the study period (in the study area) were included. Cases were confirmed by histology in the most part.

Checks of GP level data supplemented cancer registry data searches.

Some cases will have been missed despite GP level searches and confirming with histology. Those missed are likely to have been the most serious cases.

This analysis only includes patients from Exeter and so not necessarily comparable to other regions of the country. Certain factors including socio-demographic spread and ethnic composition of the population are likely to have an influence on cancer and differ between Exeter and other geographical regions.

There are no data available concerning the smoking, drinking or other lifestyle factors for the patients.

Coding of features of cancer was performed by research assistants and although the same person coded all patients for a single practice, there is still a chance of differences between different coders, or errors in the coding process.

Appendix C

Cancer medcodes

This appendix details the relevant medcodes used in the investigation of cancers in this thesis:

- Table C.1 displays a key for the codes.
- Table C.2 contains a list of medcodes with a description.

Table C.1: Keys for codes and their cancer sites for used in the table of medcodes (Table C.2).

Code	Cancer site
1	Bladder
2	Breast
3	Cervix
4	Prostate
5	Kidney
6	Leukaemia
7	Lymphoma
8	Myeloma
9	Oesophagus
10	Pancreas
11	Stomach
12	Testis
13	Uterus
14	Brain
15	Colorectal
16	Lung
17	Ovary
18	Oral
19	Skin (non-melanoma)
20	Other

Table C.2: Description of medcodes with reference ID fields.

Cancer site	Medcode	Description	Read code
1	7187	Carcinoma in situ of bladder	B837.00
1	16926	Neoplasm of unspecified nature of bladder	BA04.00
1	9712	[M]Papillary transitional cell carcinoma	BB4A.00
1	779	Malignant neoplasm of urinary bladder	B49..00
1	97091	[X]2ndry malignant neoplasm/bladder+oth+unsp urinary organs	ByuC500
1	35963	Malignant neoplasm of lateral wall of urinary bladder	B492.00
1	42023	Malignant neoplasm of urachus	B497.00
1	31102	Malignant neoplasm of urinary bladder NOS	B49z.00
1	28241	Malignant neoplasm of ureteric orifice	B496.00
1	47801	Malignant neoplasm, overlapping lesion of bladder	B49y000
1	44996	Malignant neoplasm of dome of urinary bladder	B491.00
1	6436	[M]Transitional cell carcinoma NOS	BB43.00
1	42012	Malignant neoplasm of posterior wall of urinary bladder	B494.00
1	21652	[M]Transitional cell carcinoma in situ	BB42.00
1	38862	Malignant neoplasm of trigone of urinary bladder	B490.00
1	41571	Malignant neoplasm of bladder neck	B495.00
1	58798	[M]Transitional cell carcinoma, spindle cell type	BB47.00
1	36949	Malignant neoplasm of other site of urinary bladder	B49y.00
1	22146	Secondary malignant neoplasm of bladder	B581100
2	10387	Lobular carcinoma in situ of breast	B830000
2	16760	Secondary malignant neoplasm of breast	B58y000
2	18694	Intraductal carcinoma in situ of breast	B830100
2	8351	[M]Infiltrating duct carcinoma	BB91.00
2	30543	Malignant neoplasm of skin of breast	B335200
2	8647	Carcinoma in situ of skin of breast	B825000
2	19423	Malignant neoplasm of male breast	B35..00
2	23399	Malignant neoplasm of upper-outer quadrant of female breast	B344.00
2	23380	Malignant neoplasm of nipple of female breast	B340000
2	95057	Malignant neoplasm of ectopic site of female breast	B34y000
2	40359	[M]Juvenile breast carcinoma	BB94.00
2	3968	Malignant neoplasm of female breast	B34..00
2	54202	Malignant neoplasm of other site of male breast	B35z.00
2	37969	Malignant neoplasm of skin of chest, excluding breast	B335100
2	68480	Malignant neoplasm of nipple of male breast	B350000
2	9956	[M]Intraductal carcinoma and lobular carcinoma in situ	BB9E000
2	27728	[M]Intraductal carcinoma, noninfiltrating NOS	BB90.00
2	53803	[X]Other carcinoma in situ of breast	ByuFG00
2	31546	Malignant neoplasm of central part of female breast	B341.00
2	12499	[X]Malignant neoplasm of breast	Byu6.00

C. Cancer medcodes

2	42070	Malignant neoplasm of lower-outer quadrant of female breast	B345.00
2	42542	[M]Paget's disease and infiltrating breast duct carcinoma	BB9K.00
2	48809	Malignant neoplasm of male breast NOS	B35zz00
2	59831	Malignant neoplasm of nipple or areola of female breast NOS	B340z00
2	64686	Malignant neoplasm of areola of female breast	B340100
2	49148	Malignant neoplasm, overlapping lesion of breast	B347.00
2	12427	[M]Lobular carcinoma NOS	BB9F.00
2	29826	Malignant neoplasm of upper-inner quadrant of female breast	B342.00
2	12300	[M]Paget's disease, mammary	BB9J.00
2	20685	Malignant neoplasm of axillary tail of female breast	B346.00
2	45222	Malignant neoplasm of lower-inner quadrant of female breast	B343.00
2	7833	Carcinoma in situ of breast	B830.00
2	54494	Malignant neoplasm of nipple and areola of male breast	B350.00
2	45906	Neoplasm of unspecified nature of breast	BA03.00
2	9505	Secondary malignant neoplasm of skin of breast	B582600
2	26853	Malignant neoplasm of nipple and areola of female breast	B340.00
2	12480	[M]Paget's disease and intraductal carcinoma of breast	BB9K000
2	38475	Malignant neoplasm of other site of female breast NOS	B34yz00
2	56715	Malignant neoplasm of other site of female breast	B34y.00
2	348	Ca female breast	B34..11
2	67701	[M]Secretory breast carcinoma	BB94.11
2	39760	[M]Infiltrating duct and lobular carcinoma	BB91100
2	9470	Malignant neoplasm of female breast NOS	B34z.00
2	60803	[M]Paget's disease, breast	BB9J.11
2	67884	Malignant neoplasm of areola of male breast	B350100
2	95323	Malignant neoplasm of ectopic site of male breast	B35z000
3	50297	Malignant neoplasm of exocervix	B411.00
3	50285	Malignant neoplasm of endocervix NOS	B410z00
3	28311	Malignant neoplasm of cervix uteri NOS	B41z.00
3	44534	[M]Intraepit neop,grade III,of cervix, vulva and vagina	BB2N.00
3	58094	Malignant neoplasm, overlapping lesion of cervix uteri	B412.00
3	3230	Cervical carcinoma (uterus)	B41..11
3	97832	Secondary cancer of the cervix	B58y211
3	53103	Malignant neoplasm of endocervical gland	B410100
3	50126	Carcinoma in situ of exocervix	B831100
3	95505	Malignant neoplasm of cervical stump	B41y000
3	24228	Carcinoma in situ of endocervix	B831000
3	43435	Malignant neoplasm of other site of cervix NOS	B41yz00
3	73616	Secondary malignant neoplasm of cervix uteri	B58y200
3	4087	CIN III - carcinoma in situ of cervix	B831.11
3	48820	Malignant neoplasm of endocervix	B410.00

3	57235	Malignant neoplasm of endocervical canal	B410000
3	2747	Malignant neoplasm of cervix uteri	B41..00
3	32955	Malignant neoplasm of other site of cervix	B41y.00
3	3279	Carcinoma in situ of cervix uteri	B831.00
3	72695	[X]Carcinoma in situ of other parts of cervix	ByuFA00
3	57719	Malignant neoplasm of squamocolumnar junction of cervix	B41y100
4	6328	Carcinoma in situ of prostate	B834.00
4	21590	Secondary malignant neoplasm of prostate	B58y500
4	780	Malignant neoplasm of prostate	B46..00
5	1952	Secondary malignant neoplasm of kidney	B580.00
5	13559	Malig neop of kidney and other unspecified urinary organs	B4A..00
5	15419	[M]Hypernephroma	BB5a012
5	17314	[M]Wilms' tumour	BBL7112
5	54594	[M]Mesoblastic nephroma	BBL7000
5	54184	Malignant neoplasm of renal pelvis NOS	B4A1z00
5	21681	[M]Nephroblastoma NOS	BBL7100
5	10668	[M]Renal cell carcinoma	BB5a000
5	18712	Renal malignant neoplasm	B4A..11
5	18771	[M]Clear cell sarcoma of kidney	BBLJ.00
5	1599	Malignant neoplasm of kidney parenchyma	B4A0.00
5	27697	[M]Hypernephroid tumour	BB5Y.00
5	12389	Malignant neoplasm of renal pelvis	B4A1.00
5	11178	Warthin's tumour	B702300
5	65159	Malignant neoplasm of perinephric tissue	B180100
5	52266	[M]Grawitz tumour	BB5a011
5	27540	Malignant neoplasm of renal calyces	B4A1000
5	7978	Hypernephroma	B4A0000
5	29462	Malignant neoplasm of kidney or urinary organs NOS	B4Az.00
5	19162	Malignant neoplasm of anterior wall of urinary bladder	B493.00
6	12146	[M]Lymphoid leukaemia NOS	BBr2000
6	57316	[M]Acute promyelocytic leukaemia	BBr6600
6	67700	Monoblastic leukaemia	B66..12
6	16416	Chronic leukaemia NOS	B681.00
6	31701	Chronic granulocytic leukaemia	B651.11
6	27664	Acute promyelocytic leukaemia	B65y100
6	66694	[M]Naegeli-type monocytic leukaemia	BBr6311
6	35875	Monocytic leukaemia	B66..00
6	20635	[M]Lymphatic leukaemia	BBr2011
6	99015	Other monocytic leukaemia	B66y.00
6	41734	[M]Leukaemia NOS	BBr0000
6	7176	Myeloid leukaemia	B65..00
6	41500	[M]Chronic lymphoid leukaemia	BBr2300
6	38914	Lymphoid leukaemia NOS	B64z.00

6	48049	[M]Chronic myelomonocytic leukaemia	BBr6800
6	67029	[X]Other lymphoid leukaemia	ByuD500
6	93342	Monocytic leukaemia NOS	B66z.00
6	70935	[M]Erythroleukaemia	BBr4000
6	37461	Adult T-cell leukaemia	B64y200
6	5137	Leukaemic reticuloendotheliosis	B624.11
6	4072	Acute leukaemia NOS	B680.00
6	54585	[M]Acute myeloid leukaemia	BBr6100
6	27330	Leukaemic reticuloendotheliosis	B624.00
6	63570	[M]Stem cell leukaemia	BBr0113
6	94174	Other and unspecified leukaemia	B67y.00
6	19974	Acute monocytic leukaemia	B660.00
6	61500	Acute myelomonocytic leukaemia	B690.00
6	27790	Chronic lymphatic leukaemia	B641.11
6	50928	[M]Burkitt's cell leukaemia	BBr2600
6	38331	Other lymphoid leukaemia NOS	B64yz00
6	71377	[M]Eosinophilic leukaemia	BBr8000
6	54793	Subacute leukaemia NOS	B682.00
6	6316	[M]Acute leukaemia NOS	BBr0100
6	64963	[M]Blastic leukaemia	BBr0112
6	39187	Plasma cell leukaemia	B631.00
6	19372	Lymphoid leukaemia	B64..00
6	8625	Chronic lymphoid leukaemia	B641.00
6	33344	Myeloid leukaemia NOS	B65z.00
6	27458	Chronic monocytic leukaemia	B661.00
6	66089	Other myeloid leukaemia NOS	B65yz00
6	46048	[M]Prolymphocytic leukaemia	BBr2500
6	49725	Other lymphoid leukaemia	B64y.00
6	37723	[M]Granulocytic leukaemia NOS	BBr6011
6	73777	Leukaemic reticuloendotheliosis NOS	B624z00
6	37272	Other specified leukaemia	B67..00
6	34692	Other leukaemia of unspecified cell type	B68y.00
6	25191	Leukaemia of unspecified cell type	B68..00
6	52942	[M]Chronic myeloid leukaemia	BBr6300
6	87335	Hairy cell leukaemia	B624.12
6	30632	Other specified leukaemia NOS	B67z.00
6	71850	[M]Myeloid leukaemia NOS	BBr6000
6	59929	[M]Leukaemia unspecified, NOS	BBr0z00
6	89762	[X]Other monocytic leukaemia	ByuD700
6	57671	Megakaryocytic leukaemia	B672.00
6	31586	Prolymphocytic leukaemia	B64y100
6	72197	Lymphosarcoma cell leukaemia	B67y000
6	4251	Acute lymphoid leukaemia	B640.00
6	72774	Subacute lymphoid leukaemia	B642.00
6	46263	[M]Acute myelomonocytic leukaemia	BBr6700

6	65777	Thrombocytic leukaemia	B672.11
6	65123	Leukaemic reticuloend of intra-abdominal lymph nodes	B624300
6	4222	Lymphatic leukaemia	B64..11
6	37410	[M]Acute lymphoid leukaemia	BBr2100
6	61693	[X]Other myeloid leukaemia	ByuD600
6	4413	Acute myeloid leukaemia	B650.00
6	73088	[M]Monocytic leukaemia NOS	BBr9000
6	42297	[M]Leukaemia NOS	BBrz.00
6	72310	[M]Aleukaemic leukaemia NOS	BBr0400
6	65721	Mast cell leukaemia	B673.00
6	4250	Leukaemia NOS	B68z.00
6	27520	Chronic myeloid leukaemia NOS	B651z00
6	69299	[M]Thrombocytic leukaemia	BBrA111
6	22071	[M]Blast cell leukaemia	BBr0111
6	65122	Leukaemic reticuloendotheliosis of unspecified sites	B624000
6	22050	Chronic myelomonocytic leukaemia	B691.00
6	89329	[X]Other specified leukaemias	ByuD800
6	5915	[M]Hairy cell leukaemia	BBrA400
6	65165	[X]Other leukaemia of unspecified cell type	ByuD900
6	10726	Chronic myeloid leukaemia	B651.00
6	72179	[M]Subacute leukaemia NOS	BBr0200
6	63475	Subacute myeloid leukaemia	B652.00
6	72222	[M]Megakaryocytic leukaemia	BBrA100
6	49327	[M]Acute megakaryoblastic leukaemia	BBrA500
6	42539	Acute erythraemia and erythroleukaemia	B670.00
6	31750	[M]Chronic leukaemia NOS	BBr0300
7	61251	[M]Malign lymphoma,lymphocytic,intermediate differn, diffuse	BBgN.00
7	95012	Mycosis fungoides of lymph nodes of multiple sites	B621800
7	99012	Hodgkin's disease NOS of lymph nodes inguinal region and leg	B61z500
7	21549	Follicular non-Hodgkin's lymphoma	B627C00
7	71304	Burkitt's lymphoma NOS	B602z00
7	43415	[X]Other Hodgkin's disease	ByuD000
7	12464	Peripheral T-cell lymphoma	B62x200
7	71619	[M]Malignant lymphoma, large cell, noncleaved, diffuse	BBgT.00
7	46967	[M]Mycosis fungoides	BBl..00
7	58082	Nodular lymphoma of lymph nodes of multiple sites	B620800
7	63054	Hodgkin's disease, nodular sclerosis NOS	B614z00
7	46877	[M]Malignant lymphoma, small lymphocytic NOS	BBgL.00
7	66367	HIV dis resulting oth types of non-Hodgkin's lymphoma	A789700
7	56041	[M]Hodgkin's disease, lymphocytic predominance	BBj1.00
7	69301	[M]Malignant lymphoma, convoluted cell type NOS	BBg5.00
7	92380	Burkitt's lymphoma of lymph nodes of inguinal region and leg	B602500

7	19140	Hodgkin's nodular sclerosis of lymph nodes of multiple sites	B614800
7	64336	[X]Other specified types of non-Hodgkin's lymphoma	ByuD300
7	95049	Hodgkin's lymphocytic depletion of unspecified site	B616000
7	17182	Follicular lymphoma NOS	B627C11
7	68039	Hodgkin's sarcoma of lymph nodes of axilla and upper limb	B612400
7	34089	Malignant lymphoma NOS of lymph nodes of axilla and arm	B62y400
7	91900	Hodgkin's disease NOS of lymph nodes of axilla and arm	B61z400
7	72725	Malignant lymphoma NOS of intrathoracic lymph nodes	B62y200
7	59755	Hodgkin's disease NOS of intrathoracic lymph nodes	B61z200
7	69980	[M]Malignant lymphoma, lymphocytic, well differentiated NOS	BBgC.00
7	15504	Malignant lymphoma NOS of lymph nodes of multiple sites	B62y800
7	64515	[X]Diffuse non-Hodgkin's lymphoma, unspecified	ByuDC00
7	63699	[M]Malignant lymphoma, nodular NOS	BBk0.00
7	67703	Hodgkin's disease, lymphocytic depletion	B616.00
7	91674	Mycosis fungoides of intra-abdominal lymph nodes	B621300
7	92068	Nodular lymphoma of intra-abdominal lymph nodes	B620300
7	44617	HIV disease resulting in Burkitt's lymphoma	A789600
7	66327	Nodular lymphoma of unspecified site	B620000
7	16774	[M] Cutaneous lymphoma	BBmD.00
7	94005	Hodgkin's disease, mixed cellularity NOS	B615z00
7	17460	Diffuse non-Hodgkin's lymphoblastic (diffuse) lymphoma	B627700
7	65584	[M]Hodgkin,s disease, lymphocytic predominance, diffuse	BBj1000
7	55303	Hodgkin's nodular sclerosis of head, face and neck	B614100
7	59115	Burkitt's lymphoma of lymph nodes of head, face and neck	B602100
7	8649	[X]Non-Hodgkin's lymphoma, unspecified type	ByuDF00
7	72714	Mycosis fungoides of lymph nodes of inguinal region and leg	B621500
7	39906	[M]Malignant lymphoma, centrocytic	BBgE.00
7	58015	[M]Malignant lymphomatous polyposis	BBgQ.00
7	60092	Malignant lymphoma NOS of spleen	B62y700
7	42198	[M]Hodgkin's disease, nodular sclerosis NOS	BBj6.00
7	46931	[M]Malignant lymphoma, stem cell type	BBg4.00
7	39798	Diffuse non-Hodgkin's lymphoma, unspecified	B627X00
7	50696	Malignant lymphoma NOS of lymph nodes of head, face and neck	B62y100
7	89230	[M]Hodgkin's granuloma	BBj9.00
7	65489	Hodgkin's paragranuloma	B610.00
7	64036	Hodgkin's sarcoma	B612.00

7	60275	[M]Malignant lymphoma, centroblastic type NOS	BBgJ.00
7	98961	[M]Malignant lymphoma, centroblastic-centrocytic, follicular	BBk2.00
7	51285	[M]Hodgkin's disease, mixed cellularity	BBj2.00
7	97577	Burkitt's lymphoma of intra-abdominal lymph nodes	B602300
7	21463	[M]Lymphocytic lymphoma NOS	BBgC.11
7	28639	Follicular non-Hodgkin's small cleaved cell lymphoma	B627000
7	49605	Hodgkin's disease, mixed cellularity	B615.00
7	95630	True histiocytic lymphoma	B62x600
7	53397	Hodgkin's disease NOS	B61z.00
7	40508	[M]Hodgkin,s disease, nodular sclerosis, lymphocytic predom	BBj6000
7	41754	[M]Malignant lymphoma, lymphoplasmacytoid type	BBg7.00
7	44318	Oth and unspecif peripheral & cutaneous T-cell lymphomas	B62xX00
7	53551	Diffuse non-Hodgkin's immunoblastic (diffuse) lymphoma	B627600
7	48253	[M]Malignant lymphoma, immunoblastic type	BBg8.00
7	12335	Malignant lymphoma NOS	B62y.00
7	65180	Diffuse non-Hodgkin's lymphoma undifferentiated (diffuse)	B627800
7	69767	[X]HIV disease resulting in other non-Hodgkin's lymphoma	AyuC600
7	35014	Sezary's disease	B622.00
7	7940	[X]Non-Hodgkin's lymphoma NOS	ByuDF11
7	31726	[M]Malignant lymphoma, small cleaved cell, diffuse	BBgM.00
7	93951	Hodgkin's, lymphocytic-histiocytic pred inguinal and leg	B613500
7	94279	Hodgkin's disease NOS of spleen	B61z700
7	3371	[M]Non Hodgkins lymphoma	BBg2.11
7	68964	[M]Malignant lymphoma, centroblastic-centrocytic, diffuse	BBgA.00
7	98596	[X]Other types of diffuse non-Hodgkin's lymphoma	ByuD200
7	57225	Hodgkin's disease, nodular sclerosis of unspecified site	B614000
7	96183	[M]Hodgkin's disease,lymphocytic depletion,diffuse fibrosis	BBj4.00
7	40766	[M] Peripheral T-cell lymphoma NOS	BBm5.00
7	45264	Nodular lymphoma of lymph nodes of head, face and neck	B620100
7	71262	Malignant lymphoma NOS of intrapelvic lymph nodes	B62y600
7	67518	[X]Other types of follicular non-Hodgkin's lymphoma	ByuD100
7	71142	Hodgkin's, lymphocytic-histiocytic predominance unspec site	B613000
7	63625	Hodgkin's lymphocytic depletion lymph nodes axilla and arm	B616400
7	42461	Hodgkin's disease NOS	B61zz00

7	31741	[M]Hodgkin,s disease, nodular sclerosis, lymphocytic deplet	BBj6200
7	71117	[M]Malignant lymphoma, undifferentiated cell type NOS	BBg3.00
7	92245	Hodgkin's, lymphocytic-histiocytic pred intrathoracic nodes	B613200
7	68330	Hodgkin's, lymphocytic-histiocytic pred of head, face, neck	B613100
7	15027	Malignant lymphoma NOS	B62yz00
7	29876	Hodgkin's, lymphocytic-histiocytic predominance NOS	B613z00
7	57427	Malignant lymphoma NOS of unspecified site	B62y000
7	23711	[M]Malignant lymphoma, diffuse NOS	BBg1000
7	31537	[M]Hodgkin,s disease, lymphocytic predominance, nodular	BBj1100
7	42769	[M]Hodgkin's disease NOS	BBjz.00
7	73532	Hodgkin's, lymphocytic-histiocytic pred intra-abdominal node	B613300
7	96379	Mycosis fungoides of lymph nodes of axilla and upper limb	B621400
7	12006	Mycosis fungoides	B621.00
7	42579	Malignant lymphoma NOS of intra-abdominal lymph nodes	B62y300
7	1483	[M]Lymphoma NOS	BBg1.11
7	16460	[M]Malignant lymphoma, non Hodgkin's type	BBg2.00
7	61662	Hodgkin's disease NOS, unspecified site	B61z000
7	57544	[M]True histiocytic lymphoma	BBm4.00
7	94995	Nodular lymphoma of lymph nodes of inguinal region and leg	B620500
7	71652	[M]Malignant lymphoma, mixed small and large cell, diffuse	BBgP.00
7	63994	[M]Malignant lymphoma, large cell, cleaved, diffuse	BBgS.00
7	51680	[M]Malignant lymphoma, small cell, noncleaved, diffuse	BBgV.00
7	57737	Lymphoepithelioid lymphoma	B62x100
7	61997	[M]Hodgkin's disease NOS	BBj0.00
7	38939	Hodgkin's disease, lymphocytic-histiocytic predominance	B613.00
7	97756	[M]Sezary's disease	BBl1.00
7	95338	Hodgkin's, lymphocytic-histiocytic pred intrapelvic nodes	B613600
7	99200	[M]Hodgkin's disease, nodular sclerosis, cellular phase	BBj7.00
7	44196	Hodgkin's granuloma	B611.00
7	94407	Hodgkin's mixed cellularity of lymph nodes head, face, neck	B615100
7	98840	Hodgkin's paragranuloma of intra-abdominal lymph nodes	B610300
7	31794	Unspecified B-cell non-Hodgkin's lymphoma	B627W00
7	38005	Mycosis fungoides NOS	B621z00

7	70509	Diffuse non-Hodgkin's centroblastic lymphoma	B627D00
7	64947	[M]Brill - Symmers' disease	BBk0.11
7	34352	[M]Lymphoblastic lymphoma NOS	BBgG.12
7	51852	[M]Malig lymphoma, lymphocytic, intermediate different NOS	BBgD.00
7	65483	Hodgkin's nodular sclerosis of lymph nodes of axilla and arm	B614400
7	50695	Diffuse non-Hodgkin mixed sml & lge cell (diffuse) lymphoma	B627500
7	3604	Non - Hodgkin's lymphoma	B627.00
7	49301	Malignant neoplasm lymphatic or haematopoietic tissue NOS	B6z..00
7	2462	Hodgkin's disease	B61..00
7	70842	Follicular non-Hodg mixed sml cleavd & lge cell lymphoma	B627100
7	18383	[M] Large cell lymphoma	BBmH.00
7	29335	[M]Adult T-cell leukaemia/lymphoma	BBr2700
7	31576	Other types of follicular non-Hodgkin's lymphoma	B627B00
7	33869	[M]Malignant lymphoma, large cell, diffuse NOS	BBgR.00
7	31749	[M]Monocytoid B-cell lymphoma	BBv0.00
7	49262	Follicular non-Hodgkin's large cell lymphoma	B627200
7	59778	Hodgkin's disease NOS of lymph nodes of head, face and neck	B61z100
7	63375	[X]Unspecified B-cell non-Hodgkin's lymphoma	ByuDE00
7	31492	[M] Monocytoid B-cell lymphoma	BBm9.00
7	97863	Hodgkin's disease, mixed cellularity of unspecified site	B615000
7	41841	[M]Malignant lymphoma, follicular centre cell NOS	BBgB.00
7	90201	T-zone lymphoma	B62x000
7	61149	Hodgkin's nodular sclerosis of intra-abdominal lymph nodes	B614300
7	72196	[M]Malignant lymphoma, lymphocytic, poorly different NOS	BBgG.00
7	64343	[M]Hodgkin,s disease, nodular sclerosis, mixed cellularity	BBj6100
7	97852	[M]Malignant lymphoma, centroblastic type, follicular	BBk7.00
7	95464	[M]Mycosis fungoides	BBl0.00
7	97746	Hodgkin's disease NOS of lymph nodes of multiple sites	B61z800
7	63105	Malignant lymphoma NOS of lymph node inguinal region and leg	B62y500
7	27965	[M]AngiocentricT-cell lymphoma	BBv2.00
7	49253	[M]Giant follicular lymphoma	BBk0.13
7	21402	Burkitt's lymphoma	B602.00
7	67506	Hodgkin's nodular sclerosis of intrathoracic lymph nodes	B614200
7	29178	Hodgkin's disease, nodular sclerosis	B614.00
7	5179	Nodular lymphoma (Brill - Symmers disease)	B620.00
7	98909	Hodgkin's granuloma of lymph nodes of head, face and neck	B611100

7	50668	Diffuse non-Hodgkin's small cell (diffuse) lymphoma	B627300
7	36114	[M]Malignant lymphoma NOS	BBg1.00
7	65701	Nodular lymphoma NOS	B620z00
7	95949	Mycosis fungoides of unspecified site	B621000
7	3710	[M]Adenolymphoma	BBB1.00
7	58684	Hodgkin's mixed cellularity of intrathoracic lymph nodes	B615200
8	73135	[M]Solitary myeloma	BBn2.12
8	63864	[M]Plasmacytoma NOS	BBn2.00
8	4944	Multiple myeloma	B630.00
8	26135	[M] Alpha heavy chain disease	BBm6.00
8	3672	[M]Myeloma NOS	BBn0.12
8	43450	Immunoproliferative neoplasm or myeloma NOS	B63z.00
8	52593	[M] Gamma heavy chain disease	BBmE.00
8	46042	Lambda light chain myeloma	B630300
8	31671	[M]Plasma cell myeloma	BBn0.00
8	22158	Malignant plasma cell neoplasm, extramedullary plasmacytoma	B630000
8	9172	[M]Waldenstrom's macroglobulinaemia	BBmK.00
8	38321	Plasmacytoma NOS	B936.12
8	19028	Solitary myeloma	B630100
8	18744	[M]Multiple myeloma	BBn0.11
8	49530	[M] T-gamma lymphoproliferative disease	BBmC.00
8	53647	[M]Myelomatosis	BBn0.13
8	39490	[M]Plasmacytic myeloma	BBn0.14
8	15211	Myelomatosis	B630.12
8	43312	Myeloma - solitary	B936.11
8	21329	Plasmacytoma NOS	B630200
9	94278	Malignant neoplasm of gastro-oesophageal junction	B110111
9	30700	Malignant neoplasm of oesophagus NOS	B10z.00
9	56077	Carcinoma in situ of lower 1/3 oesophagus	B801200
9	63470	Malignant neoplasm of abdominal oesophagus	B102.00
9	41362	Malignant neoplasm of thoracic oesophagus	B101.00
9	42416	Malignant neoplasm of lower third of oesophagus	B105.00
9	50789	Malignant neoplasm of upper third of oesophagus	B103.00
9	64274	Carcinoma in situ of middle 1/3 oesophagus	B801100
9	8244	Carcinoma in situ of oesophagus	B801.00
9	61695	Malignant neoplasm of cervical oesophagus	B100.00
9	53591	Malignant neoplasm of other specified part of oesophagus	B10y.00
9	4865	Oesophageal cancer	B10z.11
9	99155	Carcinoma in situ of upper 1/3 oesophagus	B801000
9	44228	Carcinoma in situ of oesophagus NOS	B801z00
9	67497	Malignant neoplasm, overlapping lesion of oesophagus	B106.00
10	16931	Carcinoma in situ of pancreas	B80z000
10	97875	Malignant neoplasm, overlapping lesion of pancreas	B175.00
10	63102	[M]Islet cell carcinoma	BB5B100

10	51656	[M]Mucinous cystadenocarcinoma NOS	BB81E00
10	39870	Malignant neoplasm of tail of pancreas	B172.00
10	55663	[M]Apudoma	BB5y300
10	40810	Malignant neoplasm of body of pancreas	B171.00
10	58022	[M]Glucagonoma NOS	BB5B400
10	9224	[M]Insulinoma NOS	BB5B200
10	95150	[M]Serous surface papillary carcinoma	BB81B00
10	8771	Malignant neoplasm of head of pancreas	B170.00
10	49629	[M]Gastrinoma, malignant	BB5C100
10	65051	[M]Papillary cystadenocarcinoma, NOS	BB81500
10	44930	[M]Papillary serous cystadenocarcinoma	BB81800
10	34388	Malignant neoplasm of pancreas NOS	B17z.00
10	48537	Malignant neoplasm of other specified sites of pancreas	B17y.00
10	96635	Malignant neoplasm of ectopic pancreatic tissue	B17y000
10	98825	[M]Mixed islet cell and exocrine adenocarcinoma	BB5B600
10	95783	Malignant neoplasm of specified site of pancreas NOS	B17yz00
10	35718	[M]Gastrinoma NOS	BB5C000
10	11469	[M]Nesidioblastoma	BB5B011
10	21792	Carcinoma in situ of ampulla of Vater	B808600
10	38442	[M]Serous cystadenocarcinoma, NOS	BB81200
10	35535	Malignant neoplasm of pancreatic duct	B173.00
10	55675	Endocrine tumour of pancreas	B717011
10	10949	Malignant neoplasm of ampulla of Vater	B162.00
10	54749	[M]Papillary mucinous cystadenocarcinoma	BB81H00
10	66876	[M]Pseudomucinous adenocarcinoma	BB81E11
10	32294	[M]Glucagonoma, malignant	BB5B500
10	95609	[M]Insulinoma, malignant	BB5B300
10	35795	Malignant neoplasm of Islets of Langerhans	B174.00
11	72947	Carcinoma in situ of fundus of stomach	B802100
11	41215	Malignant neoplasm of pyloric canal of stomach	B111100
11	21620	Malignant neoplasm of pylorus of stomach	B111.00
11	27440	[M]Linitis plastica	BB55.00
11	14800	Malignant neoplasm of stomach NOS	B11z.00
11	54171	Malignant neoplasm of middle third of oesophagus	B104.00
11	59092	Malignant neoplasm of pylorus of stomach NOS	B111z00
11	32362	Malignant neoplasm of fundus of stomach	B113.00
11	22894	Malignant neoplasm of cardio-oesophageal junction of stomach	B110100
11	37859	Malignant neoplasm of cardia of stomach NOS	B110z00
11	37774	Carcinoma in situ of stomach NOS	B802z00
11	55434	Malignant neoplasm of greater curve of stomach unspecified	B116.00
11	65312	Malignant neoplasm of anterior wall of stomach NEC	B11y000
11	51690	Malignant neoplasm, overlapping lesion of stomach	B117.00
11	63087	Carcinoma in situ of body of stomach	B802200

11	17258	Carcinoma in situ of cardia of stomach	B802000
11	19318	Malignant neoplasm of pyloric antrum of stomach	B112.00
11	65372	Malignant neoplasm of other specified site of stomach NOS	B11yz00
11	43572	Malignant neoplasm of body of stomach	B114.00
11	42193	Malignant neoplasm of lesser curve of stomach unspecified	B115.00
11	32022	Malignant neoplasm of cardia of stomach	B110.00
11	17093	Carcinoma in situ of stomach	B802.00
11	55019	Malignant neoplasm of other specified site of stomach	B11y.00
11	48237	Malignant neoplasm of prepylorus of stomach	B111000
11	96802	Malignant neoplasm of posterior wall of stomach NEC	B11y100
12	9476	Teratoma of descended testis	B471100
12	91509	Malignant neoplasm of descended testis NOS	B471z00
12	15148	Malignant neoplasm of testis	B47..00
12	35223	[M]Spermatocytic seminoma	BBQ1100
12	21786	Seminoma of descended testis	B471000
12	8177	Carcinoma in situ of testis	B836000
12	15989	Teratoma of testis	B47z.12
12	38510	Malignant neoplasm of testis NOS	B47z.00
12	47668	Malignant neoplasm of tunica vaginalis	B48y100
12	7740	Seminoma of undescended testis	B470200
12	57084	[M]Seminoma, anaplastic type	BBQ1000
12	19475	Malignant neoplasm of descended testis	B471.00
12	34145	Secondary malignant neoplasm of testis	B58y600
12	96429	Malignant neoplasm of undescended testis NOS	B470z00
12	36325	Teratoma of undescended testis	B470300
12	64602	Malignant neoplasm of undescended testis	B470.00
12	9859	[M]Seminoma NOS	BBQ1z00
12	2961	Seminoma of testis	B47z.11
12	21319	[M]Testicular stromal tumour	BBC0.13
13	7904	Carcinoma in situ of endometrium	B832000
13	49400	Malignant neoplasm of endometrium	B430211
13	59097	Malignant neoplasm of lower uterine segment	B431000
13	43940	Malignant neoplasm of isthmus of uterine body	B431.00
13	33617	Malignant neoplasm of body of uterus NOS	B43z.00
13	49828	Malignant neoplasm of fallopian tube	B441.00
13	45793	Malignant neoplasm of myometrium of corpus uteri	B430300
13	68155	Malignant neoplasm of fundus of corpus uteri	B430100
13	45490	Malignant neoplasm of corpus uteri NOS	B430z00
13	10588	[M]Leiomyosarcoma NOS	BBK0200
13	97996	Malignant neoplasm of other site of uterine adnexa	B44y.00
13	72723	Malignant neoplasm of cornu of corpus uteri	B430000
13	64596	[M]Myxoid leiomyosarcoma	BBK0700
13	55090	Secondary malignant neoplasm of uterus	B58y100

13	2890	Malignant neoplasm of endometrium of corpus uteri	B430200
13	3213	Malignant neoplasm of corpus uteri, excluding isthmus	B430.00
13	9447	[M]Endometrioid carcinoma	BB5j200
13	61803	Carcinoma in situ of body of uterus	B832.11
13	70729	Malignant neoplasm of isthmus of uterine body NOS	B431z00
13	7046	Malignant neoplasm of body of uterus	B43..00
13	65106	Malignant neoplasm of uterine adnexa NOS	B44z.00
13	2744	Malignant neoplasm of uterus, part unspecified	B40..00
13	64497	[X]Malignant neoplasm of uterine adnexa, unspecified	Byu7000
13	59499	Carcinoma in situ of fallopian tube	B833100
13	34030	[M]Endometrial stromal sarcoma	BBL0.00
13	31608	Malignant neoplasm of other site of uterine body	B43y.00
13	29898	Carcinoma in situ of other and unspecified parts of uterus	B832.00
13	16967	Malignant neoplasm of overlapping lesion of corpus uteri	B432.00
13	46153	Malignant neoplasm of parametrium	B443.00
13	56740	[M]Leiomyoblastoma	BBK0311
14	46789	Malignant neoplasm of choroid plexus	B515000
14	27744	[M]Oligodendroglioma NOS	BBbQ.00
14	61783	[M]Juvenile astrocytoma	BBbG.11
14	68479	[M]Neuroastrocytoma	BBc7.11
14	59170	Malignant neoplasm of corpus callosum	B51y000
14	52751	[M]Ependymoma, anaplastic type	BBb8.00
14	18617	Malignant neoplasm of brain	B51..00
14	39386	[M]Mixed glioma	BBb2.11
14	44089	Malignant neoplasm of brain stem	B517.00
14	33843	Secondary malignant neoplasm of brain and spinal cord	B583.00
14	8550	Malignant neoplasm of pituitary gland	B542000
14	9575	[M]Glioblastoma multiforme	BBbL.11
14	50151	[M]Pineoblastoma	BBa3.00
14	38551	[M]Gliomatosis cerebri	BBb1.00
14	93537	Malignant neoplasm of midbrain	B517200
14	41520	Malignant neoplasm of brain NOS	B51z.00
14	42460	Malignant neoplasm of pineal gland	B543.00
14	98800	[M]Piloid astrocytoma	BBbG.12
14	49132	Malignant neoplasm of medulla oblongata	B517100
14	96798	[M]Meningothelial sarcoma	BBd2.12
14	47848	[M]Meningioma NOS	BBdz.00
14	94267	[M]Subependymal glioma	BBb3.00
14	46490	[M]Angioblastic meningioma	BBd7.11
14	23083	[M]Glioblastoma NOS	BBbL.00
14	54133	Malignant neoplasm of cerebrum NOS	B510z00
14	52511	Malignant neoplasm of cerebral ventricles	B515.00
14	30273	[M]Pilocytic astrocytoma	BBbG.00
14	59718	Malig neop pituitary gland or craniopharyngeal duct NOS	B542z00

14	8328	[M]Astrocytoma, anaplastic type	BBbC.00
14	90487	[M]Subependymal astrocytoma NOS	BBb3.11
14	5199	Cerebral metastasis	B583200
14	50235	[M]Astroblastoma	BBbK.00
14	43114	[M]Myxopapillary ependymoma	BBbA.00
14	60347	[M]Leptomeningeal sarcoma	BBd2.11
14	31574	[M]Glioma, malignant	BBb0.00
14	15991	Malignant neoplasm of choroid	B506.00
14	48073	Malignant neoplasm of basal ganglia	B510000
14	66064	[M]Giant cell glioblastoma	BBbM.00
14	28919	Malignant neoplasm of cerebral meninges	B521.00
14	10851	Cerebral tumour - malignant	B51..11
14	46404	[M]Oligodendroblastoma	BBbS.00
14	70942	Malignant neoplasm of hypothalamus	B510400
14	5198	Secondary malignant neoplasm of brain	B583000
14	63925	[X]Malignant neoplasm of meninges, unspecified	ByuA200
14	91240	Malignant neoplasm of pons	B517300
14	47556	Malignant neoplasm of temporal lobe NOS	B512z00
14	28344	[M]Subependymal astrocytoma NOS	BBb3.12
14	49875	Malignant neoplasm of meninges, unspecified	B52X.00
14	63973	[M]Microglioma	BBm0.00
14	31767	[M]Medullomyoblastoma	BBbV.00
14	46769	[M]Ependymoblastoma	BBb8.11
14	71139	Malignant neoplasm of other parts of brain	B51y.00
14	67587	[M]Pleomorphic xanthoastrocytoma	BBbZ.00
14	45154	Malignant neoplasm of cerebellum	B516.00
14	42426	Malignant neoplasm of frontal lobe	B511.00
14	45909	Carcinoma in situ of pituitary gland	B8yy300
14	64557	Malignant neoplasm of cerebral peduncle	B517000
14	59823	Malignant neoplasm pituitary gland and craniopharyngeal duct	B542.00
14	61399	Malignant neoplasm of cerebral cortex	B510100
14	1044	Neoplasm of unspecified nature of brain	BA06.00
14	37473	[M]Cerebellar sarcoma NOS	BBbW.00
14	49168	[M]Subependymal giant cell astrocytoma	BBb4.00
14	65241	Malignant neoplasm, overlapping lesion of brain	B51y200
14	8547	[M]Astrocytoma NOS	BBbB.00
14	41695	[M]Primitive neuroectodermal tumour	BBba.00
14	47633	[X]Malig neopl, overlap lesion brain & other part of CNS	ByuA300
14	70104	Malignant neoplasm of cerebral meninges NOS	B521z00
14	65952	[M]Desmoplastic medulloblastoma	BBbU.00
14	7319	[M]Infiltrating ductular carcinoma	BB9G.00
14	98677	[M]Meningiomatosis NOS	BBd1.00
14	19226	Malignant neoplasm of parietal lobe	B513.00
14	62126	Malignant neoplasm of thalamus	B510500

14	67236	Malignant neoplasm of hippocampus	B512000
14	31629	[M]Ganglioglioma	BBc6.00
14	27748	[M]Astrocytic glioma	BBbB.11
14	38870	[M]Psammomatous meningioma	BBd5.00
14	27363	[M]Meningioma, malignant	BBd2.00
14	39088	Malignant neoplasm of occipital lobe	B514.00
14	95108	[M]Diffuse meningiomatosis	BBd1.11
14	68641	Malignant neoplasm of brain stem NOS	B517z00
14	34763	[M]Medulloblastoma NOS	BBbT.00
14	59375	Secondary malignant neoplasm of brain or spinal cord NOS	B583z00
14	15711	Malignant neoplasm cerebrum (excluding lobes and ventricles)	B510.00
14	34252	[M]Gliosarcoma	BBb0.12
14	8523	[M]Glioma NOS	BBb0.11
14	49186	[M]Oligodendroglioma, anaplastic type	BBbR.00
14	27846	[M]Fibrillary astrocytoma	BBbF.00
14	45531	[M]Gemistocytic astrocytoma	BBbE.00
14	68808	[M]Mixed glioma	BBb2.00
15	93478	Malignant neoplasm, overlapping lesion of colon	B138.00
15	27811	Carcinoma in situ of rectosigmoid junction	B804000
15	62909	Secondary malignant neoplasm of rectum	B575100
15	48231	Malignant neoplasm of other specified sites of colon	B13y.00
15	22163	Carcinoma of caecum	B134.11
15	17144	Carcinoma in situ of sigmoid colon	B803300
15	39080	Carcinoma in situ of hepatic flexure of colon	B803000
15	7219	Carcinoma of rectum	B141.11
15	10864	Malignant neoplasm of descending colon	B132.00
15	29975	Carcinoma in situ of rectum	B804100
15	11628	Cancer of bowel	B1z0.11
15	10946	Malignant neoplasm of ascending colon	B136.00
15	33561	Carcinoma in situ of colon NOS	B803z00
15	31893	Carcinoma in situ of ascending colon	B803600
15	50974	Malignant neoplasm rectum,rectosigmoid junction and anus NOS	B14z.00
15	36200	Secondary malig neop of large intestine or rectum NOS	B575z00
15	27855	Malignant neoplasm of rectosigmoid junction	B140.00
15	47667	Carcinoma in situ of descending colon	B803200
15	28163	Malignant neoplasm of colon NOS	B13z.00
15	44529	Secondary malignant neoplasm of large intestine and rectum	B575.00
15	18619	Malignant neoplasm of splenic flexure of colon	B137.00
15	22699	Carcinoma in situ of splenic flexure of colon	B803700
15	2815	Malignant neoplasm of sigmoid colon	B133.00
15	3811	Malignant neoplasm of caecum	B134.00

15	38883	Carcinoma in situ of rectum or rectosigmoid junction NOS	B804z00
15	5901	Rectal carcinoma	B141.12
15	9088	Malignant neoplasm of hepatic flexure of colon	B130.00
15	30165	Malignant neoplasm of mesorectum	B18y200
15	9118	Colonic cancer	B13z.11
15	6935	Malignant neoplasm of transverse colon	B131.00
15	55659	Malig neop other site rectum, rectosigmoid junction and anus	B14y.00
15	60477	Carcinoma in situ of rectum and rectosigmoid junction	B804.00
15	6903	Carcinoma in situ of colon	B803.00
15	28727	Secondary malignant neoplasm of colon	B575000
15	1800	Malignant neoplasm of rectum	B141.00
15	37125	Carcinoma in situ of transverse colon	B803100
15	16916	Carcinoma in situ of caecum	B803400
16	21770	[M]Mesothelioma, unspecified	BBPX.00
16	52373	Carcinoma in situ of lower lobe bronchus and lung	B812400
16	40595	[X]Malignant neoplasm of bronchus or lung, unspecified	Byu2000
16	52178	[M]Intravascular bronchial alveolar tumour	BBTL.00
16	41523	Malignant neoplasm of middle lobe bronchus	B223000
16	27509	[M]Mesothelioma, malignant	BBP1.00
16	86820	[M]Mesothelioma, biphasic type, malignant	BBP7.00
16	10358	Malignant neoplasm of upper lobe, bronchus or lung	B222.00
16	31700	Malignant neoplasm of upper lobe bronchus	B222000
16	9156	[M]Oat cell carcinoma	BB1K.00
16	31573	Malignant neoplasm of pleura	B23..00
16	38756	[M]Cystic mesothelioma	BBP9.00
16	31268	Malignant neoplasm of middle lobe, bronchus or lung	B223.00
16	62124	Secondary and unspec malig neop bronchopulmonary lymph nodes	B561800
16	69392	Secondary and unspec malig neop inferior tracheo-bronchial LN	B561700
16	67797	Secondary and unspec malig neop superfic tracheo-bronchial LN	B561600
16	9267	Carcinoma in situ of bronchus and lung	B812.00
16	29283	Malignant neoplasm of other site of respiratory tract	B2zy.00
16	20170	Pancoast's syndrome	B222.11
16	25886	Malignant neoplasm of upper lobe of lung	B222100
16	44169	Malignant neoplasm of upper lobe, bronchus or lung NOS	B222z00
16	39923	Malignant neoplasm of middle lobe of lung	B223100
16	17391	Malignant neoplasm of carina of bronchus	B221000
16	37579	Carcinoma in situ of upper lobe bronchus and lung	B812200
16	4137	Secondary malignant neoplasm of lung	B570.00
16	30526	[X]Mesothelioma, unspecified	Byu5100
16	64810	Malignant neoplasm of thorax NOS	B551z00

16	36371	Malignant neoplasm of overlapping lesion of bronchus & lung	B225.00
16	34015	[M]Bronchiolo-alveolar adenocarcinoma	BB5S200
16	21715	[X]Mesothelioma of lung	Byu5011
16	9600	Mesothelioma of pleura	B232.00
16	18678	Malignant neoplasm of lower lobe bronchus	B224000
16	2587	Lung cancer	B22z.11
16	31188	Malignant neoplasm of lower lobe, bronchus or lung	B224.00
16	25372	Carcinoma in situ of bronchus or lung NOS	B812z00
16	12870	Malignant neoplasm of main bronchus	B221.00
16	21698	Malignant neoplasm of main bronchus NOS	B221z00
16	3903	Malignant neoplasm of bronchus or lung NOS	B22z.00
16	47897	Carcinoma in situ of middle lobe bronchus and lung	B812300
16	35058	Carcinoma in situ of main bronchus	B812100
16	38961	Malignant neoplasm of other sites of bronchus or lung	B22y.00
16	48348	[M]Pulmonary blastoma	BBLM.00
16	16723	[M]Bronchiolar carcinoma	BB5S212
16	47734	[M]Epithelioid mesothelioma, malignant	BBP5.00
16	7484	Mesothelioma	B226.00
16	66646	Malignant neoplasm, overlap lesion of resp & intrathor orgs	B26..00
16	33444	Malignant neoplasm of hilus of lung	B221100
16	49159	Carcinoma in situ of carina of bronchus	B812000
16	47286	Malignant neoplasm of thorax	B551.00
16	42566	Malignant neoplasm of lower lobe, bronchus or lung NOS	B224z00
16	54134	Malignant neoplasm of middle lobe, bronchus or lung NOS	B223z00
16	13243	Malignant neoplasm of trachea, bronchus and lung	B22..00
16	12582	Malignant neoplasm of lower lobe of lung	B224100
17	19141	Malignant neoplasm of ovary and other uterine adnexa	B44..00
17	18065	[M]Sertoli-Leydig cell tumour	BBC7.00
17	21435	[M]Ovarian stromal tumour	BBC0.12
17	21173	[M]Mullerian mixed tumour	BBL5.00
17	71301	[M]Struma ovarii, malignant	BBQA100
17	17137	Carcinoma in situ of ovary	B833000
17	70383	[M]Brenner tumour, malignant	BBM0100
17	44793	[M]Brenner tumour NOS	BBM0z00
17	44615	Secondary malignant neoplasm of ovary	B586.00
17	53694	[M]Krukenberg tumour	BB85111
17	95373	[M]Leydig cell tumour, malignant	BBCC100
17	71490	[M]Brenner tumour, borderline malignancy	BBM0000
17	60530	[M]Thecoma NOS	BBC1000
17	73623	[M]Leydig cell tumour NOS	BBCCz00
17	40742	[M]Thecoma, luteinized	BBC1200
17	6751	[M]Granulosa cell tumour NOS	BBC3.00

C. Cancer medcodes

17	7805	Malignant neoplasm of ovary	B440.00
17	1986	Cancer of ovary	B440.11
17	29945	[M]Malignant teratoma, trophoblastic	BBR4.00
17	31609	[M]Granulosa cell tumour, malignant	BBC4.00
17	52306	[M]Struma ovarii NOS	BBQA000
17	59995	[M]Lipid cell tumour of ovary	BBCE.00
17	48957	[M]Granulosa cell-theca cell tumour	BBC5.00
18	91895	Malignant neoplasm of glossoepiglottic fold	B064100
18	36716	Malignant neoplasm of floor of mouth NOS	B04z.00
18	96869	Malignant neoplasm of posterior wall of nasopharynx NOS	B071z00
18	61510	Malignant neoplasm of palatoglossal arch	B062200
18	57248	Malignant neoplasm aryepiglottic fold, hypopharyngeal aspect	B082.00
18	55066	Malignant neoplasm of tonsillar pillar	B062.00
18	55833	[M]Odontogenic tumour NOS	BBZ1.00
18	45824	Secondary malignant neoplasm of tongue	B58y900
18	72443	[M]Odontogenic tumour, malignant	BBZ2.00
18	33388	Malignant neoplasm of adenoid	B071000
18	51926	Malignant neoplasm of faucial pillar	B062000
18	50296	Malignant neoplasm of upper lip, lipstick area	B000100
18	37590	Malignant neoplasm of hard palate	B052.00
18	66384	Malignant neoplasm of lower lip, external	B001000
18	27944	Carcinoma in situ of tongue	B800100
18	63957	Carcinoma in situ of skin of lip	B820.00
18	43781	Malignant neoplasm of dorsum of tongue NOS	B011z00
18	58550	[M]Adenoameloblastoma	BBZD.11
18	45189	[M]Squamous odontogenic tumour	BBZJ.00
18	59274	[M]Calcifying epithelial odontogenic tumour	BBZP.00
18	22441	Malignant neoplasm of subglottis	B212.00
18	57864	[M]Melanoameloblastoma	BBa4.11
18	71147	Malignant neoplasm of lower lip, inner aspect	B003.00
18	67504	Malignant neoplasm of lower lip, buccal aspect	B003000
18	98740	Malignant neoplasm of upper lip, vermilion border NOS	B000z00
18	99001	Malignant neoplasm of upper lip, frenulum	B002100
18	94441	Malignant neoplasm of lower lip, oral aspect	B003300
18	37724	Malignant neoplasm of retromolar area	B056.00
18	40557	Malignant neoplasm of tongue NOS	B01z.00
18	49360	Malignant neoplasm of lower gum	B031.00
18	97530	Malignant neoplasm of lower buccal sulcus	B051100
18	7697	Carcinoma in situ of vocal fold - glottis	B810800
18	36948	Carcinoma in situ of cricoid cartilage	B810100
18	98861	[M]Odontoameloblastoma	BBZH.00
18	43642	Malignant neoplasm of dorsal surface of tongue	B011.00
18	37505	Carcinoma in situ of oral cavity	B800.11

18	95429	Malignant neoplasm of posterior wall of nasopharynx	B071.00
18	38488	Malignant neoplasm of ventral tongue surface NOS	B013z00
18	55374	Malignant neoplasm of epiglottis NOS	B215.00
18	96783	Malignant neoplasm of commissure of lip	B005.00
18	69951	Malignant neoplasm of roof of mouth	B055100
18	44663	Carcinoma in situ of hypopharynx	B800900
18	43431	Malignant neoplasm of base of tongue	B010.00
18	98500	Malignant neoplasm of upper lip, mucosa	B002200
18	37516	Malignant neoplasm of uvula	B054.00
18	45408	Malignant neoplasm of anterior portion of floor of mouth	B040.00
18	45986	Malignant neoplasm of lateral portion of floor of mouth	B041.00
18	73962	Malignant neoplasm of upper lip, vermilion border	B000.00
18	69671	Malignant neoplasm of posterior third of tongue	B010.11
18	37916	Malignant neoplasm of other specified mouth parts	B05y.00
18	36104	Carcinoma in situ of nasopharynx	B800700
18	30966	Carcinoma in situ of palate	B800600
18	40292	Malignant neoplasm of soft palate	B053.00
18	37940	Malignant neoplasm of pharyngeal recess	B072000
18	55630	Malignant neoplasm of other specified site of nasopharynx	B07y.00
18	59004	Malignant neoplasm of lateral wall of nasopharynx	B072.00
18	91035	Malignant neoplasm of fixed part of tongue NOS	B010z00
18	67323	Malignant neoplasm of oropharynx, other specified sites	B06y.00
18	73439	Malignant neoplasm of anterior epiglottis NOS	B064z00
18	26134	Malignant neoplasm of epiglottis, free border	B064000
18	95480	Malignant neoplasm of lower lip, lipstick area	B001100
18	62840	Malignant neoplasm of ventral surface of tongue	B013.00
18	66422	Malignant neoplasm, overlapping lesion of nasopharynx	B074.00
18	53884	Malignant neoplasm tonsil NOS	B060z00
18	26448	Malignant neoplasm of faucial tonsil	B060000
18	68399	Malignant neoplasm of lip unspecified, mucosa	B004200
18	70819	Malignant neoplasm of palate unspecified	B055.00
18	46548	Malignant neoplasm of pharyngeal tonsil	B071100
18	34409	Malignant neoplasm of base of tongue dorsal surface	B010000
18	30402	Malignant neoplasm of buccal mucosa	B050.11
18	28559	Malignant neoplasm of palate NOS	B055z00
18	24801	Carcinoma in situ of floor of mouth	B800400
18	98483	[M]Odontogenic fibrosarcoma	BBZN.11
18	63979	Malignant neoplasm of frenulum linguae	B013100
18	37549	Kaposi's sarcoma of palate	B05z000
18	16297	Malignant neoplasm of pharynx unspecified	B0z0.00
18	318	Malignant neoplasm of glottis	B210.00
18	50419	Carcinoma in situ of oropharynx	B800800
18	24397	Malignant neoplasm of tonsillar fossa	B061.00
18	94251	Malignant neoplasm of lip, unspecified, lipstick area	B00z100

18	39430	Malignant neoplasm of lip, oral cavity and pharynx NOS	B0zz.00
18	56709	Malignant neoplasm of other sites of floor of mouth	B04y.00
18	95772	Malignant neoplasm of upper buccal sulcus	B051000
18	44139	Malignant neoplasm of anterior wall of nasopharynx	B073.00
18	93842	Malignant neoplasm of palatopharyngeal arch	B062300
18	40467	[M]Ameloblastoma NOS	BBZF.00
18	42129	Carcinoma in situ of pharynx	B800.12
18	31860	Carcinoma in situ of aryepiglottic fold	B810600
18	37096	Malignant neoplasm of tongue, junctional zone	B015.00
18	56355	Malignant neoplasm of lateral wall of oropharynx	B066.00
18	90124	Malignant neoplasm of posterior wall of oropharynx	B067.00
18	46741	[M]Ameloblastic odontosarcoma	BBZC.00
18	18882	Malignant neoplasm of overlapping lesion of lip	B006.00
18	64462	Malignant neoplasm of posterior pharynx	B083.00
18	91037	Malignant neoplasm of other specified site of oropharynx NOS	B06yz00
18	28665	Malignant neoplasm of nasopharynx NOS	B07z.00
18	47205	Malignant overlapping lesion of tongue	B017.00
18	17912	Malignant neoplasm, overlapping lesion of floor of mouth	B042.00
18	50288	Carcinoma in situ of salivary glands	B800200
18	47737	Carcinoma in situ of lip	B800000
18	88362	Malignant neoplasm of other specified hypopharyngeal site	B08y.00
18	53460	Carcinoma in situ of epiglottis	B810200
18	32024	Malignant neoplasm of upper gum	B030.00
18	93218	Malignant neoplasm of gum NOS	B03z.00
18	49758	Malignant neoplasm of other sites lip, oral cavity, pharynx	B0zy.00
18	95016	Malignant neoplasm of Waldeyer's ring	B0z1.00
18	96782	Malignant neoplasm of lower lip, inner aspect NOS	B003z00
18	66270	Malignant neoplasm of upper lip, external	B000000
18	36161	Malignant neoplasm of tongue, tip and lateral border	B012.00
18	89909	Malignant neoplasm of lower lip, mucosa	B003200
18	94390	Malignant neoplasm of roof of nasopharynx	B070.00
18	90610	Malignant neoplasm of upper lip, oral aspect	B002300
18	91843	Malignant neoplasm of lower lip, frenulum	B003100
18	57866	Carcinoma in situ of gums	B800300
18	46728	Malignant neoplasm of anterior epiglottis	B064.00
18	73614	Malignant neoplasm of lip unspecified, buccal aspect	B004000
18	39897	Malignant neoplasm of pyriform sinus	B081.00
18	31364	Malignant neoplasm of cheek mucosa	B050.00
18	37187	Carcinoma in situ of lip, oral cavity and pharynx NOS	B800z00
18	67446	Malignant neoplasm of lower lip, vermilion border	B001.00
18	28451	Malignant neoplasm of hypopharynx NOS	B08z.00
18	26165	Malignant neoplasm of supraglottis	B211.00

18	55015	Malignant neoplasm of mouth NOS	B05z.00
18	95390	Carcinoma in situ of lip, oral cavity and pharynx	B800.00
18	43548	Malignant neoplasm of postcricoid region	B080.00
18	48519	Malignant neoplasm of junctional region of epiglottis	B065.00
18	10375	Carcinoma in situ of glottis	B810811
18	61692	Malignant neoplasm of lip unspecified, inner aspect	B004.00
18	24852	Malignant neoplasm of lingual tonsil	B016.00
18	18245	Malignant neoplasm of skin of lip	B330.00
18	58121	Malignant neoplasm of anterior 2/3 of tongue unspecified	B014.00
18	69761	Malignant neoplasm of lip, vermilion border NOS	B00zz00
18	43200	Malignant neoplasm of oropharynx NOS	B06z.00
18	99185	Malignant neoplasm of glossopalatine fold	B062100
18	96003	Malignant neoplasm of junction of hard and soft palate	B055000
19	57446	Malignant neoplasm of skin of trunk, excluding scrotum	B335.00
19	51054	Carcinoma in situ of anal canal	B805.00
19	29282	[M]Basal cell tumour	BB30.00
19	12273	Carcinoma in situ of anus NOS	B806.00
19	69720	Carcinoma in situ of skin of eyebrow	B823100
19	49403	Malignant neoplasm of skin of chin	B333100
19	59100	[M]Mucoepidermoid tumour	BB70.00
19	3135	Carcinoma in situ of skin of nose	B823400
19	46469	Carcinoma in situ of skin of thigh	B827100
19	52328	Carcinoma in situ of skin of trunk, excluding scrotum	B825.00
19	50189	Carcinoma in situ skin of ear and external auricular canal	B822.00
19	47767	Malignant neoplasm of scrotum	B486.00
19	42212	Carcinoma in situ of skin of abdominal wall	B825400
19	67755	Carcinoma in situ of skin of foot	B827400
19	69601	Carcinoma in situ of skin of knee	B827200
19	19678	[M]Intraepithelial squamous cell carcinoma	BB29.13
19	54790	Carcinoma in situ of skin of lower arm	B826200
19	57358	Carcinoma in situ of skin of groin	B825500
19	27370	Malignant neoplasm skin of other and unspecified parts face	B333.00
19	65782	Malignant neoplasm of skin of toe	B337800
19	64270	Malignant neoplasm of skin of ankle	B337500
19	30576	Malignant neoplasm of skin of forehead	B333300
19	56954	Malignant neoplasm of skin of knee	B337200
19	38032	Carcinoma in situ of skin of back	B825300
19	68447	[M]Blue naevus, malignant	BBEV.00
19	61321	Carcinoma in situ of skin of buttock	B825700
19	12084	Carcinoma in situ of skin	B82..00
19	30853	[M]Basal cell neoplasm NOS	BB3z.00
19	67914	Malignant neoplasm of skin of great toe	B337900
19	43761	Malignant neoplasm of labia majora	B451.00
19	30747	Malignant neoplasm of skin of upper limb and shoulder	B336.00

19	53515	Malignant neoplasm skin of ear and external auricular canal	B332.00
19	37016	Malignant neoplasm of sebaceous gland	B33..14
19	19665	Carcinoma in situ of scalp	B824000
19	60162	[X]Malignant neoplasm overlapping lesion of skin	Byu5A00
19	21327	Malignant neoplasm of skin of temple	B333500
19	42429	Malignant neoplasm overlapping lesion of skin	B33X.00
19	62939	Carcinoma in situ of skin of chest wall NOS	B825100
19	29524	[M]Basal cell carcinoma, fibroepithelial type	BB34.00
19	27542	Carcinoma in situ of skin of lower leg	B827300
19	32249	Carcinoma in situ of ear	B822.11
19	708	Carcinoma in situ of skin of leg	B827.11
19	62305	Malignant neoplasm of skin of buttock	B335800
19	20539	Neoplasm of unspecified nature of skin	BA02200
19	9885	[M]Basal cell carcinoma, morphoea type	BB33.00
19	24370	Malignant neoplasm of anal canal	B142.00
19	49358	Carcinoma in situ of skin of hand	B826300
19	24551	[M]Melanocarcinoma	BBE1.11
19	36731	Malignant neoplasm of canthus	B331000
19	59614	Carcinoma in situ of skin of auricle	B822000
19	61103	Carcinoma in situ of skin of cheek	B823300
19	70918	[M]Mucoepidermoid neoplasm NOS	BB7z.00
19	56554	Carcinoma in situ of skin of shoulder	B826000
19	67912	[M]Papillary epidermoid carcinoma	BB26.11
19	8917	[M]Bowen's disease	BB2L.00
19	46568	Carcinoma in situ of skin of upper limb and shoulder	B826.00
19	47789	Carcinoma in situ of skin of forehead skin	B823000
19	56374	Carcinoma in situ of perianal skin	B825800
19	42707	Malignant neoplasm of skin of upper arm	B336100
19	25245	Malignant neoplasm of skin of finger	B336400
19	33271	Malignant neoplasm of pinna NEC	B332200
19	34823	Carcinoma in situ of cheek	B800500
19	57450	Carcinoma in situ of other specified sites of skin	B82y.00
19	2467	Bowen's disease	B8...11
19	43122	Malignant neoplasm of skin of shoulder	B336000
19	15868	Malignant neoplasm of skin of trunk, excluding scrotum, NOS	B335z00
19	55670	Malignant neoplasm of skin of eyebrow	B333200
19	49254	Carcinoma in situ of skin of other parts of face	B823.00
19	60563	Carcinoma in situ of skin of trunk NOS	B825z00
19	60526	Malignant neoplasm of skin of upper limb or shoulder NOS	B336z00
19	28625	[M]Mucoepidermoid carcinoma	BB71.00
19	61194	Malignant neoplasm of skin of lower limb or hip NOS	B337z00
19	66447	Malignant neoplasm of skin of scapular region	B335A00

19	33682	Malignant neoplasm of skin of lower leg	B337400
19	71655	Carcinoma in situ of skin of hip	B827000
19	48182	[M]Epidermoid carcinoma in situ	BB29.11
19	14815	Carcinoma in situ of skin of lower limb and hip	B827.00
19	70380	Malignant neoplasm of skin of axillary fold	B335000
19	23480	Malignant neoplasm of perianal skin	B335900
19	62399	Malig neop skin of ear and external auricular canal NOS	B332z00
19	68783	[M]Skin appendage carcinoma	BB60100
19	67748	Malignant neoplasm of skin of umbilicus	B335400
19	58601	Malignant neoplasm of skin of thigh	B337100
19	43087	Malignant neoplasm of eyelid including canthus	B331.00
19	68197	Malignant neoplasm of skin of popliteal fossa area	B337300
19	70295	Carcinoma in situ skin of ear/external auricular canal NOS	B822z00
19	46008	Malignant neoplasm skin other and unspec part of face NOS	B333z00
19	45077	Malignant neoplasm of skin of back	B335700
19	31511	Carcinoma in situ of skin of temple	B823500
19	30577	Malignant neoplasm of skin of fore-arm	B336200
19	43717	[M]Verrucous epidermoid carcinoma	BB24.11
19	56121	[X]Malignant neoplasm of skin, unspecified	Byu4300
19	90339	Carcinoma in situ of skin of upper limb or shoulder NOS	B826z00
19	41958	Malignant neoplasm of lower eyelid	B331200
19	64406	Malignant neoplasm of skin of thumb	B336500
19	57513	[M]Epidermoid carcinoma, keratinising type	BB2C.11
19	64630	Carcinoma in situ of skin of lower limb or hip NOS	B827z00
19	43930	Secondary malignant neoplasm of skin of head	B582000
19	4632	Other malignant neoplasm of skin	B33..00
19	37165	Malignant neoplasm of scalp	B334000
19	29787	[M]Squamous cell carcinoma, keratinising type NOS	BB2C.00
19	27897	Malignant neoplasm of anus unspecified	B143.00
19	70587	Malignant neoplasm of skin of foot	B337700
19	39390	Carcinoma in situ of skin of axilla	B825200
19	70988	Malignant neoplasm of skin of hip	B337000
19	57442	Malignant neoplasm of skin of lower limb and hip	B337.00
19	54352	Malignant neoplasm of skin of hand	B336300
19	57284	Carcinoma in situ of skin of upper arm	B826100
19	66319	Malignant neoplasm of skin of groin	B335500
19	18618	Malignant neoplasm of skin of abdominal wall	B335300
19	876	Basal cell carcinoma	B33..11
19	1940	Rodent ulcer	B33..13
19	62080	Malignant neoplasm of skin of external auditory meatus	B332100
19	68787	Malignant neoplasm of back NOS	B55y000
19	63142	Carcinoma in situ of skin NOS	B82z.00
19	16202	Malignant neoplasm of skin of nose (external)	B333400

19	2492	Malignant neoplasm of skin NOS	B33z.00
19	58879	Carcinoma in situ of scrotum	B836300
20	42856	Malignant neoplasm of nasal cavities NOS	B200z00
20	26846	Carcinoma in situ of ethmoidal sinus	B81y700
20	11403	Carcinoma in situ of larynx	B810.00
20	49463	Malignant neoplasm of tarsus of eyelid	B310400
20	64918	Secondary and unspec malig neop of superficial parotid LN	B560000
20	50475	Malignant neoplasm of major salivary gland NOS	B02z.00
20	26813	Malignant neoplasm of larynx, other specified site	B21y.00
20	73023	[M]Retinal angle tumour	BBa4.13
20	98911	Malignant neoplasm of nasal conchae	B200100
20	62182	Malignant neoplasm of vestibule of nose	B200300
20	64427	Unspec malig neop lymphoid/histiocytic lymph node head/neck	B62z100
20	95458	Malignant neoplasm of nasal bone	B300300
20	73537	Malig neop auditory tube, middle ear, mastoid air cells NOS	B201z00
20	97332	Malignant neoplasm of laryngeal cartilage NOS	B213z00
20	62761	Malignant neoplasm of septum of nose	B200200
20	50579	Malignant neoplasm, overlapping lesion of larynx	B214.00
20	98537	Malignant neoplasm of tympanic cavity	B201100
20	59036	Malignant neoplasm of bones of skull and face	B300.00
20	43380	Carcinoma in situ of maxillary sinus	B81y600
20	70928	Malignant neoplasm of sublingual gland	B022.00
20	58903	Malignant neoplasm of head, neck and face NOS	B550z00
20	73760	Malignant neoplasm of scalp or skin of neck NOS	B334z00
20	43475	Malig neop of connective and soft tissue head, face and neck	B310.00
20	32174	Malignant neoplasm of maxillary sinus	B202.00
20	27594	[M]Warthin's tumour	BBB1.11
20	50299	Malignant neoplasm of zygomatic bone	B300900
20	53599	Malignant neoplasm of frontal bone	B300100
20	39590	Malignant neoplasm, overlapping lesion of accessory sinuses	B206.00
20	70696	Malignant neoplasm of other major salivary glands	B02y.00
20	68135	Carcinoma in situ of mastoid air cells	B81y500
20	43619	Malignant neoplasm of skin of neck	B334100
20	54234	Malignant neoplasm of scalp and skin of neck	B334.00
20	12490	Malignant neoplasm of nose NOS	B550200
20	39717	Carcinoma in situ of nasal cavity	B81y100
20	59382	Malignant neoplasm of soft tissue of head	B310000
20	15684	Malignant neoplasm of frontal sinus	B204.00
20	9237	Malignant neoplasm of larynx NOS	B21z.00

20	73718	Malig neop connective and soft tissue head, face, neck NOS	B310z00
20	71238	Lymphosarcoma of lymph nodes of head, face and neck	B601100
20	23389	Malignant neoplasm of nasal cavities	B200.00
20	53594	Malignant neoplasm of ethmoid bone	B300000
20	4388	Malignant neoplasm of parotid gland	B020.00
20	51786	Malignant neoplasm of submandibular gland	B021.00
20	71946	Malignant neoplasm of mastoid air cells	B201300
20	71204	Malignant neoplasm of cartilage of nose	B200000
20	54636	Malignant neoplasm of ethmoid sinus	B203.00
20	49214	Secondary and unspec malig neop lymph nodes head/face/neck	B560.00
20	39064	Carcinoma in situ of sphenoidal sinus	B81y900
20	65215	Malignant neoplasm of sphenoidal sinus	B205.00
20	43111	Malignant neoplasm of laryngeal cartilage	B213.00
20	68236	Malignant neoplasm of head, neck and face	B550.00
20	54140	Carcinoma in situ of skin of neck	B824100
20	71584	Malignant neoplasm of lacrimal duct	B507.00
20	65222	Carcinoma in situ of skin of jaw	B823600
20	71031	Reticulosarcoma of lymph nodes of head, face and neck	B600100
20	17475	Malignant neoplasm of maxilla	B300A00
20	59426	Carcinoma in situ of nasal sinuses	B81y.11
20	33833	Malignant neoplasm of mandible	B301.00
20	48517	Malignant neoplasm of soft tissue of neck	B310200
20	67129	Secondary unspec malig neop lymph nodes head/face/neck NOS	B560z00
20	319	Malignant neoplasm of larynx	B21..00
20	55098	Malignant neoplasm of head NOS	B550000
20	16280	Malignant neoplasm of neck NOS	B550400
20	64817	Malignant neoplasm of lacrimal gland	B502.00
20	39084	Malignant neoplasm of laryngopharynx	B0z2.00
20	55550	Malignant neoplasm of upper eyelid	B331100
20	40014	Malignant neoplasm of soft tissue of face	B310100
20	69345	Carcinoma in situ of scalp and skin of neck	B824.00
20	53882	Carcinoma in situ of larynx NOS	B810z00
20	65357	Malignant neoplasm of nasolacrimal duct	B507100
20	35999	Secondary malignant neoplasm of skin of neck	B582200
20	17841	Malignant neoplasm of glans penis	B481.00
20	19321	Malignant neoplasm of connective and soft tissue of hand	B311300
20	57191	[X]Malignant neoplasm/other specified male genital organs	Byu8000
20	38777	Carcinoma in situ of skin of perineum	B825600
20	55101	Malignant neoplasm of pelvis NOS	B553z00
20	70126	Malignant neoplasm of optic nerve	B520100
20	62610	Carcinoma in situ of respiratory organ NOS	B81z.00

20	44915	Carcinoma in situ other and unspecified female genital organ	B833.00
20	56718	Malignant neoplasm of eyeball NOS	B500z00
20	48952	[M]Retinoblastoma NOS	BBc9z00
20	8627	[M]Tumour cells, malignant	BB07.00
20	86997	[X]Malignant neoplasm/ill-defined sites within resp system	Byu2400
20	39734	[M]Hilar cell tumour	BBCD.00
20	28059	Secondary and unspec malig neop of facial lymph nodes	B560600
20	37618	Malignant neoplasm of axilla NOS	B551000
20	98104	Malignant neoplasm of other specified pleura	B23y.00
20	62396	[M]Epithelioid cell sarcoma	BBF6.00
20	94614	[M]Arrhenoblastoma NOS	BBC6z11
20	6701	Secondary and unspec malig neop intrapelvic lymph nodes	B565.00
20	8930	[M]Adenocarcinoma NOS	BB52.00
20	59381	Malignant neoplasm of iris	B500100
20	52963	[M]Juvenile granulosa cell tumour	BBC3000
20	8693	Carcinoma of other and unspecified sites	B5...11
20	40240	[M]Hepatocellular carcinoma NOS	BB5D500
20	63659	[M]Juxtacortical chondrosarcoma	BBW6.00
20	63598	[X]Malignant neoplasms/independent (primary) multiple sites	ByuE.00
20	46792	Malignant neoplasm of temporal lobe	B512.00
20	64195	Malig neop of endocrine gland or related structure NOS	B54z.00
20	72803	Secondary and unspec malig neop intrapelvic LN NOS	B565z00
20	17874	Mesothelioma of peritoneum	B181.00
20	97954	Carcinoma in situ of other specified part respiratory system	B81y.00
20	54679	Secondary malignant neoplasm of unknown site	B594.00
20	5842	Secondary malignant neoplasm of other specified sites	B58..00
20	37919	Secondary and unspec malig neop internal mammary lymph nodes	B561000
20	30988	[M]Small cell carcinoma, intermediate cell	BB1M.00
20	39388	[M]Olfactory neuroblastoma	BBcC.11
20	70374	Reticulosarcoma of intra-abdominal lymph nodes	B600300
20	51115	Malignant neoplasm of spinal cord	B522.00
20	89258	Malignant neoplasm of peripheral nerve of low limb, incl hip	B524200
20	94975	Malignant neoplasm of pericardium	B241300
20	46771	[M]Hepatocellular carcinoma, fibrolamellar	BB5D800
20	35071	[M]Mixed germ cell tumour	BBQB.00
20	67203	[M]Lymphoblastic lymphosarcoma NOS	BBgG.11
20	65124	Malignant neoplasm of interlobular bile ducts	B151000
20	55946	Secondary malignant neoplasm of duodenum	B574000

20	67018	[M]Aortic body tumour	BBD5.00
20	58888	[M]Polycythaemia rubra vera	BBs0.11
20	27449	Malignant neoplasm of upper limb NOS	B554.00
20	46905	Malignant neoplasm of coccygeal body	B545200
20	68055	Malignant neoplasm of long bones of leg	B307.00
20	65047	[M]Phaeochromocytoma, malignant	BBDA.00
20	21217	[M]Small cell-large cell carcinoma	BB1N.00
20	96072	[M]Haemangiopericytic neoplasm NOS	BBTDz00
20	61643	Malignant neoplasm of intrahepatic bile ducts NOS	B151z00
20	6966	[M]Spindle cell carcinoma	BB1D.00
20	17098	[M]Pseudomyxoma peritonei	BB83.00
20	29160	Malignant neoplasm of connective and soft tissue of axilla	B313000
20	58692	Secondary and unspec malig neop paratracheal lymph nodes	B561500
20	17292	[M]Melanocytoma of eyeball	BBE8.11
20	62828	Secondary malignant neoplasm of other urinary organ NOS	B581z00
20	93384	Unspec malig neop lymphoid/histiocytic of intrathoracic node	B62z200
20	55468	[M]Mucocarcinoid tumour, malignant	BB5R600
20	46939	Malignant neoplasm of cervical vertebra	B302000
20	68611	Secondary and unspec malig neop deep cervical LN	B560900
20	58131	[M]Comedocarcinoma NOS	BB93.00
20	22156	[M]Malignant tumour, small cell type	BB08.00
20	21914	[M]Intraepithelial carcinoma NOS	BB11.11
20	45814	[M]G cell tumour NOS	BB5C011
20	35474	[M]Giant cell carcinoma	BB1C.00
20	69981	[M]Neurilemmoma, malignant	BBe7.00
20	67575	HIV disease resulting in unspecified malignant neoplasm	A788W00
20	72433	[M]Reticulosarcoma NOS	BBh0.00
20	95671	[X]Malignant neoplasm of peritoneum, unspecified	Byu5700
20	62492	[M]Osteoblastoma	BBV8.00
20	60045	[M]Tubular adenocarcinoma	BB5M100
20	51934	Carcinoma in situ of biliary system	B808.11
20	42082	[M]Alveolar rhabdomyosarcoma	BBK3700
20	65253	Secondary and unspec malignant neoplasm occipital lymph node	B560300
20	67339	[M]Malignant mastocytosis	BBp2.00
20	67970	[M]Small cell carcinoma, fusiform cell type	BB1L.00
20	57471	Malig neop of connective and soft tissue trunk unspecified	B316.00
20	67019	[M]Angiomyosarcoma	BBK1100
20	17212	[M]Rhabdoid sarcoma	BBLH.00
20	58953	[M]Malig lymp,follicular centre cell,noncleaved,follicular	BBk8.00
20	28628	[M]Liposarcoma, well differentiated type	BBJ3.00

20	28599	[M]Liposarcoma NOS	BBJ1.00
20	37553	Malignant neoplasm of lip, unspecified	B007.00
20	1969	[M]Haemangioblastoma	BBTF.00
20	70463	Malignant neoplasm of connective and soft tissue of buttock	B315000
20	69300	[M]Polygonal cell carcinoma	BB1F.00
20	51965	Malignant neoplasm of connective and soft tissue of pelvis	B315.00
20	64116	Secondary and unspec malig neop intrathoracic lymph nodes	B561.00
20	44435	[M]Phaeochromocytoma NOS	BBD9.00
20	45071	Malignant neoplasm of connective and soft tissue of abdomen	B314.00
20	66908	Malignant neoplasm of coccygeal vertebra	B306400
20	56640	Carcinoma in situ of other specified site NOS	B8yyz00
20	27827	[M]Adenocarcinoma in situ	BB51.00
20	58902	[M]Olfactory neurogenic tumour	BBcA.00
20	40740	[X]Malignant neoplasms of lymphoid, haematopoietic and rela	ByuD.00
20	45364	[M]Giant cell tumour of soft parts NOS	BBX2.00
20	60134	Secondary malignant neoplasm of ureter	B581000
20	72522	Malignant neoplasm of great vessels	B313200
20	60335	Secondary malignant neoplasm of vulva	B58y400
20	72713	Secondary and unspec malig neop superficial mesenteric LN	B562100
20	40592	[X]Malignant neoplasm of mesothelial and soft tissue	Byu5.00
20	64848	Malignant neoplasm of ulna	B304400
20	95421	Malignant neoplasm of other specified female genital organ	B45y.00
20	59362	Malignant neoplasm of labia majora NOS	B451z00
20	61542	[M]Malignant teratoma, undifferentiated type	BBQ7400
20	10134	[M]Squamous cell carcinoma in situ NOS	BB29.00
20	1056	Malignant neoplasm of other and unspecified site NOS	B5z..00
20	35348	[M]Papillary adenocarcinoma NOS	BB5T100
20	96231	[M]Fibroxanthoma, malignant	BBGJ.00
20	36876	[M]Eosinophil carcinoma	BB5V311
20	50035	Malignant neoplasm of aortic body and other paraganglia	B545.00
20	52029	[X]Malignant neoplasm without specification of site	ByuC800
20	50298	Malignant neoplasm of orbital bone	B300500
20	53989	Malig neop connective and soft tissue upper limb/shoulder	B311.00
20	99096	[X]Malignant neopl/overlapping les/resp+intrathoracic organs	Byu2300
20	34228	Neoplasm of unspecified nature NOS	BAz..00
20	43614	Malignant neoplasm/bones+articular cartilage/limb,unspfd	B30X.00

20	59919	[M]Multicentric basal cell carcinoma	BB32.00
20	66088	Malig neop of connective and soft tissue of hip and leg	B312.00
20	69443	[M]Gigantiform cementoma	BBZ7.00
20	27715	Malignant neoplasm of anterior mediastinum	B242.00
20	31772	[M]Dermatofibrosarcoma NOS	BBGM.00
20	63286	[M]Clear cell sarcoma of tendons and aponeuroses	BBN5.00
20	55096	Secondary malignant neoplasm of skin NOS	B582z00
20	21741	[M]Follicular adenocarcinoma NOS	BB5f100
20	46594	Carcinoma in situ of gall bladder	B808300
20	36870	[M]Adenosarcoma	BBL7111
20	44074	[M]Mucin-producing adenocarcinoma	BB84.00
20	31421	[M]Rhabdomyosarcoma NOS	BBK3100
20	27971	[M]Germinoma	BBQ2.00
20	40814	Malignant neoplasm of tibia	B307200
20	41515	[X]Malignant neoplasm/central nervous system, unspecified	ByuA100
20	38938	Malignant neoplasm of pelvis, sacrum or coccyx NOS	B306z00
20	68757	[M]Nonencapsulated sclerosing carcinoma	BB5f700
20	99240	Reticulosarcoma NOS	B600z00
20	13569	Disseminated malignancy NOS	B590.00
20	12497	[M]Mucinous adenocarcinoma	BB82100
20	51921	Malignant neoplasm of pubis	B306200
20	39883	[M]Malig lymp, follicular centre cell, cleaved, follicular	BBk5.00
20	66444	[X]Malignant neoplasm/overlap lesion/heart,mediastinm+pleura	Byu2100
20	35034	[M]Fibroxanthosarcoma	BBGJ.11
20	16704	Malignant neoplasm of vertebral column	B302.00
20	36530	[M]Alveolar cell carcinoma	BB5S211
20	16677	[M]Medullary carcinoma NOS	BB9B.00
20	49811	[M]Mesodermal mixed tumour	BBL6.00
20	95792	Lymphoid and histiocytic malignancy NOS	B62zz00
20	47330	[M]Histiocytic medullary reticulosis	BBm2.00
20	64516	Malignant neoplasm of parietal peritoneum	B18y400
20	58061	Malignant neoplasm of labia minora	B452.00
20	39531	Malig neo, overlapping lesion of heart, mediastinum & pleura	B25..00
20	27439	[M]Kaposi's sarcoma	BBTA.00
20	92720	Malignant neoplasm of posterior mediastinum	B243.00
20	35186	[X]Malignant neoplasm of ill-defined, secondary and unspeci	ByuC.00
20	97463	[M]Giant cell sarcoma (except of bone)	BBF4.00
20	61588	[M]Signet ring cell carcinoma	BB85000
20	41349	[M]Angioblastoma	BBTF.11
20	59284	[M]Mucous adenocarcinoma	BB82114

20	63896	Secondary malignant neoplasm of skin of shoulder and arm	B582400
20	31004	[M]Adenoid squamous cell carcinoma	BB2G.00
20	24539	[M]Chondroblastic osteosarcoma	BBV2.00
20	61716	Malignant neoplasm of peripheral nerve,upp limb,incl should	B524100
20	67107	Malignant neoplasm of parietal pleura	B230.00
20	70736	Secondary malignant neoplasm of vagina	B58y300
20	55429	[M]Mucoid adenocarcionoma	BB82113
20	73530	Malignant neoplasm of hand bones	B305.00
20	60035	Malignant neoplasm of cartilage of ear	B310300
20	28272	[M]Adenocarcinoma, intestinal type	BB57.00
20	69927	Malignant neoplasm of first metatarsal bone	B308800
20	12388	[M]Urothelial carcinoma	BB43.11
20	59240	[M]Carcinoma, diffuse type	BB58.00
20	68524	[M]Chondroblastoma NOS	BBW7.00
20	93762	Malignant neoplasm of placenta	B42..00
20	4473	[M]Ewing's sarcoma	BBY0.00
20	38477	[M]Giant cell tumour of bone NOS	BBX0.00
20	29385	[M]Osteoclastoma	BBX0.11
20	27617	Malignant neoplasm of overlapping lesion of vulva	B45y000
20	9622	Malignant neoplasm of cauda equina	B525.00
20	46761	[M]Papillary and follicular adenocarcinoma	BB5f600
20	46423	Cystosarcoma phyllodes	B933.11
20	49023	[M]Endothelial bone sarcoma	BBY0.11
20	67430	[M]Tibial adamantinoma	BBY1.11
20	49491	Malignant neoplasm of sternum	B303100
20	40622	[M]Mucoid cell carcinoma	BB5V711
20	93175	[M]Intraosseous carcinoma	BBZ2.11
20	49701	Malignant neoplasm of vertebral column NOS	B302z00
20	47840	Malignant neoplasm of aortic body	B545100
20	48828	Secondary malignant neoplasm of skin of hip and leg	B582500
20	63460	Malignant neoplasm of arytenoid cartilage	B213000
20	67217	Malignant neoplasm of trunk NOS	B55y100
20	15221	Malignant neoplasm of trachea	B220.00
20	29337	[M] Small cell osteosarcoma	BBVA.00
20	45667	Malignant neoplasm of orbit	B501.00
20	67934	[M]Carcinosarcoma, embryonal type	BBLA.00
20	66639	Malignant neoplasm of clavicle	B303200
20	97200	Carcinoma in situ of Eustachian tube	B81y400
20	16105	Malignant neoplasm of gallbladder	B160.00
20	44452	Malignant neoplasm of vomer	B300C00
20	9366	[M]Secondary carcinoma	BB13.11
20	69497	Malignant histiocytosis of unspecified site	B623000

20	50904	Secondary and unspec malig neop infraclavicular lymph nodes	B563200
20	98559	[M]Chondroblastoma, malignant	BBW8.00
20	59520	Malignant neoplasm of malar bone	B300200
20	40443	Malignant neoplasm of sweat gland	B33..15
20	7982	Malignant neoplasm of common bile duct	B161200
20	64700	Carcinoma in situ of spleen	B80z100
20	4852	[M]Verrucous squamous cell carcinoma	BB24.12
20	21447	[M]Fibroblastic osteosarcoma	BBV3.00
20	4555	Malig neop of other and unspecified female genital organs	B45..00
20	8711	[M]Cholangiocarcinoma	BB5D100
20	2481	Polycythaemia vera	B934.00
20	94355	Malignant neoplasm of flank NOS	B55y200
20	36401	Secondary malignant neoplasm of adrenal gland	B587.00
20	6471	Metastases of respiratory and/or digestive systems	B57..11
20	64670	Lymphosarcoma of intra-abdominal lymph nodes	B601300
20	59918	[M]Follicular adenocarcinoma, well differentiated type	BB5f200
20	56513	Malignant neoplasm of femur	B307000
20	8695	[M]Carcinoma NOS	BB12.00
20	8088	[M]Fibromyxosarcoma	BBG3.00
20	35772	Carcinoma in situ of thyroid cartilage	B810000
20	51255	Malignant neoplasm of digestive tract and peritoneum NOS	B1zz.00
20	60247	Malig neop of connective and soft tissue of abdomen NOS	B314z00
20	60511	Carcinoma in situ of digestive organs	B80..00
20	94286	[M]Congenital fibrosarcoma	BBG8.11
20	44108	Malignant neoplasm of retroperitoneum and peritoneum	B18..00
20	95378	Secondary and unspec malig neop diaphragmatic lymph nodes	B561200
20	72464	Malignant neoplasm of metacarpal bones	B305.12
20	59223	Malignant neoplasm of ischium	B306100
20	7473	Carcinoma in situ	B8...00
20	35364	Secondary malignant neoplasm of retroperitoneum	B576000
20	10541	[M]Papillary carcinoma NOS	BB22.00
20	73538	Secondary and unspec malig neop axilla and upper limb LN NOS	B563z00
20	20160	Malignant neoplasm of eye	B50..00
20	22524	Secondary malignant neoplasm of other specified site NOS	B58yz00
20	57854	Malignant neoplasm of inguinal region NOS	B553000
20	4118	[M]Myxoid chondrosarcoma	BBV9.00
20	32472	[M]Inflammatory carcinoma	BB9H.00
20	54691	Malignant neoplasm of lumbar vertebra	B302200
20	41953	[M]Gangliocytoma	BBc0011
20	38736	Malignant neoplasm of other and unspecified site OS	B5y..00

20	98626	Secondary and unspec malig neop supratrochlear lymph nodes	B563100
20	41313	[M]Bile duct cystadenocarcinoma	BB5D300
20	38481	[M]Epithelioid haemangioendothelioma, malignant	BBTK.00
20	21732	[M]Myxosarcoma	BBH1.00
20	26393	Malignant neoplasm of liver unspecified	B152.00
20	50519	[M]Adrenal rest tumour	BBCF.00
20	8958	Carcinoma in situ of thyroid gland	B8yy000
20	91896	[X]Mal neoplasm/connective+soft tissue of trunk,unspecified	Byu5800
20	55268	[M]Myosarcoma	BBK2100
20	93778	Malignant neoplasm of spleen NOS	B1z1z00
20	20564	[M]Carcinoma in situ NOS	BB11.00
20	52190	Secondary and unspec malig neop pulmonary lymph nodes	B561900
20	3923	[M]Carcinoid tumours	BB5R.00
20	70716	Immunoproliferative neoplasm	B62zz11
20	34878	Malignant neoplasm of medial cuneiform	B308300
20	26814	[M]Hepatoma, malignant	BB5D512
20	69146	Malignant neoplasm of bones of skull and face NOS	B300z00
20	57988	Malignant neoplasm of carpal bone - scaphoid	B305000
20	37477	[M]Schwannoma, malignant	BBe7.11
20	39629	Granulocytic sarcoma	B653100
20	60312	Malignant neoplasm other gallbladder/extrahepatic bile duct	B16y.00
20	46458	Malignant neoplasm of skin of perineum	B335600
20	11009	Malig neop oth/ill-defined sites digestive tract/peritoneum	B1z..00
20	69821	Malignant neoplasm of the pouch of Douglas	B18y600
20	94810	[M]Adenocarcinoma with spindle cell metaplasia	BBB4.00
20	5542	Polycythaemia rubra vera	B934.11
20	28069	Malignant neoplasm of retina	B505.00
20	50777	Malignant neoplasm,overlap lesion periph nerve & auton ns	B524600
20	68332	[X]2ndry malignant neoplasm/oth+unspec parts/nervous system	ByuC600
20	54627	[M]Choriocarcinoma combined with teratoma	BBR3.00
20	91457	[X]Malignant neoplasm/connective + soft tissue,unspecified	Byu5900
20	67354	[M]Sweat gland tumour NOS	BB61100
20	39554	Malignant neoplasm of vallecula	B063.00
20	16922	[M]Polycythaemia vera	BBs0.00
20	31673	[M]Osteoclastoma, malignant	BBX1.12
20	73992	Malignant neoplasm of cornea	B504.00
20	90290	Malignant neoplasm of mesentery	B18y700

20	71609	Unspec malig neop lymphoid/histiocytic nodes inguinal/leg	B62z500
20	42218	Malignant neoplasm of other specified sites	B55y.00
20	69104	Malignant neoplasm of carpal bone - lunate	B305100
20	11991	Primary vulval cancer	B454.11
20	63723	Lymphosarcoma NOS	B601z00
20	61390	Malignant neoplasm of adrenal cortex	B540000
20	63988	Malignant neoplasm of connective and soft tissue of thumb	B311500
20	62104	Malignant neoplasm of temporal bone	B300800
20	48048	[M]Giant cell and spindle cell carcinoma	BB1B.00
20	60127	[M]Myxoliposarcoma	BBJ5.12
20	24048	Malignant neoplasm of retrocaecal tissue	B180200
20	59041	Malignant neoplasm of ciliary body	B500000
20	93716	Secondary and unspec malig neop intrathoracic LN NOS	B561z00
20	40671	[X]Malignant neoplasm of male genital organs	Byu8.00
20	58016	Carcinoma in situ of parathyroid gland	B8yy200
20	21682	[M]Malignant teratoma, intermediate type	BBQ7500
20	66775	Secondary and unspec malignant neoplasm mastoid lymph nodes	B560100
20	63239	[M]Malignant histiocytosis	BBm1.00
20	57550	Carcinoma in situ of skin of eyelid including canthus	B821.00
20	54120	Secondary malignant neoplasm of other part of nervous system	B584.00
20	54631	Malignant neoplasm of pelvic bones, sacrum and coccyx	B306.00
20	18632	Malignant neoplasm of appendix	B135.00
20	63915	Secondary and unspec malig neop inguinal and lower limb LN	B564.00
20	63430	Malignant neoplasm of endocardium	B241000
20	73213	Secondary malignant neoplasm of other urinary organs	B581.00
20	37680	[M]Fibrous histiocytoma, malignant	BBGF.00
20	64345	Malignant neoplasm of connective and soft tissue, upper arm	B311100
20	34395	[M]Verrucous carcinoma NOS	BB24.00
20	62256	[M]Adrenal cortical tumours NOS	BB5hz00
20	72684	[M]Papillary cystic tumour	BB81L00
20	50289	Malignant neoplasm of heart NOS	B241z00
20	46409	Secondary and unspec malig neop pectoral lymph nodes	B563300
20	61467	[M]Follicular adenocarcinoma, trabecular type	BB5f300
20	37805	Malignant neoplasm of cricoid cartilage	B213100
20	53349	Carcinoma in situ of other specified site	B8yy.00
20	44805	Malig neop of connective and soft tissue thigh and upper leg	B312100
20	50379	[M]Synovial sarcoma NOS	BBN1.00
20	25961	[M]Large cell carcinoma NOS	BB17.00

20	3969	[M]Intracystic carcinoma NOS	BB9M.00
20	30542	Malig neop of connective and soft tissue of lower leg	B312300
20	58088	Malignant neoplasm of intrahepatic gall duct	B151400
20	38979	[M]Sertoli cell tumour	BBC9.13
20	33497	[M]Squamous cell carcinoma, microinvasive	BB2J.00
20	52493	[M]Teratoblastoma, malignant	BBQ7213
20	67712	[M]Choriocarcinoma	BBR2.00
20	16126	Primary carcinoma of liver	B150000
20	72224	Fibrosarcoma of spleen	B1z1100
20	92371	Malignant neoplasm of radius	B304300
20	45217	Carcinoma in situ of ileum	B807200
20	68358	Carcinoma in situ of urinary organs NOS	B83z.00
20	95008	[M]Gelatinous adenocarcinoma	BB82112
20	62437	Malignant reticulosis	B62x400
20	60242	Reticulosarcoma of unspecified site	B600000
20	59310	[M]Osteochondrosarcoma	BBV1.12
20	30645	Malignant neoplasm of skin of cheek, external	B333000
20	57680	[M]Spinous cell carcinoma	BB2A.12
20	64567	Other immunoproliferative neoplasms	B63y.00
20	33951	[M]Pineocytoma	BBa2.00
20	49825	[M]Reticulum cell sarcoma NOS	BBh0.11
20	54186	Malignant neoplasm of diaphragm	B313100
20	58836	Malig neop of connective and soft tissue of pelvis NOS	B315z00
20	39433	Secondary and unspec malig neop submandibular lymph nodes	B560500
20	40492	[M]Triton tumour, malignant	BBe9.00
20	6170	Carcinomatosis	B590.11
20	62584	Secondary malignant neoplasm of other respiratory organs	B573.00
20	55953	Malignant neoplasm of occipital bone	B300400
20	40966	Malignant neoplasm of sacral vertebra	B306300
20	98408	Malig neop of connective and soft tissue of thorax NOS	B313z00
20	66166	Malignant neoplasm, overlapping lesion of small intestine	B124.00
20	63995	Malignant neoplasm of Meckel's diverticulum	B123.00
20	45573	[M]Carcinoid tumours NOS	BB5Rz00
20	61115	[M]Sex cord tumour with annular tubules	BBC0000
20	54253	[X]Secondary malignant neoplasm of other specified sites	ByuC700
20	60775	[M]Adrenal cortical carcinoma	BB5h100
20	40608	[X]Malignant neoplasm of thyroid and other endocrine glands	ByuB.00
20	27528	Malignant neoplasm of ribs, sternum and clavicle	B303.00
20	54182	Naevoid basal cell carcinoma syndrome	B33z100
20	55658	[M]Orchioblastoma	BBQ4.12
20	44267	Malignant histiocytosis	B623.00
20	54276	[M]Pseudosarcomatous carcinoma	BB1E.00

20	34879	[M]Cylindroid adenocarcinoma	BB5J.11
20	64309	[X]Malignant neoplasm of endocrine gland, unspecified	ByuB100
20	37081	[M]Smooth muscle tumour NOS	BBK3800
20	38575	[M]Apocrine adenocarcinoma	BB62100
20	58883	Carcinoma in situ of pyloric canal	B802400
20	51237	Malignant neoplasm of rib, sternum and clavicle NOS	B303z00
20	53504	Malig neopl, overlap lesion brain & other part of CNS	B52W.00
20	54654	[M]Sex cord-stromal tumour	BBC0.00
20	64874	[M]Melanotic neuroectodermal tumour	BBa4.00
20	42553	[M]Adenocarcinoma with cartilaginous and osseous metaplasia	BBB3.00
20	16915	Malignant neoplasm of intrahepatic bile ducts	B151.00
20	8154	Malignant ascites	B576200
20	70724	Myeloid sarcoma	B653.00
20	40749	[X]Malignant neoplasm of bone and articular cartilage	Byu3.00
20	49862	[M]Osteoblastic sarcoma	BBV1.11
20	52570	Malignant neoplasm, overlapping lesion of penis	B487.00
20	55947	[M]Pleomorphic liposarcoma	BBJ7.00
20	18658	Secondary and unspec malig neop common iliac lymph nodes	B562300
20	45458	[M]Squamous cell carcinoma, spindle cell type	BB2F.00
20	34713	[M]Ganglioneuroma	BBc0000
20	87003	[M]Mesenchymoma, malignant	BBLC100
20	60504	[M]Lymphocytic lymphosarcoma NOS	BBgC.12
20	73988	Malignant neoplasm of peripheral nerve of pelvis	B524500
20	98009	[M]Granulocytic sarcoma	BBrA312
20	39413	Malignant neoplasm of pelvic peritoneum	B18y500
20	34269	[M]Sebaceous adenocarcinoma	BB69100
20	49525	Kaposi's sarcoma, unspecified	B59zX00
20	23861	Malignant neoplasm of chest wall NOS	B551100
20	24456	Malig neop auditory tube, middle ear and mastoid air cells	B201.00
20	58871	Malignant histiocytosis NOS	B623z00
20	60052	Malignant neoplasm of specified site NOS	B55yz00
20	65642	Malignant histiocytosis of intra-abdominal lymph nodes	B623300
20	61579	Neoplasm of unspecified nature of respiratory system	BA01.00
20	21609	[M]Carcinoma, undifferentiated type, NOS	BB18.00
20	67396	Secondary malig neop of retroperitoneum and peritoneum	B576.00
20	37501	Carcinoma in situ of hepatic duct	B808200
20	66541	[M]Round cell carcinoma	BB1J.12
20	39312	[M]Cystosarcoma phyllodes NOS	BBM8.00
20	44066	[M]Eccrine acrospiroma	BB63.00
20	45262	[X]Malignant neoplasm of male genital organ, unspecified	Byu8200

20	94415	Malignant histiocytosis of lymph nodes head, face and neck	B623100
20	89657	Malignant mast cell tumour NOS	B626z00
20	43968	Neoplasm of unspecified nature of digestive system	BA00.00
20	47810	Malignant neoplasm of unspecified site	B59..00
20	35039	Malignant neoplasm, overlapping lesion of biliary tract	B163.00
20	11754	[M]Sclerosing stromal tumour	BBCG.00
20	24375	Dermatofibrosarcoma protuberans	B339.00
20	88144	Malignant neoplasm of other specified part of nervous system	B52y.00
20	44931	Secondary and unspec malig neop intra-abdominal LN NOS	B562z00
20	29789	Histiocytic tumour NOS	B935.11
20	54267	Malignant neoplasm of unspecified site NOS	B59z.00
20	68730	[M]Ameloblastic fibrosarcoma	BBZN.00
20	73556	Malignant neoplasm of hand bones NOS	B305z00
20	59651	[M]Mixed type liposarcoma	BBJ8.00
20	32213	[M]Malignant tumour, fusiform cell type	BB0A.00
20	40598	[X]Malignant neoplasm of female genital organs	Byu7.00
20	57677	[M]Hepatoblastoma	BBL8.00
20	69210	[M]Goblet cell tumour	BB5R611
20	34110	[M]Carcinoid tumour, malignant	BB5R100
20	28806	[M]Oncocytoma	BB5W013
20	42569	Malignant neoplasm of respiratory tract NOS	B2zz.00
20	4554	Malignant neoplasm of vulva unspecified	B454.00
20	92703	Secondary and unspec malig neop deep parotid lymph nodes	B560400
20	88593	[M]Placental site trophoblastic tumour	BBR6.00
20	63247	[M]Sarcoma botryoides	BBK3611
20	89916	Malignant neoplasm of presacral region	B553100
20	50946	[M]Medullary carcinoma with amyloid stroma	BB9C.00
20	94083	[M]Solid carcinoma NOS	BB5P.00
20	64050	Carcinoma in situ of respiratory system	B81..00
20	48275	[M]Embryonal rhabdomyosarcoma	BBK3600
20	69208	Carcinoma in situ of female genital organs NOS	B833z00
20	37810	Malignant neoplasm of trachea NOS	B220z00
20	30189	[M]Intraductal papillary adenocarcinoma with invasion	BB91000
20	28941	[M]Embryonal carcinoma NOS	BBQ3.00
20	95644	Malignant neoplasm of heart	B241.00
20	88022	[X]Secondary malignant neoplasm/oth+unspcfd digestive organs	ByuC400
20	49054	Malignant neoplasm of scapula	B304000
20	56600	[M]Epidermoid carcinoma NOS	BB2A.11
20	95559	Carcinoma in situ of specified parts respiratory system NOS	B81yz00

20	62941	[M]Neurofibrosarcoma	BBe2.00
20	52684	[M]Mesenchymal chondrosarcoma	BBW9.00
20	72445	Malignant neoplasm of cystic duct	B161000
20	54613	Malignant neoplasm of tympanic antrum	B201200
20	50681	Malignant neoplasm of prepuce (foreskin)	B480.00
20	26413	[M]Pleomorphic carcinoma	BB1A.00
20	39027	[X]Malignant neoplasm of other specified sites	ByuC000
20	36790	Primary polycythaemia	B934.12
20	12609	[M]Carcinoma, anaplastic type, NOS	BB19.00
20	58949	Malignant neoplasm of phalanges of foot	B308D00
20	15182	Malignant neoplasm of connective and soft tissue, site NOS	B31z.00
20	10913	[M]Paraganglioma NOS	BBD0.00
20	50569	Neo/uncertn+unknwn behav/lymph,h'matopetc+rel tiss,unspcf	B93X.00
20	89593	Malignant neoplasm of intrahepatic biliary passages	B151200
20	67949	Malignant neoplasm of other male genital organ	B48y.00
20	73076	Carcinoma in situ of vestibular fold	B810700
20	98797	[M]Embryonal sarcoma	BBLD.00
20	38611	Carcinoma in situ of other and unspecified sites	B8y..00
20	61289	Secondary and unspec malig neop deep inguinal lymph nodes	B564100
20	31324	Mast cell malignancy of lymph nodes of multiple sites	B626800
20	34000	[M]Cystadenocarcinoma NOS	BB80100
20	31399	Malignant neoplasm of lower limb NOS	B555.00
20	12119	Carcinoma in situ of vulva	B833300
20	27651	Secondary carcinoma of other specified sites	B58..11
20	54965	Malig neop connective and soft tissue of popliteal space	B312200
20	37540	Secondary and unspec malig neop axillary lymph nodes	B563000
20	38918	Secondary malignant neoplasm of spinal cord	B583100
20	73296	[X]Malignant neoplasm/bones+articular cartilage/limb,unspfd	Byu3100
20	67763	Malignant neoplasm of costo-vertebral joint	B303400
20	10995	Malignant neoplasm of other and unspecified sites	B5...00
20	5136	Choriocarcinoma	B911013
20	49292	[X]Malignant neoplsm/ill-defin sites within digestive system	Byu1300
20	64089	Carcinoma in situ of common bile duct	B808500
20	92382	Malignant neoplasm of fourth metatarsal bone	B308B00
20	22187	Hepatocellular carcinoma	B150300
20	46613	Malignant neoplasm of specified parts of peritoneum	B18y.00
20	18314	Malignant neoplasm of bone and articular cartilage	B30..00
20	57729	[M]Lymphangiosarcoma	BBU1.00
20	18266	[M]Granular cell tumour NOS	BBf0.00
20	68456	[M]Chromophobe carcinoma	BB5V100

20	51352	Malignant neoplasms of independent (primary) multiple sites	B592.00
20	65233	Malig neop connective and soft tissue other specified site	B31y.00
20	16298	Malignant neoplasm of retroperitoneum and peritoneum NOS	B18z.00
20	70740	[M]Malignant reticulosis	BBm1.11
20	66750	Malignant neoplasm of heart, thymus and mediastinum NOS	B24z.00
20	72503	Neoplasm uncert / unkn behav brain, infratentorial	B925300
20	59152	Malignant neoplasm of connective and soft tissue of perineum	B315200
20	9291	[M]Small cell carcinoma NOS	BB1J.00
20	15036	Malignant mast cell tumours	B626.00
20	27853	HIV disease resulting in Kaposi's sarcoma	A789500
20	31393	Carcinoma gallbladder	B160.11
20	38770	[M]Epithelial-myoepithelial carcinoma	BBB7.00
20	98322	[M]Haemangioendothelioma, malignant	BBT7100
20	60631	[M]Osteosarcoma in Paget's disease of bone	BBV5.00
20	54222	Malignant neoplasm of connective and soft tissue of foot	B312400
20	65605	Malignant neoplasm of myocardium	B241200
20	19437	Osteosarcoma	B30z000
20	92275	[M]Gonadal stromal tumour	BBC0.11
20	65861	[M]Dermoid cyst with malignant transformation	BBQ9.00
20	65466	Kaposi's sarcoma of multiple organs	B592X00
20	45267	Malignant neoplasm of other and ill defined site NOS	B55z.00
20	33997	Malignant neoplasm of skin of auricle (ear)	B332000
20	54190	[M] Angioimmunoblastic lymphadenopathy	BBm8.00
20	50292	[X]Malignant neoplasm of mediastinum, part unspecified	Byu2500
20	94272	Malig neoplasm of connective and soft tissues of lumb spine	B314100
20	36147	Secondary malignant neoplasm of liver	B153.00
20	61082	[M]Pneumoblastoma	BBLA.11
20	31323	[M]Fibrosarcoma NOS	BBG1.00
20	21833	[M]Duct carcinoma NOS	BB91.11
20	57756	[X]Malignant neoplasm/other specified female genital organs	Byu7100
20	66083	Secondary malig neop of respiratory or digestive system NOS	B57z.00
20	59415	[M]Thymoma, malignant	BBB6100
20	12580	[M]Adenosquamous carcinoma	BBB0.00
20	65458	Malig neop of other and unspecified parts of nervous system	B52..00
20	71625	Lymphosarcoma of unspecified site	B601000
20	27416	Lymphosarcoma	B601.00
20	28148	Malignant neoplasm of adrenal gland	B540.00

20	95058	Reticulosarcoma of spleen	B600700
20	42528	[M]Hibernoma	BBJD.00
20	43392	Malignant neoplasm of penis, part unspecified	B483.00
20	63518	[M]Adenosarcoma	BBLE.00
20	1481	Reticulosarcoma	B600.00
20	43479	Malignant neoplasm of jejunum	B121.00
20	22290	Malignant neoplasm of connective and soft tissue of thorax	B313.00
20	8085	[M]Sarcoma NOS	BBF1.00
20	20807	[M]Papillary squamous cell carcinoma	BB26.00
20	64106	Malignant neoplasm of specified parts of peritoneum NOS	B18yz00
20	46497	Carcinoma in situ of pleura	B81y000
20	38444	[M]Carcinoid tumour NOS	BB5R000
20	50290	Kaposi's sarcoma of lymph nodes	B6z0.00
20	57184	[X]Oth malignant neoplasm/skin of oth+unspecfd parts of face	Byu4200
20	65490	Secondary cancer of the vulva	B58y411
20	10698	Malignant neoplasm of vaginal vault	B450100
20	50605	[M]Glomangiosarcoma	BBDB.00
20	33250	Ca-in-situ of G.I. tract	B80..11
20	37842	Malignant neoplasm of rib	B303000
20	50222	Malignant neoplasm of connective and soft tissue of shoulder	B311000
20	27311	Carcinoma in situ of penis	B835.00
20	27931	Kaposi's sarcoma of skin	B33z000
20	48223	[M]Scirrhous adenocarcinoma	BB54.00
20	57336	[M]Epithelioma, malignant	BB16.00
20	53528	Secondary malignant neoplasm of urethra	B581200
20	66000	[M]Adenocarcinoma with apocrine metaplasia	BBB5.00
20	44399	Primary malignant neoplasm of liver NOS	B150z00
20	45953	[M]Glomus jugulare tumour	BBD4.00
20	31026	[M]Spindle cell sarcoma	BBF3.00
20	68410	Primary angiosarcoma of liver	B150200
20	73164	Carcinoma in situ of cystic duct	B808400
20	50140	[M]Cribriform carcinoma	BB5K.00
20	41144	Secondary malignant neoplasm of skin of trunk	B582300
20	51748	Carcinoma in situ of pyloric antrum	B802300
20	51795	Malignant neoplasm of glomus jugulare	B545000
20	96893	[M]Myeloid sarcoma	BBrA300
20	19945	Secondary malignant neoplasm of skin	B582.00
20	63568	Malignant neoplasm of peripheral nerves of head, face & neck	B524000
20	72500	[X]Mal neoplasm/lymphoid,haematopoietic+related tissu,unspcf	ByuDB00

20	70728	Carcinoma in situ other and unspecified small intestine NOS	B807z00
20	32641	[M]Merkel cell carcinoma	BB5RA00
20	50199	Secondary and unspec malig neop axilla and upper limb LN	B563.00
20	74896	Malignant neoplasm of extrahepatic bile ducts NOS	B161z00
20	55463	Secondary and unspec malig neop post mediastinal lymph nodes	B561400
20	43390	Malignant neoplasm of small intestine NOS	B12z.00
20	56345	Secondary malignant neoplasm of other digestive organ	B57y.00
20	16500	Secondary malignant neoplasm of other specified site NOS	B58z.00
20	72440	[M]Acinar cell tumour	BBA1.00
20	34742	Malignant neoplasm of pleura NOS	B23z.00
20	55246	Malignant neoplasm of accessory sinus NOS	B20z.00
20	64680	Secondary malignant neoplasm of small intestine and duodenum	B574.00
20	58962	Malignant immunoproliferative small intestinal disease	B62x500
20	71810	Malignant neoplasm of scapula and long bones of upper arm	B304.00
20	70026	Secondary malig neop of small intestine or duodenum NOS	B574z00
20	66673	Carcinoma in situ of liver and biliary system	B808.00
20	45922	Malignant neoplasm, overlapping lesion of eye and adnexa	B508.00
20	16213	Secondary malignant neoplasm of pleura	B572.00
20	61064	Malignant neoplasm of mediastinum, part unspecified	B24X.00
20	24301	Secondary carcinoma of respiratory and/or digestive systems	B57..12
20	98540	Carcinoma in situ of liver or biliary system NOS	B808z00
20	35053	Secondary malig neop of respiratory and digestive systems	B57..00
20	20159	Secondary and unspec malig neop lymph nodes multiple sites	B56y.00
20	94750	[M]Mastocytoma NOS	BBp0.00
20	62380	Lymphosarcoma of intrathoracic lymph nodes	B601200
20	63300	[X]Malignant neoplasm/overlap lesion/bone+articulr cartilage	Byu3200
20	51878	[M]Aesthesioneuroblastoma	BBcC.00
20	71869	[M]Alveolar soft part sarcoma	BBf2.00
20	40437	Malignant neoplasm of other specified site of eye	B50y.00
20	51551	Secondary malignant neoplasm of mediastinum	B571.00
20	34891	[M]Sarcomatosis NOS	BBF2.00
20	47656	Carcinoma in situ of appendix	B803500
20	98361	[X]Kaposi's sarcoma of other sites	Byu5B00

20	66163	[X]2ndry+unspcf malignant neoplasm lymph nodes/multi regions	ByuC200
20	18613	Malignant neoplasm of duodenum	B120.00
20	51714	Carcinoma in situ of trachea	B811.00
20	65793	Malig neop of upper respiratory tract, part unspecified	B2z0.00
20	69132	Secondary and unspec malig neop external iliac lymph nodes	B562400
20	22392	Carcinoma in situ of other and unspecified small intestine	B807.00
20	57481	[X]Secondary malignant neoplasm/oth+unspc respiratory organs	ByuC300
20	72912	[M]Tubular androblastoma with lipid storage	BBCB.00
20	15103	Secondary malignant neoplasm of liver	B577.00
20	41691	Secondary and unspec malig neop coeliac lymph nodes	B562000
20	71895	[M]Superficial spreading adenocarcinoma	BB56.00
20	68161	Malignant neoplasm of seminal vesicle	B48y000
20	72127	Malignant neoplasm of epididymis	B484.00
20	86996	Malignant neoplasm of connective tissue of orbit	B501000
20	94438	[M]Signet ring carcinoma NOS	BB85z00
20	52736	Secondary and unspec malig neop intra-abdominal lymph nodes	B562.00
20	72212	Malignant neoplasm of calcaneum	B308200
20	97593	[M]Ameloblastoma, malignant	BBZG.00
20	25366	Secondary and unspec malig neop ant mediastinal lymph nodes	B561300
20	65880	Malig neop of scapula and long bones of upper arm NOS	B304z00
20	7830	Lymph node metastases	B56..11
20	54278	Secondary and unspec malig neop superficial inguinal LN	B564000
20	18354	Malignant neoplasm of other specified skin sites	B33y.00
20	33871	Malignant neoplasm of ileum	B122.00
20	37688	[M]Acinar cell carcinoma	BBA2.00
20	9618	Secondary and unspecified malignant neoplasm of lymph nodes	B56..00
20	53129	[M]Oncytic adenocarcinoma	BB5W112
20	95024	[M]Infantile fibrosarcoma	BBG8.00
20	97672	Secondary malig neop of retroperitoneum or peritoneum NOS	B576z00
20	18231	Phaeochromocytoma	B540.11
20	3152	[M]Carcinoma, metastatic, NOS	BB13.00
20	52316	Malignant neoplasm of pelvis	B553.00
20	73916	[M]Epithelioid leiomyosarcoma	BBK0400
20	46159	Malignant neoplasm of cloacogenic zone	B142000
20	97547	Malignant neoplasm of intrathoracic site NOS	B551200
20	36124	[M]Eccrine dermal cylindroma	BB5H.00
20	41816	[M]Squamous cell carcinoma, small cell, non-keratinising	BB2E.00
20	51818	Malignant neoplasm of jaw NOS	B550300

C. Cancer medcodes

20	40449	[M]Gonadoblastoma	BBQ6.00
20	2123	[M]Neuroblastoma NOS	BBc1.00
20	41931	Malignant neoplasm of cheek NOS	B550100
20	24293	[M]Squamous cell carcinoma, metastatic NOS	BB2B.00
20	9030	Malignant neoplasm of other and ill-defined sites	B55..00
20	55588	[X]Malignant neoplasm of female genital organ, unspecified	Byu7300
20	59251	[M]Cystosarcoma phyllodes, malignant	BBM9.00
20	87113	Malignant neoplasm-pluriglandular involvement,unspecified	B54X.00
20	67211	Malignant neoplasm of spinal meninges NOS	B523z00
20	15907	Malignant neoplasm gallbladder/extrahepatic bile ducts NOS	B16z.00
20	39121	[M]Ganglioneuroblastoma	BBc0100
20	67324	Malig neop of connective and soft tissue of inguinal region	B315100
20	8660	[M]Osteosarcoma NOS	BBV1.00
20	33395	Secondary and unspec malig neop superficial cervical LN	B560200
20	99219	[M]Interstitial cell tumour NOS	BBCCz11
20	30646	Malignant neoplasm lymphatic or haematopoietic tissue OS	B6y..00
20	53910	Malignant neoplasm of clitoris	B453.00
20	26253	[M]Neuroendocrine carcinoma	BB5R900
20	70824	Malignant neoplasm of adrenal gland NOS	B540z00
20	35136	Carcinoma in situ NOS	B8z..00
20	94239	[M]Mast cell sarcoma	BBp1.00
20	37542	[M]Teratocarcinoma	BBQ7300
20	5637	Malignant neoplasm of thyroid gland	B53..00
20	56490	Malignant neoplasm of nervous system NOS	B52z.00
20	16146	[M]Adenocarcinoma with squamous metaplasia	BBB2.00
20	86046	Malignant neoplasm of peripheral nerve of abdomen	B524400
20	28003	Choriocarcinoma	B420.00
20	34946	Carcinoma in situ of vagina	B833200
20	37354	[M]Clear cell adenocarcinoma NOS	BB5X100
20	16902	[M]Basal cell adenocarcinoma	BB5y000
20	34096	[M]Granular cell carcinoma	BB5b.00
20	94220	Malignant neoplasm of adrenal medulla	B540100
20	96226	[X]Malignant neoplasm/overlap lesion/other+ill-defined sites	ByuC100
20	26034	Other malignant neoplasm NOS	B591.00
20	27391	Secondary malignant neoplasm of peritoneum	B576100
20	36495	Carcinoma common bile duct	B161211
20	21330	Malignant neoplasm of retroperitoneum	B180.00
20	61928	[M]Squamous cell ca-in-situ, questionable stromal invasion	BB2H.00
20	61741	Malignant neoplasm of humerus	B304200

20	16692	[M]Carcinomatosis	BB14.00
20	20166	Malignant neoplasm of female genital organ NOS	B45z.00
20	65953	Carcinoma in situ of arytenoid cartilage	B810300
20	44157	[M]Melanosarcoma NOS	BBE1.13
20	21847	[M]Follicular carcinoma	BB5f111
20	26454	Malignant neoplasm/overlapping lesion/feml genital organs	B45X.00
20	67034	[X]Mesothelioma of other sites	Byu5000
20	39899	Malignant neoplasm of craniopharyngeal duct	B542100
20	7654	Secondary malignant neoplasm of bone and bone marrow	B585.00
20	61677	Secondary and unspec malig neop inferior mesenteric LN	B562200
20	50859	[M]Giant cell bone sarcoma	BBX1.11
20	27849	[M]Villous adenocarcinoma	BB5U200
20	5052	[M]Osteogenic sarcoma NOS	BBV1.13
20	52537	Malignant neoplasm of hepatic duct	B161100
20	49145	Secondary malignant neoplasm of penis	B58y700
20	47899	Malignant neoplasm of greater vestibular (Bartholin's) gland	B451000
20	73510	Malignant neoplasm of supraclavicular fossa NOS	B550500
20	50898	Malignant neoplasm of omentum	B18y300
20	44217	[M]Mixed tumour NOS	BBL3.12
20	49900	[M]Klatskin's tumour	BB5y200
20	61984	[M]Spheroidal cell carcinoma	BB1G.00
20	66488	Malig neop of connective and soft tissue of abdominal wall	B314000
20	56794	[M]Adenocarcinoid tumour	BB5R800
20	39130	[M]Carcinoid tumour, argentaffin, NOS	BB5R200
20	47920	[M]C cell carcinoma	BB9B.11
20	37328	Malignant neoplasm of vagina	B450.00
20	60186	[M]Refractory anaemia+excess of blasts with transformation	BBmB.00
20	68027	[X]Malignant neoplasm/other and unspecified cranial nerves	ByuA000
20	38343	Secondary and unspec malig neop submental lymph nodes	B560700
20	17559	Malignant neoplasm of intestinal tract, part unspecified	B1z0.00
20	63695	Malignant neoplasm of peripheral nerve of thorax	B524300
20	24511	[M]Malignant tumour, giant cell type	BB09.00
20	57482	Malignant neoplasm of connective and soft tissue of forearm	B311200
20	33636	[M]Teratoma, malignant, NOS	BBQ7200
20	98781	[M]Trabecular adenocarcinoma	BB5F.00
20	60772	Malignant neoplasm of vagina NOS	B450z00
20	23433	Malignant neoplasm of extrahepatic bile ducts	B161.00
20	25641	[M]Liver cell carcinoma	BB5D513

20	60403	Malignant neoplasm of costal cartilage	B303300
20	59388	Malignant neoplasm of mesocaecum	B18y100
20	31210	Hepatoblastoma of liver	B150100
20	71627	[M]Sweat gland adenocarcinoma	BB61200
20	46581	[M]Pleomorphic cell sarcoma	BBF4.11
20	56918	Malignant neoplasm other spec digestive tract and peritoneum	B1zy.00
20	72277	[M]Basophil carcinoma	BB5V700
20	65434	Malignant neoplasms of lymphoid and histiocytic tissue NOS	B62z.00
20	32442	[M]Chondromatous giant cell tumour	BBW7.11
20	54493	Malignant neoplasm of xiphoid process	B303500
20	54747	Malignant neoplasm of parietal bone	B300600
20	11035	Primary malignant neoplasm of unknown site	B593.00
20	4218	Malignant neoplasm of parathyroid gland	B541.00
20	25535	Primary malignant neoplasm of liver	B150.00
20	43151	[X]Malignant neoplasm/bone+articular cartilage, unspecified	Byu3300
20	31090	[M]Pigmented dermatofibrosarcoma protuberans	BBGP.00
20	44356	Malig neop other/ill-defined sites resp/intrathoracic organs	B2z..00
20	4403	Liver metastases	B577.11
20	60181	Neoplasm of unspecified nature of other specified sites	BA0y.00
20	98883	[M]Medullary carcinoma with lymphoid stroma	BB9D.00
20	92329	Malignant neoplasm of other male genital organ NOS	B48yz00
20	5455	[M]Adenocarcinoma, metastatic, NOS	BB53.00
20	48743	Malignant neoplasm of body of penis	B482.00
20	40438	[M]Bile duct carcinoma	BB5D111
20	47366	Secondary and unspec malig neop sacral lymph nodes	B565300
20	44166	Carcinoma in situ of other and unspecified digestive organs	B80z.00
20	63331	Malignant neoplasm of spermatic cord	B485.00
20	86812	Malignant neoplasm of phalanges of hand	B305D00
20	68824	Malignant neoplasm, overlapping lesion male genital orgs	B48y200
20	63224	Malignant neoplasm of penis and other male genital organ NOS	B48z.00
20	65216	[M]Cloacogenic carcinoma	BB49.00
20	22561	[M]Telangiectatic osteosarcoma	BBV4.00
20	59143	[M]Squamous cell carcinoma, large cell, non-keratinising	BB2D.00
20	93665	[X]Kaposi's sarcoma, unspecified	Byu5300
20	45510	[M]Lymphoepithelial carcinoma	BB2M.00
20	47862	Malignant neoplasm of thyroid cartilage	B213300
20	32372	Malignant neoplasm of thoracic vertebra	B302100
20	9491	Anal carcinoma	B142.11
20	71497	[M]Oxyphilic adenocarcinoma	BB5W100

20	68956	[M]Giant cell tumour of bone, malignant	BBX1.00
20	91586	Malignant neoplasm of connective and soft tissue of finger	B311400
20	27483	Malignant neoplasm of thymus	B240.00
20	16075	Malignant neoplasm of bone and articular cartilage NOS	B30z.00
20	25310	Carcinoma in situ of liver	B808000
20	29008	[M]Hurthle cell adenocarcinoma	BB5W111
20	35457	[M]Basosquamous carcinoma	BB35.00
20	27562	[M]Follicular lymphosarcoma NOS	BBk0.12
20	1624	[M]Squamous cell carcinoma NOS	BB2A.00
20	21861	[M]Lobular carcinoma in situ	BB9E.00
20	38593	[M]Adamantinoma of long bones	BBY1.00
20	19041	[M]Intraepidermal carcinoma NOS	BB29.12
20	19334	[M]Carcinosarcoma NOS	BBL9.00
20	63104	Malignant neoplasm of orbit NOS	B501z00
20	52591	[M]Lymphoblastoma NOS	BBgG.13
20	33775	[M]Adenoid cystic carcinoma	BB5J.00
20	7856	[M]Dedifferentiated liposarcoma	BBJH.00
20	28178	[M]Craniopharyngioma	BBa0.00
20	49714	Malignant neoplasm of spinal meninges	B523.00
20	49797	[M]Argentaffinoma NOS	BB5R211
20	18616	Secondary malignant neoplasm of other specified sites	B58y.00
20	45070	Carcinoma in situ of duodenum	B807000
20	68220	[M]Fibrochondrosarcoma	BBW4.11
20	37621	[M]Endodermal sinus tumour	BBQ4.00
20	65599	Malignant neoplasm of acoustic nerve	B520200
20	65460	Malignant neoplasm of spleen NEC	B1z1.00
20	57796	[M]Synovial sarcoma, biphasic type	BBN4.00
20	64971	Malignant neoplasm of olfactory bulb	B520000
20	70747	Secondary and unspec malig neop of inguinal and leg LN NOS	B564z00
20	62630	Malignant neoplasm of long bones of leg NOS	B307z00
20	39038	[M]Signet ring carcinoma	BB85.00
20	38454	[M]Basaloid carcinoma	BB48.00
20	62348	[M]Haemangiosarcoma	BBT1.00
20	44420	Refractory anaemia with excess of blasts with transformation	B937300
20	66607	[M]Mixed tumour, malignant, NOS	BBL4.00
20	90659	Malignant neoplasm of other specified endocrine gland	B54y.00
20	14825	Neoplasm of unspecified nature	BA0..00
20	57802	[M]Alveolar adenocarcinoma	BB5S400
20	94427	Malignant neoplasm of fifth metacarpal bone	B305C00
20	48282	Neoplasm of unspecified nature of bone	BA02000
20	54874	[M]Metastatic signet ring cell carcinoma	BB85100
20	7941	[M]Chondrosarcoma NOS	BBW4.00

20	22650	[M]Angiosarcoma	BBT1.11
20	43490	[X]Other specified carcinomas of liver	Byu1100
20	50402	Malignant neoplasm of fibula	B307100
20	58837	[M]Small cell sarcoma	BBF5.00
20	63804	Carcinoma in situ of jejunum	B807100
20	67451	Malignant neoplasm/overlap lesion/bone+articulr cartilage	B30W.00
20	10258	[M]Neoplasms NOS	BB0..00
20	21868	[M]Neoplasm, malignant	BB02.00
20	3197	[M]Neoplasm, metastatic	BB03.00
20	6985	[M]Secondary neoplasm	BB03.11
20	64897	[X]Malignant neoplasms/independent(primary)multiple sites	ByuE000
20	90546	Malig neop connective and soft tissue hip and leg NOS	B312z00
20	12265	Primary thrombocythaemia	B937411
20	59787	[M]Tubular androblastoma NOS	BBC9.00
20	15507	Secondary and unspec malig neop lymph nodes NOS	B56z.00
20	30416	[M]Colloid adenocarcinoma	BB82111
20	24235	Malig neopl peripheral nerves and autonomic nervous system	B524.00
20	26630	Unspecified nature neoplasm	BA...00
20	58124	Carcinoma in situ of eye	B8y0.00
20	55595	Malignant neoplasm of sphenoid bone	B300700
20	69844	[M]Round cell sarcoma	BBF5.11
20	44609	Malignant neoplasm of ilium	B306000
20	57505	[M]Pleomorphic rhabdomyosarcoma	BBK3200
20	63657	Malignant neoplasm of conjunctiva	B503.00
20	38651	[M]Papillary carcinoma in situ	BB21.00
20	35285	[X]Malignant neoplasm of eye, brain and other parts of cent	ByuA.00
20	94776	Malignant neoplasm, overlapping lesion of digestive system	B1z2.00
20	3028	[M]Basal cell carcinoma NOS	BB31.00
20	56676	[M]Myxoid liposarcoma	BBJ5.00
20	3348	[M]Histiocytoma NOS	BBGK.12
20	93852	[M]Lymphangiomyoma	BBU5.00
20	96445	Malignant neoplasm of turbinate	B300B00
20	95182	Malignant neoplasm of talus	B308100
20	44627	Secondary and unspec malig neop anterior cervical LN	B560800
20	95818	[M]Paraganglioma, malignant	BBD1.00
20	29580	[M]Sertoli cell carcinoma	BBCA.00
20	84368	Secondary and unspec malig neop internal iliac lymph nodes	B565000
20	61555	Malignant neoplasm of retroperitoneum NOS	B180z00

20	98813	Malig neop eyeball excl conjunctiva, cornea, retina, choroid	B500.00
20	62871	[M]Comedocarcinoma, noninfiltrating	BB92.00
20	36209	[M]Carotid body tumour	BBD6.00
20	67966	[M]Naevocarcinoma	BBE1.14
20	13574	[M]Metatypical carcinoma	BB36.00
20	15976	Malignant neoplasm of abdomen	B552.00
20	57047	Malignant neoplasm of carotid body	B544.00
20	45766	[X]Malignant neoplasm of intestinal tract, part unspecified	Byu1200
20	54956	Malignant neoplasm of eye NOS	B50z.00
20	46478	Carcinoma in situ of adrenal gland	B8yy100

Appendix D

Symptom codes

This appendix details the relevant medcodes used in the investigation of symptoms of cancer in this thesis:

- Table D.1 displays a key for the codes.
- Table D.2 contains a list of medcodes with a description.

Table D.1: Keys for codes and their symptoms used in the table of medcodes (Table C.2).

Code	Symptom
1	Abdominal pain
2	Iron-deficient anaemia
3	Diarrhoea
4	Constipation
5	Rectal bleed
6	Change in bowel habit
7	Abdominal mass
8	Cough
9	Fatigue
10	Shortness of breath
11	Chest pain
12	Loss of appetite
13	Weight loss
14	Lower abdominal pain

Table D.2: Description of symptoms with medcode reference ID fields.

Medcode	Description	Symptom code
177	Abdominal pain	1
628	[D]Hypochondrial pain	1
716	[D]Abdominal cramps	1
1239	[D]Colic NOS	1
1336	[D] Groin pain	1
1763	[D]Abdominal pain	1
1976	Abdominal pain type	1
2056	[D]Abdominal colic	1
2234	[D]Recurrent acute abdominal pain	1
2383	Abdominal discomfort	1
3338	[D]Abdominal pain NOS	1
4617	Central abdominal pain	1
4771	[D]Umbilical pain	1
5691	Non-colicky abdominal pain	1
5782	O/E - abdomen tender	1
5960	Site of abdominal pain	1
6357	Subcostal pain	1
6395	Type of GIT pain - symptom	1
7726	[D]Right upper quadrant pain	1
7812	Colicky abdominal pain	1
8362	[D]Left upper quadrant pain	1
9695	Right upper quadrant pain	1
11070	General abdominal pain-symptom	1
12639	O/E - abd.pain-R.hypochondrium	1
14916	O/E -abd.pain on palpation NOS	1
14989	Type of GIT pain	1
15180	O/E - abdo. pain on palpation	1
15908	Appendicular colic	1
16402	[D]Abdominal tenderness	1
17324	Griping pain	1
17636	O/E - umbilical pain on palp.	1
19283	[D]Nonspecific abdominal pain	1
23756	[D]Evening colic	1
23872	Left subcostal pain	1
24627	O/E - abd. pain - umbilical	1
24661	Generalised abdominal pain	1
24821	Right subcostal pain	1
25118	Site of GIT pain	1
28285	[D]Gas pain (abdominal)	1
29352	Abdominal wall pain	1
29922	Site of GIT pain NOS	1
31062	[D]Other specified abdominal pain	1
37101	O/E - abd.pain-L.hypochondrium	1
42195	O/E - guarding-R.hypochondrium	1
42211	O/E - abd. pain - hypogastrium	1
47423	[D]Other abdominal and pelvic symptoms	1
50590	O/E - abdominal rigidity	1
52402	[X]Other and unspecified abdominal pain	1
56084	O/E - guarding - hypogastrium	1
56094	O/E - guarding-L.hypochondrium	1
62927	O/E - rebound - umbilical	1
62933	O/E - guarding - umbilical	1
73235	O/E - abdominal rigidity NOS	1

D. Symptom codes

Medcode	Description	Symptom code
539	Microcytic - hypochromic anaemia	2
795	Iron deficiency anaemias	2
882	Hypochromic - microcytic anaemia	2
4839	Microcytic hypochromic anaemia	2
4858	[X]Other iron deficiency anaemias	2
9537	Other specified iron deficiency anaemia NOS	2
15439	Iron deficiency anaemia NOS	2
18137	Unspecified iron deficiency anaemia	2
21127	Iron deficiency anaemia due to dietary causes	2
27726	Iron deficiency anaemia due to chronic blood loss	2
33420	Other specified iron deficiency anaemia	2
48338	Iron deficiency anaemia due to blood loss	2
192	Diarrhoea	3
1695	Loose stools	3
2133	Dysenteric diarrhoea	3
2182	Diarrhoea & vomiting, symptom	3
4343	Diarrhoea	3
4542	Infectious diarrhoea	3
5036	Functional diarrhoea	3
5090	Diarrhoea of presumed infectious origin	3
5134	Diarrhoea symptoms	3
6007	Diarrhoea due to Campylobacter jejuni	3
6016	Noninfective diarrhoea	3
6083	Incontinent of faeces symptom	3
6685	Chronic diarrhoea	3
7644	Diarrhoea and vomiting	3
10158	Spurious diarrhoea	3
11155	Travellers' diarrhoea	3
13387	Allergic diarrhoea	3
14665	Diarrhoea & vomiting -? infect	3
14695	Diarrhoea symptom NOS	3
14881	[D] Stools loose	3
15289	Viral diarrhoea	3
15371	Psychogenic diarrhoea	3
15555	[D]Incontinence of faeces NOS	3
17017	Diarrhoea - presumed non-infectious	3
17162	Dietetic diarrhoea	3
21294	Spurious (overflow) diarrhoea	3
29835	Irritable bowel syndrome with diarrhoea	3
30321	Presumed noninfectious diarrhoea	3
36613	Epidemic diarrhoea	3
43316	[X]Psychogenic diarrhoea	3
48313	Infectious diarrhoea NOS	3
52750	[X]Diarrhoea+gastroenteritis of presumed infectious origin	3
53739	Diarrhoea due to staphylococcus	3
56957	Diarrhoea due to Pseudomonas pyocyanea	3
61519	Diarrhoea due to staphylococcal toxin	3
1028	Constipation	4
1709	Constipation - functional	4
2004	Constipation symptom	4
2264	Manual removal of impacted faeces from rectum	4
5803	Constipation NOS	4
6364	Chronic constipation with overflow	4
8541	Painful defaecation	4
10687	Faecal impaction	4
15939	Psychogenic constipation	4
17652	Constipated	4
20450	Constipation NOS	4

Medcode	Description	Symptom code
6364	Chronic constipation with overflow	4
8541	Painful defaecation	4
10687	Faecal impaction	4
15939	Psychogenic constipation	4
17652	Constipated	4
20450	Constipation NOS	4
20464	Removal of impacted faeces	4
23641	Acute constipation	4
24180	Other specified constipation	4
25797	Chronic constipation without overflow	4
26022	Drug induced constipation	4
97021	Difficulty in ability to defaecate	4
98785	O/E - defaec.ref.abn.-constip.	4
621	Rectal bleeding	5
2873	Blood in stool	5
3872	Bleeding PR	5
5462	Blood in faeces	5
6151	Blood in faeces symptom	5
6554	PRB - Rectal bleeding	5
6574	Rectal haemorrhage	5
7096	Perianal haematoma	5
9968	Blood on toilet paper	5
11698	Painless rectal bleeding	5
11718	Painful rectal bleeding	5
14256	Blood in faeces	5
19271	Haemorrhage of rectum and anus	5
20859	Blood in stools altered	5
27862	Altered blood in stools	5
32446	Anal haemorrhage	5
45911	Blood on pants	5
46479	Haemorrhage of rectum and anus NOS	5
61761	Perinatal rectal haemorrhage	5
910	Change in bowel habit	6
16665	[D]Change in bowel habit	6
19690	Altered bowel habit	6
3015	[D]Abdominal mass	7
5838	[D]Swelling, mass or lump within abdomen or pelvis	7
7073	[D]Abdominal lump	7
8731	O/E - abdominal mass palpated	7
20827	Right iliac fossa mass	7
24527	Epigastric mass	7
56675	O/E - abd. mass palpated NOS	7
92	Cough	8
292	Chesty cough	8
1025	Bronchial cough	8
1160	[D]Cough	8
1234	Productive cough NOS	8
1273	C/O - cough	8
1612	Chronic cough	8
3068	Night cough present	8
3628	Persistent cough	8
3645	Coughing up phlegm	8
4070	Morning cough	8
4836	Nocturnal cough / wheeze	8
4931	Dry cough	8
7706	Productive cough -clear sputum	8
7707	Cough symptom NOS	8

D. Symptom codes

Medcode	Description	Symptom code
7708	Productive cough-yellow sputum	8
7773	Productive cough -green sputum	8
8239	[D]Cough with haemorrhage	8
16717	Smokers' cough	8
29318	Evening cough	8
43795	Unexplained cough	8
1147	[D]Tiredness	9
1371	[D]Lethargy	9
1404	Fatigue	9
1582	Nervous exhaustion	9
1688	[D]Fatigue	9
2855	Weakness present	9
3192	Exhaustion due to exposure	9
4546	Chronic fatigue syndrome	9
5583	Lethargy - symptom	9
5751	Tired all the time	9
5794	Tiredness symptom	9
5814	[D]Lassitude	9
6029	Weakness symptoms	9
6190	Postviral fatigue syndrome	9
6242	Fatigue - symptom	9
7235	Tired all the time	9
7529	CFS - Chronic fatigue syndrome	9
9127	Post-viral fatigue syndrome	9
9220	Exhaustion	9
9435	O/E - apathetic	9
9656	[X]Fatigue syndrome	9
9889	[D]General weakness	9
15516	C/O - tired all the time	9
16479	O/E - weakness	9
17083	Excessive exertion exhaustion	9
17526	Maternal exhaustion	9
17736	Malaise/lethargy	9
22923	Heat exhaustion, unspecified	9
24382	[D]Senile exhaustion	9
27877	PVFS - Postviral fatigue syn	9
29181	[D]Post polio exhaustion	9
29292	Tiredness symptom NOS	9
43550	Combat fatigue	9
44215	[D]Malaise and fatigue	9
52583	Fatigue during pregnancy	9
64099	Fatigue during pregnancy unspecified	9
68930	Fatigue during pregnancy - delivered	9
70779	[X]Combat fatigue	9
73727	Fatigue during pregnancy NOS	9
97284	Moderate chronic fatigue syndrome	9
98512	Mild chronic fatigue syndrome	9
98734	Severe chronic fatigue syndrome	9
99956	Fatigue during pregnancy with postnatal complication	9
99980	Fatigue during pregnancy - not delivered	9
735	[D]Breathlessness	10
741	[D]Shortness of breath	10
1429	Breathlessness	10
2575	Short of breath on exertion	10
2931	Difficulty breathing	10
3092	[D]Dyspnoea	10
4822	Shortness of breath	10

Medcode	Description	Symptom code
5175	Breathlessness symptom	10
5349	Shortness of breath symptom	10
5896	Dyspnoea - symptom	10
6326	Breathless - moderate exertion	10
6434	Paroxysmal nocturnal dyspnoea	10
7000	O/E - dyspnoea	10
7683	Breathless - lying flat	10
7932	Breathless - mild exertion	10
12474	SOBOE	10
18116	Nocturnal dyspnoea	10
19426	MRC Breathlessness Scale: grade 3	10
19427	MRC Breathlessness Scale: grade 2	10
19429	MRC Breathlessness Scale: grade 5	10
19430	MRC Breathlessness Scale: grade 4	10
19432	MRC Breathlessness Scale: grade 1	10
21801	Breathlessness NOS	10
22094	Short of breath dressing/undressing	10
23663	Respiratory symptom NOS	10
24889	Breathless - strenuous exertion	10
31143	Breathless - at rest	10
53771	Dyspnoea on exertion	10
57903	CLASP shortness of breath score	10
374	Chest pain	11
544	[D]Chest pain, unspecified	11
726	Atypical chest pain	11
1059	Pleuritic pain	11
1270	[D]Chest tightness	11
2519	Rib pain	11
2584	[D]Chest pain	11
3518	[D]Musculoskeletal chest pain	11
3796	[D]Chest pain NOS	11
7346	Central chest pain	11
7844	[D]Non-cardiac chest pain	11
7878	[D]Chest discomfort	11
8349	Costal margin chest pain	11
9340	[D]Non cardiac chest pain	11
9698	Anterior chest wall pain	11
10370	Chest pain NOS	11
14823	[D]Anterior chest wall pain	11
15528	[D]Chest pressure	11
18134	Chest pain on exertion	11
18183	[D]Pleuritic pain	11
19199	[D]Parasternal chest pain	11
21082	[D]Retrosternal chest pain	11
24321	Chest wall tenderness	11
24704	Chest wall pain	11
29490	[D] Retrosternal chest pain	11
50477	[D]Central chest pain	11
53806	[X]Other chest pain	11
912	[D]Anorexia	12
1855	[D]Appetite loss	12
6607	Loss of appetite - symptom	12
7608	Appetite loss - anorexia	12
7743	Other and unspecified non-organic eating disorders	12
7744	Anorexia symptom	12
13081	Reduced appetite	12
17203	[X]Psychogenic loss of appetite	12

D. Symptom codes

Medcode	Description	Symptom code
22820	Non-organic loss of appetite	12
25189	Restricting food intake	12
53746	[D]Anorexia NOS	12
126	O/E - Underweight	13
654	Weight decreasing	13
3647	[D]Abnormal loss of weight	13
4663	Abnormal weight loss	13
5812	Abnormal weight loss - symptom	13
12398	Complaining of weight loss	13
12530	[D]Underweight	13
22005	O/E - cachexic	13
24068	[D] Cachexia	13
26473	O/E - weight > 20% below ideal	13
29029	O/E -weight 10-20% below ideal	13
32914	Body Mass Index low K/M2	13
37937	Weight loss from baseline weight	13
38995	Abnormally thin	13
53801	[D]Cachexia NOS	13
421	Iliac fossa pain	14
1181	Right iliac fossa pain	14
2781	C/O pelvic pain	14
2982	Left iliac fossa pain	14
4706	Bony pelvic pain	14
9061	[D]Left lower quadrant pain	14
9811	[d] pelvic pain	14
9920	[D]Pelvic and perineal pain	14
11647	O/E - abd. pain - R. iliac	14
16806	[D]Pain in right iliac fossa	14
16868	[D]Pain in left iliac fossa	14
17223	O/E - iliac pain on palpation	14
19360	[D]Right lower quadrant pain	14
21583	O/E - abd. pain - L. iliac	14
22608	Lower abdominal pain	14
35876	O/E - guarding - R.iliac	14
50662	[X]Pain localized to other parts of lower abdomen	14
56085	O/E - guarding - L.iliac	14

References

Abdel-Rahman, M., Stockton, D., Rachet, B., Hakulinen, T. & Coleman, M.P. (2009). What if cancer survival in Britain were the same as in Europe: how many deaths are avoidable? *British Journal of Cancer*, **101 Suppl**, S115–24. 17

Allensworth, S.K., Langstraat, C.L., Martin, J.R., Lemens, M.A., McGree, M.E., Weaver, A.L., Dowdy, S.C., Podratz, K.C. & Bakkum-Gamez, J.N. (2013). Evaluating the prognostic significance of preoperative thrombocytosis in epithelial ovarian cancer. *Gynecologic Oncology*, **130**, 499–504. 11, 142

Aoe, K., Hiraki, A., Ueoka, H., Kiura, K., Tabata, M., Tanaka, M. & Tanimoto, M. (2004). Thrombocytosis as a useful prognostic indicator in patients with lung cancer. *Respiration*, **71**, 170–173. 11

Arslan, C. & Coskun, H. (2005). Thrombocytosis in solid tumors: review of the literature. *Turkish Journal of Haematology*, **22**, 59–64. 49

Astin, M., Griffin, T., Neal, R.D., Rose, P. & Hamilton, W. (2011). The diagnostic value of symptoms for colorectal cancer in primary care: a systematic review. *The British Journal of General Practice*, **61**, e231–43. 151

Bain, B.J. (1996). Ethnic and sex differences in the total and differential white cell count and platelet count. *Journal of Clinical Pathology*, **49**, 664–6. 8, 138, 149

Balduini, C. & Noris, P. (2014). Platelet count and ageing. *Haematologica*, **99**, 953–955. 78

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**, 232–240. 21, 23, 150

Barrett, J., Sharp, D.J., Stapley, S., Stabb, C. & Hamilton, W. (2010). Pathways to the diagnosis of ovarian cancer in the UK: a cohort study in primary care. *BJOG*, **117**, 610–614. 32

Benchimol, E.I., Manuel, D.G., To, T., Griffiths, A.M., Rabeneck, L. & Guttmann, A. (2010). Development and use of reporting guidelines for assessing the quality of validation studies of health administrative data. *Journal of Clinical Epidemiology*. 124

Berrino, F., De Angelis, R., Sant, M., Rosso, S., Lasota, M.B., Coebergh, J.W. & Santaquilani, M. (2007). Survival for eight major cancers and all cancers combined for European adults diagnosed in 1995-99: results of the EUROCARE-4 study. *The Lancet Oncology*, **8**, 773–783. 16

Biino, G., Santimone, I., Minelli, C., Sorice, R., Frongia, B., Traglia, M., Ulivi, S., Di Castelnuovo, A., Gögele, M., Nutile, T., Francavilla, M., Sala, C., Pirastu, N., Cerletti, C., Iacoviello, L., Gasparini, P., Toniolo, D., Ciullo, M., Pramstaller, P., Pirastu, M., de Gaetano, G. & Balduini, C.L. (2013). Age- And Sex-Related Variations in Platelet Count in Italy: A Proposal of Reference Ranges Based on 40987 Subjects' Data. *PLoS ONE*, **8**. 8, 138, 149, 152

Blay, J.Y., Negrier, S., Combaret, V., Attali, S., Goillot, E., Merrouche, Y., Mercatello, A., Ravault, A., Tourani, J.M. & Moskovtchenko, J.F. (1992). Serum level of interleukin 6 as a prognosis factor in metastatic renal cell carcinoma. *Cancer Research*, **52**, 3317–22. 13

Bleeker, J.S. & Hogan, W.J. (2011). Thrombocytosis: diagnostic evaluation, thrombotic risk stratification, and risk-based management strategies. *Thrombosis*, **2011**, 536062. 10

Boggon, R., van Staa, T.P., Chapman, M., Gallagher, A.M., Hammad, T.A. & Richards, M.A. (2013). Cancer recording and mortality in the General Practice Research Database and linked cancer registries. *Pharmacoepidemiology Drug Safety*, **22**, 168–175. 25, 26, 28, 29, 50, 55, 119, 120, 134, 136, 146

Boneu, B., Bugat, R., Boneu, A., Eche, N., Sie, P. & Combes, P.F. (1984). Exhausted platelets in patients with malignant solid tumors without evidence of active consumption coagulopathy. *European Journal of Cancer and Clinical Oncology*, **20**, 899–903. 13

Booth, N. (1994). What are the Read Codes? *Health Libraries Review*, **11**, 177–82. 25, 147

Borsig, L., Wong, R., Feramisco, J., Nadeau, D.R., Varki, N.M. & Varki, A. (2001). Heparin and cancer revisited: mechanistic connections involving platelets,

P-selectin, carcinoma mucins, and tumor metastasis. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 3352–3357. 13

Buergy, D., Wenz, F., Groden, C. & Brockmann, M.A. (2012). Tumor-platelet interaction in solid tumors. *International Journal of Cancer*, **130**, 2747–2760. 11, 12, 13

Butkiewicz, A.M., Kemona-Chetnik, I., Dymicka-Piekarska, V., Matowicka-Karna, J., Kemona, H. & Radziwon, P. (2006). Does smoking affect thrombocytopoiesis and platelet activation in women and men? *Advances in Medical Sciences*, **51**, 123–126. 62

Butler, C.A., Darragh, K.M., Currie, G.P. & Anderson, W.J.A. (2006). Variation in lung cancer survival rates between countries: Do differences in data reporting contribute? *Respiratory Medicine*, **100**, 1642–1646. 17

Butler, J., Foot, C., Bomb, M., Hiom, S., Coleman, M., Bryant, H., Vedsted, P., Hanson, J. & Richards, M. (2013). The International Cancer Benchmarking Partnership: An international collaboration to inform cancer policy in Australia, Canada, Denmark, Norway, Sweden and the United Kingdom. *Health Policy*, **112**, 148–155. 16

Cancer Research UK (2012). Cancer incidence for common cancers. Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/common-cancers-compared, (Date accessed: 15/11/2016). 69

Cancer Research UK (2014a). Cancer Survival. Available at: http://publications.cancerresearchuk.org/downloads/Product/CS_REPORT_ SURVIVAL.pdf, (Date accessed: 21/05/2016). 16

Cancer Research UK (2014b). Stages of Cancer. Available at: http://www.cancerresearchuk.org/about-cancer/what-is-cancer/stages-of-cancer, (Date accessed: 03/04/2016). 14, 15

Cancer Research UK (2014c). What is the HPV virus? Available at: http://www.cancerresearchuk.org/about-cancer/cancers-in-general/cancer-questions/what-is-the-hpv-virus, (Date accessed: 21/11/2016). 143

Cancer Research UK (2015a). Blood tests. Available at: http://www.cancerresearchuk.org/about-cancer/cancers-in-general/tests/blood-tests, (Date accessed: 13/05/2016). 9

Cancer Research UK (2015b). Cervival cancer incidence statistics. Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/cervical-cancer/survival, (Date accessed: 13/06/2016). 143

Cancer Research UK (2016). Testicular cancer incidence statistics. Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/testicular-cancer, (Date accessed: 13/06/2016). 150

Centre for Reviews and Dissemination (2009). *Systematic reviews: CRD's guidance for undertaking reviews in health care*. 36

Coleman, M.P., Forman, D., Bryant, H., Butler, J., Rachet, B., Maringe, C., Nur, U., Tracey, E., Coory, M., Hatcher, J., McGahan, C.E., Turner, D., Marrett, L., Gjerstorff, M.L., Johannesen, T.B., Adolfsson, J., Lambe, M., Lawrence, G., Meechan, D., Morris, E.J., Middleton, R., Steward, J. & Richards, M.A. (2011). Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the international cancer benchmarking partnership): An analysis of population-based cancer registry data. *The Lancet*, **377**, 127–138. 16

Crawford, S., Davis, J., Siddiqui, N., de Caestecker, L., Gillis, C., Hole, D. & Penney, G. (2002). The waiting time paradox: population based retrospective study of treatment delay and survival of women with endometrial cancer in Scotland. *BMJ*, **325**. 21, 22

Daly, M. (2011). Determinants of platelet count in humans. *Haematologica*, **96**, 10–13. 7, 8

Davis, W.M. & Mendez Ross, A.O. (1972). Thrombocytosis and Thrombocythemia: The Laboratory and Clinical Significance of an Elevated Platelet Count. *American Journal of Clinical Pathology*, **59**, 243–247. 11, 140, 141, 142

De Angelis, R., Sant, M., Coleman, M.P., Francisci, S., Baili, P., Pierannunzio, D., Trama, A., Visser, O., Brenner, H., Ardanaz, E., Bielska-Lasota, M., Engholm, G., Nennecke, A., Siesling, S., Berrino, F. & Capocaccia, R. (2013). Cancer survival in Europe 1999-2007 by country and age: results of EUROCARE-5-a population-based study. *The Lancet Oncology*, **2045**, 1–12. 16

De Vita, F., Romano, C., Orditura, M., Galizia, G., Martinelli, E., Lieto, E. & Catalano, G. (2001). Interleukin-6 Serum Level Correlates with Survival in

Advanced Gastrointestinal Cancer Patients but Is Not an Independent Prognostic Indicator. *Journal of Interferon and Cytokine Research*, **21**, 45–52. 13

Deeks, J.J., Higgins, J.P. & Altman, D.G. (2008). Analysing Data and Undertaking Meta-Analyses. In *Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series*, 243–296. 33

Department of Health (2011). Improving outcomes: a strategy for cancer. 19

Deutsch, V.R. & Tomer, A. (2006). Megakaryocyte development and platelet production. *British Journal of Haematology*, **134**, 453–466. 8, 12

Dregan, A., Moller, H., Murray-Thomas, T. & Gulliford, M.C. (2012). Validity of cancer diagnosis in a primary care database compared with linked cancer registrations in England. Population-based cohort study. *Cancer Epidemiology*, **36**, 425–429. 27, 28, 41, 50, 119, 120, 134, 135, 136, 146, 147, 151

Forbes, L.J.L., Simon, A.E., Warburton, F., Boniface, D., Brain, K.E., Dessaix, A., Donnelly, C., Haynes, K., Hvidberg, L., Lagerlund, M., Lockwood, G., Tishelman, C., Vedsted, P., Vigmostad, M.N., Ramirez, A.J. & Wardle, J. (2013). Differences in cancer awareness and beliefs between Australia, Canada, Denmark, Norway, Sweden and the UK (the International Cancer Benchmarking Partnership): do they contribute to differences in cancer survival? *British Journal of Cancer*, **108**, 292–300. 17

Gader, A., Bahakim, H., Awadalla, S. & Malaika, S. (1995). Ethnic variations in the haemostatic system: comparison between Arabs, Westerners (Europeans and Americans), Asians and Africans. *Blood Coagulation and Fibinolysis*, **6**, 537–542. 138, 149

Gao, S.P., Mark, K.G., Leslie, K., Pao, W., Motoi, N., Gerald, W.L., Travis, W.D., Bornmann, W., Veach, D., Clarkson, B. & Bromberg, J.F. (2007). Mutations in the EGFR kinase domain mediate STAT3 activation via IL-6 production in human lung adenocarcinomas. *The Journal of Clinical Investigation*, **117**, 3846–56. 9

Gay, L.J. & Felding-Habermann, B. (2011). Contribution of platelets to tumour metastasis. *Nature Reviews Cancer*, **11**, 123–34. 8, 13

Geddis, A.E. (2009). The regulation of proplatelet production. *Haematologica*, **94**, 756–759. 8

Giles, C. (1981). The platelet count and mean platelet volume. *British Journal of Haematology*, **48**, 31–7. 8

Graham, S., Traub, B. & Mink, I. (1987). Automated Platelet-Sizing Parameters on a Normal Population. *American Journal of Clinical Pathology*, **87**, 365–369. 138, 149

Green, M.S., Peled, I. & Najenson, T. (1992). Gender differences in platelet count and its association with cigarette smoking in a large cohort in Israel. *Journal of Clinical Epidemiology*, **45**, 77–84. 62

Griesshammer, M., Bangerter, M., Sauer, T., Wennauer, R., Bergmann, L. & Heimpel, H. (1999). Aetiology and clinical significance of thrombocytosis: Analysis of 732 patients with an elevated platelet count. *Journal of Internal Medicine*, **245**, 295–300. 9, 56

Grubb, R.L., Pinsky, P.F., Greenlee, R.T., Izmirlian, G., Miller, A.B., Hickey, T.P., Riley, T.L., Mabie, J.E., Levin, D.L., Chia, D., Kramer, B.S., Reding, D.J., Church, T.R., Yokochi, L.A., Kvale, P.A., Weissfeld, J.L., Urban, D.A., Buys, S.S., Gelmann, E.P., Ragard, L.R., Crawford, E.D., Prorok, P.C., Gohagan, J.K., Berg, C.D. & Andriole, G.L. (2008). Prostate cancer screening in the Prostate, Lung, Colorectal and Ovarian cancer screening trial: Update on findings from the initial four rounds of screening in a randomized trial. *BJU International*, **102**, 1524–1530. 151

Gücer, F., Moser, F., Tamussino, K., Reich, O., Haas, J., Arikan, G., Petru, E. & Winter, R. (1998). Thrombocytosis as a Prognostic Factor in Endometrial Carcinoma. *Gynecologic Oncology*, **70**, 210–214. 11

Hamilton, F., Carroll, R., Hamilton, W. & Salisbury, C. (2014). The risk of cancer in primary care patients with hypercalcaemia: a cohort study using electronic records. *British Journal of Cancer*, **111**, 1410–1412. 150, 151, 152

Hamilton, W. (2009). The CAPER studies: five case-control studies aimed at identifying and quantifying the risk of cancer in symptomatic primary care patients. *British Journal of Cancer*, **101 Suppl**, S80–6. 10

Hamilton, W. (2010). Cancer diagnosis in primary care. *British Journal of General Practice*, **60**, 121–128. 31

Hamilton, W., Peters, T.J., Round, A. & Sharp, D. (2005a). What are the clinical features of lung cancer before the diagnosis is made? A population based

case-control study. *Thorax*, **60**, 1059–1065. 9, 10, 32, 38, 41, 43, 44, 46, 47, 106, 107, 141, 148, 151

Hamilton, W., Round, A., Sharp, D. & Peters, T. (2005b). Clinical features of colorectal cancer before diagnosis: a population-based case-control study. *British Journal of Cancer*, **93**, 399–405. 10, 32, 38, 41, 43, 46, 47, 106, 107, 108

Hamilton, W., Lancashire, R., Sharp, D., Peters, T.J., Cheng, K.K. & Marshall, T. (2008). The importance of anaemia in diagnosing colorectal cancer: a casecontrol study using electronic primary care records. *British Journal of Cancer*, **98**, 323–327. 117, 137, 152

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. 38, 41, 43, 46, 47

Hauser, C.A., Stockler, M. & Tattersall, M. (2006). Prognostic factors in patients with recently diagnosed incurable cancer: a systematic review. *Support Cancer Care*, **14**, 999–1011. 10

Herrett, E., Thomas, S.L., Schoonen, W.M., Smeeth, L. & Hall, A.J. (2010). Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *British Journal of Clinical Pharmacology*, **69**, 4–14. 25

Herter, J.M., Rossaint, J. & Zarbock, A. (2014). Platelets in inflammation and immunity. *Journal of Thrombosis and Haemostasis*, **12**, 1764–1775. 8

Higgins, J.P.T., Thompson, S.G., Deeks, J.J. & Altman, D.G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, **327**, 557–560. 33

Hiom, S.C. (2015). Diagnosing cancer earlier: reviewing the evidence for improving cancer survival. *British Journal of Cancer*, **112**, S1–S5. 19, 20

Hippisley-Cox, J. & Coupland, C. (2013a). Symptoms and risk factors to identify men with suspected cancer in primary care: Derivation and validation of an algorithm. *British Journal of General Practice*, **63**. 37

Hippisley-Cox, J. & Coupland, C. (2013b). Symptoms and risk factors to identify women with suspected cancer in primary care: Derivation and validation of an algorithm. *British Journal of General Practice*, **63**. 37

Hobbs, F.D.R., Bankhead, C., Mukhtar, T., Stevens, S., Perera-Salazar, R., Holt, T. & Salisbury, C. (2016). Clinical workload in UK primary care: a

retrospective analysis of 100 million consultations in England, 2007-14. *The Lancet*, 1–8. 18

Holmes, C.E., Ramos-Nino, M.E. & Littenberg, B. (2010). An association between anti-platelet drug use and reduced cancer prevalence in diabetic patients: results from the Vermont Diabetes Information System Study. *BMC Cancer*, **10**. 12

Hong, D.S., Angelo, L.S. & Kurzrock, R. (2007). Interleukin-6 and its receptor in cancer: Implications for translational therapeutics. *Cancer*, **110**, 1911–1928. 9, 13

Ikeda, M., Furukawa, H., Imamura, H., Shimizu, J., Ishida, H., Masutani, S., Tatsuta, M. & Satomi, T. (2002). Poor prognosis associated with thrombocytosis in patients with gastric cancer. *Annals of Surgical Oncology*, **9**, 287–91. 11

Information Services Division Scotland (2014). Cancer statistics. 14, 15

Insicive Health (2014). Saving lives, averting costs. Tech. Rep. September. 21, 23, 32

Jain, S., Harris, J. & Ware, J. (2010). Platelets: Linking hemostasis and cancer. *Arteriosclerosis, Thrombosis, and Vascular Biology*, **30**, 2362–2367. 12

Kaser, A., Brandacher, G., Steurer, W., Kaser, S., Offner, F.A., Zoller, H., Theurl, I., Widder, W., Molnar, C., Ludwiczek, O., Atkins, M.B., Mier, J.W. & Tilg, H. (2001). Interleukin-6 stimulates thrombopoiesis through thrombopoietin: Role in inflammatory thrombocytosis. *Blood*, **98**, 2720–2725. 13

Kaushansky, K. (2005). The molecular mechanisms that control thrombopoiesis. *The Journal of Clinical Investigation*, **115**, 3339–47. 9

Khan, N.F., Harrison, S.E. & Rose, P.W. (2010). Validity of diagnostic coding within the General Practice Research Database: a systematic review. *The British Journal of General Practice*, **60**, e128–36. 25, 26, 50, 120

Khan, P.N., Nair, R.J., Olivares, J., Tingle, L.E. & Li, Z. (2009). Postsplenectomy reactive thrombocytosis. *Proceedings (Baylor University Medical Centre)*, **22**, 9–12. 10

Knottnerus, J.A. (2002). *The evidence base of clinical diagnosis*. BMJ Books, London. 36

Kuter, D.J. (1996). The physiology of platelet production. *Stem Cells*, **14 Suppl 1**, 88–101. 8

Levin, J. & Conley, C. (1964). Thrombocytosis is associated with malignant disease. *Archives of Internal Medicine*, **114**, 497–500. 10

Lyratzopoulos, G., Abel, G., McPhail, S., Neal, R.D. & Rubin, G.P. (2013). Measures of promptness of cancer diagnosis in primary care: secondary analysis of national audit data on patients with 18 common and rarer cancers. *British Journal of Cancer*, **108**, 1–5. 24

Lyratzopoulos, G., Vedsted, P. & Singh, H. (2015). Understanding missed opportunities for more timely diagnosis of cancer in symptomatic patients after presentation. *British Journal of Cancer*, **112 Suppl**, S84–91. 24

Machlus, K.R. & Italiano, J.E. (2013). The incredible journey: From megakaryocyte development to platelet formation. *Journal of Cell Biology*, **201**, 785–796. 8

Macleod, U., Mitchell, E.D., Burgess, C., Macdonald, S. & Ramirez, a.J. (2009). Risk factors for delayed presentation and referral of symptomatic cancer: evidence for common cancers. *British Journal of Cancer*, **101 Suppl**, S92–S101. 137

Mantovani, A., Mantovani, A., Allavena, P., Allavena, P., Sica, A., Sica, A., Balkwill, F. & Balkwill, F. (2008). Cancer-related inflammation. *Nature*, **454**, 436–44. 8

Marcus, P.M., Prorok, P.C., Miller, A.B., DeVoto, E.J. & Kramer, B.S. (2015). Conceptualizing Overdiagnosis in Cancer Screening. *Journal of the National Cancer Institute*, **107**, 1–4. 21

Maringe, C., Walters, S., Butler, J., Coleman, M.P., Hacker, N., Hanna, L., Mosgaard, B.J., Nordin, A., Rosen, B., Engholm, G., Gjerstorff, M.L., Hatcher, J., Johannesen, T.B., McGahan, C.E., Meechan, D., Middleton, R., Tracey, E., Turner, D., Richards, M.A. & Rachet, B. (2012). Stage at diagnosis and ovarian cancer survival: Evidence from the international cancer benchmarking partnership. *Gynecologic Oncology*, **127**, 75–82. 17

Maringe, C., Walters, S., Rachet, B., Butler, J., Fields, T., Finan, P., Maxwell, R., Nedrebø, B., Påhlman, L., Sjövall, A., Spigelman, A., Engholm, G., Gavin, A., Gjerstorff, M.L., Hatcher, J., Johannesen, T.B., Morris, E., McGahan, C.E., Tracey, E., Turner, D., Richards, M.a. & Coleman, M.P. (2013). Stage at diagnosis and colorectal cancer survival in six high-income countries: a population-based study of patients diagnosed during 2000-2007. *Acta Oncologica*, **52**, 919–32. 17

McGlynn, E.A., McDonald, K.M. & Cassel, C.K. (2015). Measurement Is Essential for Improving Diagnosis and Reducing Diagnostic Error: a report from the Institute of Medicine. *JAMA*, **314**, 2501–2502. 24

Moher, D., Pham, B., Lawson, M.L. & Klassen, T.P. (2003). The inclusion of reports of randomised trials published in languages other than English in systematic reviews. *Health Technology Assessment*, **7**, 1–90. 50

Møller, H., Gildea, C., Meechan, D., Rubin, G., Round, T. & Vedsted, P. (2015). Use of the English urgent referral pathway for suspected cancer and mortality in patients with cancer: cohort study. *BMJ*, h5102. 21

Nakashima, J., Tachibana, M., Horiguchi, Y., Oya, M., Ohigashi, T., Asakura, H. & Murai, M. (2000). Serum interleukin 6 as a prognostic factor in patients with prostate cancer. *Clinical Cancer Research*, **6**, 2702–6. 13

Nash, G.F., Turner, L.F., Scully, M.F. & Kakkar, A.K. (2002). Platelets and cancer. *The Lancet Oncology*, **3**, 425–430. 14, 49

National Cancer Intelligence Network (2010). Routes to Diagnosis - NCIN Data Briefing. 143

Neal, R.D., Tharmanathan, P., France, B., Din, N.U., Cotton, S., Fallon-Ferguson, J., Hamilton, W., Hendry, A., Hendry, M., Lewis, R., Macleod, U., Mitchell, E.D., Pickett, M., Rai, T., Shaw, K., Stuart, N., Torring, M.L., Wilkinson, C., Williams, B., Williams, N. & Emery, J. (2015). Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *British Journal of Cancer*, **112 Suppl**, S92–107. 22

NHS England (2016). New cancer ratings published. Available at: https://www.england.nhs.uk/cancer/2016/10/04/cancer-ratings/, (Date accessed: 02/11/2016). 148

NICE (2015). Suspected cancer: recognition and referral [NG12]. 18, 32, 55, 71, 79, 108, 109, 117, 142, 150, 151

Njølstad, T.S., Engerud, H., Werner, H.M.J., Salvesen, H.B. & Trovik, J. (2013). Preoperative anemia, leukocytosis and thrombocytosis identify aggressive endometrial carcinomas. *Gynecologic Oncology*, **131**, 410–415. 11

Office of National Statistics (2015). What are the top causes of death by age and gender? Available at: http://visual.ons.gov.uk/what-are-the-top-causes-of-death-by-age-and-gender/, (Date accessed: 24/05/2016). 15

Offner, F., Obrist, P., Stadlmann, S., Feichtinger, H., Klingler, P., Herold, M., Zwierzina, H., Hittmair, A., Mikuz, G. & Abendstein, B. (1995). IL-6 secretion by human peritoneal mesothelial and ovarian cancer cells. *Cytokine*, **7**, 542–7. 9

Pedersen, L.M. & Milman, N. (1996). Prognostic significance of thrombocytosis in patients with primary lung cancer. *The European Respiratory Journal*, **9**, 1826–1830. 11, 140, 141

Plante, M., Rubin, S.C., Wong, G.Y., Federici, M.G., Finstad, C.L. & Gastl, G.A. (1994). Interleukin-6 level in serum and ascites as a prognostic factor in patients with epithelial ovarian cancer. *Cancer*, **73**, 1882–8. 13

Price, S.J., Stapley, S., Shephard, E.A., Barraclough, K. & Hamilton, W. (2016). Is omission of free text records a possible source of data loss and bias in Clinical Practice Research Datalink studies? A case control study. *BMJ Open*, **6**, 1–7. 147

Quaresma, M., Coleman, M.P. & Rachet, B. (2015). 40-year trends in an index of survival for all cancers combined and survival adjusted for age and sex for each cancer in England and Wales, 19712011: a population-based study. *The Lancet*, **385**, 1206–1218. 16

Rashbass, J. (2014). Cancer registration as good as the world's best. *Public Health Matters blog*. 29

RevMan (2014). Review Manager (RevMan) [Computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014. 37

Richards, M. (2009). The size of the prize for earlier diagnosis of cancer in England. *British Journal of Cancer*, **101 Suppl**, S125–S129. 17, 19, 31

Rose, P.W., Rubin, G., Perera-Salazar, R., Almberg, S.S., Barisic, A., Dawes, M., Grunfeld, E., Hart, N., Neal, R.D., Pirotta, M., Sisler, J., Konrad, G., Toftegaard, B.S., Thulesius, H., Vedsted, P., Young, J. & Hamilton, W. (2015). Explaining variation in cancer survival between 11 jurisdictions in the International Cancer Benchmarking Partnership: a primary care vignette survey. *BMJ Open*, **5**, e007212. 17, 21

Rupassara, K.S., Ponnusamy, S., Withanage, N. & Milewski, P.J. (2006). A paradox explained? Patients with delayed diagnosis of symptomatic colorectal cancer have good prognosis. *Colorectal Disease*, **8**, 423–429. 22

Sansone, P., Storci, G., Tavolari, S., Guarnieri, T., Giovannini, C., Taffurelli, M., Ceccarelli, C., Santini, D., Paterini, P., Marcu, K.B., Chieco, P. & Bonafè, M. (2007). IL-6 triggers malignant features in mammospheres from human ductal breast carcinoma and normal mammary gland. *Journal of Clinical Investigation*, **117**, 3988–4002. 9

Schafer, A.I. (2002). Thrombocytosis: too much of a good thing? *Transactions of the American Clinical and Climatological Association*, **113**, 67–68. 9

Schafer, Z.T. & Brugge, J.S. (2007). IL-6 involvement in epithelial cancers. *Journal of Clinical Investigation*, **117**, 3660–3663. 9

Scheller, J., Chalaris, A., Schmidt-Arras, D. & Rose-John, S. (2011). The pro- and anti-inflammatory properties of the cytokine interleukin-6. *Biochimica et Biophysica Acta - Molecular Cell Research*, **1813**, 878–888. 13

Segal, J. & Moliterno, A. (2006). Platelet counts differ by sex, ethnicity, and age in the United States. *Annals of Epidemiology*, **16**, 123–130. 138, 149

Shephard, E., Stapley, S., Neal, R.D., Rose, P., Walter, F.M. & Hamilton, W. (2012). Clinical features of bladder cancer in primary care. *British Journal of General Practice*, **62**, e598–e604. 10, 38, 41, 43, 46

Shephard, E., Neal, R., Rose, P., Walter, F. & Hamilton, W. (2013). Clinical features of kidney cancer in primary care: a case-control study using primary care records. *British Journal of General Practice*, **63**, e250–5. 10, 32, 38, 41, 43, 44, 46, 151

Sloan, A., Gona, P. & Johnson, A.D. (2015). Cardiovascular correlates of platelet count and volume in the Framingham Heart Study. *Annals of Epidemiology*, **25**, 492–498. 62

Stapley, S., Peters, T.J., Sharp, D. & Hamilton, W. (2006). The mortality of colorectal cancer in relation to the initial symptom at presentation to primary care and to the duration of symptoms: a cohort study using medical records. *British Journal of Cancer*, **95**, 1321–1325. 32

Stapley, S., Peters, T.J., Neal, R.D., Rose, P.W., Walter, F.M. & Hamilton, W. (2012). The risk of pancreatic cancer in symptomatic patients in primary

care: a large case-control study using electronic records. *British Journal of Cancer*, **106**, 1940–1944. 10, 32, 38, 41, 43, 46

Stapley, S., Peters, T.J., Neal, R.D., Rose, P.W., Walter, F.M. & Hamilton, W. (2013). The risk of oesophago-gastric cancer in symptomatic patients in primary care: a large case-control study using electronic records. *British Journal of Cancer*, **108**, 25–31. 10, 32, 38, 41, 43, 44, 46, 151

Sterne, J.A.C., Sutton, A.J., Ioannidis, J.P.A., Terrin, N., Jones, D.R., Lau, J., Carpenter, J., Rücker, G., Harbord, R.M., Schmid, C.H., Tetzlaff, J., Deeks, J.J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D.G., Moher, D. & Higgins, J.P.T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ (Clinical research ed.)*, **343**, d4002. 51

Stevens, R. & Alexander, M. (1977). A sex difference in the platelet count. *British Journal of Haematology*, **37**, 295–300. 138, 149

Stewart, A.F. (2005). Hypercalcemia associated with cancer. *The New England Journal of Medicine*, **352**, 373–379. 152

Stone, R.L., Nick, A.M., McNeish, I.a., Balkwill, F., Han, H.D., Bottsford-Miller, J., Rupaimoole, R., Armaiz-Pena, G.N., Pecot, C.V., Coward, J., Deavers, M.T., Vasquez, H.G., Urbauer, D., Landen, C.N., Hu, W., Gershenson, H., Matsuo, K., Shahzad, M.M., King, E.R., Tekedereli, I., Ozpolat, B., Ahn, E.H., Bond, V.K., Wang, R., Drew, A.F., Gushiken, F., Lamkin, D., Collins, K., DeGeest, K., Lutgendorf, S.K., Chiu, W., Lopez-Berestein, G., Afshar-Kharghan, V. & Sood, A.K. (2012). Paraneoplastic Thrombocytosis in Ovarian Cancer. *New England Journal of Medicine*, **366**, 610–618. 10, 11, 140, 142

Suppiah, R., Shaheen, P., Elson, P., Misbah, S., Wood, L., Motzer, R., Negrier, S., Andresen, S. & Bukowski, R. (2006). Thrombocytosis as a prognostic factor for survival in patients with metastatic renal cell carcinoma. *Cancer*, **107**, 1793–1800. 11

Suwansaksri, J., Wiwanitkit, V. & Soogarun, S. (2004). Effect of smoking on platelet count and platelet parameters: an observation. *Clinical and applied thrombosis/hemostasis : official journal of the International Academy of Clinical and Applied Thrombosis/Hemostasis*, **10**, 287–8. 62

Syed, N., Usman, M. & Khurshid, M. (2007). Thrombocytosis: age dependent aetiology and analysis of platelet indices for differential diagnosis. *Indian Journal of Pathology and Microbiology*, **50**, 628–633. 56

Takeuchi, E., Ito, M., Mori, M., Yamaguchi, T., Nakagawa, M., Yokota, S., Nishikawa, H., Sakuma-Mochizuki, J., Hayashi, S. & Ogura, T. (1996). Lung cancer producing interleukin-6. *Internal Medicine*, **35**, 212–214. 13

Thomson, C.S. & Forman, D. (2009). Cancer survival in England and the influence of early diagnosis: what can we learn from recent EUROCARE results? *British Journal of Cancer*, **101 Suppl**, S102–9. 16

Todenhöfer, T., Renninger, M., Schwentner, C., Stenzl, A. & Gakis, G. (2012). A new prognostic model for cancer-specific survival after radical cystectomy including pretreatment thrombocytosis and standard pathological risk factors. *BJU International*, **110**, e533–40. 11

Voutsadakis, I.A. (2014). Thrombocytosis as a prognostic marker in gastrointestinal cancers. *World Journal of Gastrointestinal Oncology*, **6**, 34–40. 11

Waldner, M.J., Foersch, S. & Neurath, M.F. (2012). Interleukin-6 - A key regulator of colorectal cancer development. *International Journal of Biological Sciences*, **8**, 1248–1253. 9

Walker, S., Hyde, C. & Hamilton, W. (2013). Risk of uterine cancer in symptomatic women in primary care: Case-control study using electronic records. *British Journal of General Practice*, **63**, e643–e648. 10, 38, 42, 43, 44, 46, 151

Walker, S., Hyde, C. & Hamilton, W. (2014). Risk of breast cancer in symptomatic women in primary care: a case control study using primary care records. *British Journal of General Practice*, **64**, e788–93. 10, 32, 38, 41, 42, 43, 46, 143, 150

Walters, S., Maringe, C., Butler, J., Rachet, B., Barrett-Lee, P., Bergh, J., Boyages, J., Christiansen, P., Lee, M., Wärnberg, F., Allemani, C., Engholm, G., Fornander, T., Gjerstorff, M.L., Johannesen, T.B., Lawrence, G., McGahan, C.E., Middleton, R., Steward, J., Tracey, E., Turner, D., Richards, M.A. & Coleman, M.P. (2013a). Breast cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK, 2000-2007: a population-based study. *British Journal of Cancer*, **108**, 1195–208. 17

Walters, S., Maringe, C., Coleman, M.P., Peake, M.D., Butler, J., Young, N., Bergström, S., Hanna, L., Jakobsen, E., Kölbeck, K., Sundstrøm,

S., Engholm, G., Gavin, A., Gjerstorff, M.L., Hatcher, J., Johannesen, T.B., Linklater, K.M., McGahan, C.E., Steward, J., Tracey, E., Turner, D., Richards, M.a. & Rachet, B. (2013b). Lung cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK: a population-based study, 2004-2007. *Thorax*, **68**, 551–64. 17

Walters, S., Benitez-Majano, S., Muller, P., Coleman, M.P., Allemani, C., Butler, J., Peake, M., Guren, M.G., Glimelius, B., Bergström, S., Påhlman, L. & Rachet, B. (2015). Is England closing the international gap in cancer survival? *British Journal of Cancer*, **113**, 848–60. 16, 17

Wang, Y., Hunt, K., Nazareth, I., Freemantle, N. & Petersen, I. (2013). Do men consult less than women? An analysis of routinely collected UK general practice data. *BMJ Open*, **3**, e003320. 18

Weller, D., Vedsted, P., Rubin, G., Walter, F.M., Emery, J., Scott, S., Campbell, C., Andersen, R.S., Hamilton, W., Olesen, F., Rose, P., Nafees, S., van Rijswijk, E., Hiom, S., Muth, C., Beyer, M. & Neal, R.D. (2012). The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. *British Journal of Cancer*, **106**, 1262–7. 19, 23

Welsh Cancer Intelligence and Surveillance Unit (2014). Cancer in Wales. 14, 15

Whiting, P.F., Rutjes, A.W., Westwood, M.E., Mallett, S., Deeks, J.J., Reitsma, J.B., Leeflang, M.M., Sterne, J.A. & Bossuyt, P.M. (2011). QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, **155**, 529–536. 35

Williams, T., van Staa, T., Puri, S. & Eaton, S. (2012). Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. *Therapeutic Advances in Drug Safety*, **3**, 89–99. 25

World Health Organisation (2012). ICD 10. 63

Xiao da, W., Yang, M., Yang, J., Hon, K.L. & Fok, F.T. (2006). Lung damage may induce thrombocytopenia. *Platelets*, **17**, 347–349. 8, 9

Yu, C., Huang, B. & Zhou, F.L. (2015). Correlation between bone metastasis and thrombocytosis in pulmonary adenocarcinoma patients. *Oncology Letters*, **9**, 762–768. 13

References

Zucker-Franklin, D. & Philipp, C.S. (2000). Platelet Production in the Pulmonary Capillary Bed. *The American Journal of Pathology*, **157**, 69–74. 8