The presenting symptoms of cancer patients and associations with diagnostic timeliness

Minjoung Monica Koo

Thesis for the degree of Doctor of Philosophy in Epidemiology and Public Health

UNIVERSITY COLLEGE LONDON

Declaration

I, Minjoung Monica Koo, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, this has been indicated in the thesis.

Acknowledgements

I'd like to acknowledge and thank:

My principal supervisor Yoryos – to whom I owe much of my professional development; a sincere thank you for your tireless mentorship and guidance, and for always having the time to talk.

Christian – thank you for adopting me as a supervisee and for always being there to provide a fresh perspective and reassurance. Prof Greg Rubin, Dr Gary Abel, and Dr Sean McPhail, all of whom have (perhaps unwittingly!) played significant roles in my doctoral training: thank you for sharing your invaluable insight, experience, and time.

Members (past and present) of room 208 and colleagues at BSH – you know who you are! Thank you for your support and humour which has kept me grounded and cheered me on in equal measure.

My parents and my sister Maria – an enormous thank you for your never-ending support in everything that I do, your tolerance of increasingly sporadic FaceTime updates, and your wisdom and encouragement that spans time zones and distance: this thesis is for you.

이 논문은 엄마 아빠에게 바칩니다. 언제나 응원해주고, 위로해주고, 힘이 되어준 엄마 아빠 – 진심으로 고마워요.

Finally, to Liam – who has been subjected to considerable collateral damage in the past 3 years; for your patience, your compromises, your relentless encouragement, and your confidence in me and my work that often exceeds my own: thank you.

Statement about funders

I would like to acknowledge the studentship that supported my doctoral training from the Policy Research Unit in Cancer Awareness, Screening, and Early Diagnosis. The Policy Research Unit in Cancer Awareness, Screening, and Early Diagnosis is a collaboration between researchers from seven institutions (Queen Mary University of London, University College London, King's College London, London School of Hygiene and Tropical Medicine, Hull York Medical School, Durham University and University of Exeter), and receives funding from the UK Department of Health Policy Research Programme.

Abstract

Diagnosing cancer earlier is an important strand of cancer control. Interventions promoting early diagnosis such as awareness campaigns and fast-track clinical pathways are increasingly commonplace in England and other countries, but their theoretical underpinning is limited. Cancer symptoms are critical components of such interventions, but evidence regarding the presenting symptoms of individuals diagnosed with cancer and measures of diagnostic timeliness remains sparse. I sought to address this evidential gap using data from the first English National Audit of Cancer Diagnosis in Primary Care on a large and representative cohort of cancer patients.

Using symptom status (ascertained through the coding of free-text information on presenting symptoms), I identified and described a group of atypically diagnosed cancer patients before proceeding to examine the nature and frequency of presenting symptoms and associated diagnostic timeliness among symptomatic cancer patients.

I profiled the broad range of presenting symptoms beyond breast lump among women diagnosed with breast cancer, and found that non-lump symptoms were associated with longer intervals to presentation and referral.

I also described variation in diagnostic timeliness among cancer patients who presented with abdominal symptoms, and the case-mix of cancers they were subsequently diagnosed with. The relative length of time to presentation and referral varied by abdominal symptom. A considerable proportion of cancer patients who presented with abdominal symptoms were diagnosed with other solid tumours or haematological cancers, particularly for non-specific abdominal symptoms.

Lastly, I examined the association between alarm symptoms and stage at diagnosis among cancer patients who had presented promptly. While most alarm symptoms at presentation were associated with early stage disease, the extent of the association was highly variable compared to patients with other symptomatic presentations.

This thesis provides exemplar evidence regarding the epidemiology of presenting symptoms among cancer patients and associated diagnostic timeliness. Together with other evidence, the findings could contribute to the design and evaluation of symptom awareness campaigns and healthcare interventions that expedite the investigation and diagnosis of cancer.

Outline

1. Introduction
2. Aims and objectives
3. Literature review
4. Data & methods
5. Atypical diagnosis of cancer
6. The symptom signature of breast cancer
7. Abdominal symptoms and time to presentation
8. Abdominal symptoms and time to referral
9. Cancer alarm symptoms and stage at diagnosis
10. Discussion

Table of Contents

List of Figures

List of Tables

Acronyms and key concepts

2-week-wait	The fast-track clinical pathways designed for expediting the investigation and diagnosis of cancer in England; also see page 29
ACE	Accelerate, Coordinate, Evaluate programme in England
Alarm symptom	A symptom that has relatively high predictive value for cancer Clinical guidelines for the referral of suspected cancer in England have adopted an explicit PPV threshold of 3% or greater since 2015 to classify symptoms as mandating fast-track referral or investigation so this has been used as a reference point throughout the thesis; also see page 44
BCOC	Be Clear on Cancer
CI	Confidence interval
CIBH	Change in bowel habit
Early diagnosis	Early diagnosis of cancer
GP	General Practitioner; i.e. a primary care physician
IQR	Inter-quartile range
NACDPC	National Audit of Cancer Diagnosis in Primary Care (first audit of cancer in primary care in England)
NAEDI	National Awareness and Early Diagnosis Initiative
NCDA	National Cancer Diagnosis Audit (second audit of cancer in primary care in England)
NCIN	former National Cancer Intelligence Network in England
NICE	National Institute for Health and Care Excellence
ONS	Office for National Statistics
PMB	Post-menopausal bleeding
PPV	Positive predictive value
RCGP	Royal College of General Practitioners
SEER LRD	Staging system (local-regional-distant disease) used by the Surveillance, Epidemiology, and End Results (SEER) Program
TNM	Classification system for malignant tumours based T (tumour), N (lymph node involvement), and M (metastasis)
WHO	World Health Organisation

Chapter 1: Introduction

In this chapter, I describe the importance of early diagnosis of cancer, how it may be conceptualised, and current early diagnosis policy initiatives, focusing on two major complex interventions in England. Based on these considerations, I outline the relevance of the epidemiology of presenting cancer symptoms to early diagnosis interventions.

1 Introduction

1.1 The importance of early diagnosis in cancer

The disease burden caused by cancer is a major global health challenge: the number of new cases is expected to increase by 70% over the next twenty years, representing a leading cause of mortality worldwide (World Health Organisation, 2017). Cancer control strategies comprising prevention, early diagnosis, and treatment have been developed to address the associated morbidity and mortality (Olesen *et al*, 2009; Cancer Australia, 2014; NHS England, 2016; WHO, 2017).

Among these, the earlier detection and diagnosis of symptomatic cancer forms an important strand, as the majority of patients are diagnosed after developing symptoms (Elliss-Brookes *et al*, 2012; Jensen *et al*, 2014). Acknowledgement of the potential importance of early diagnosis is not new: one of the earliest mentions in MEDLINE describes it as the greatest among the "triad" of tools to prevent mortality from cancer, alongside surgical treatment ('operation') and radiotherapy ('irradiation'):

> "...We must still, perforce, rely upon already existing clinical knowledge – diagnosis, operation, irradiation – the indispensable triad. Of these three the greatest is, "diagnosis"; but this is only fully effective when it is ***early***."
>
> – AG Nicholls, editor of the Canadian Medical Association Journal, writing in the November issue of the journal in 1933[1]

The benefits of earlier diagnosis of cancer have long been examined and debated. A systematic review of the associations between intervals to diagnosis and clinical outcomes across all cancers identified substantial variation in the availability and quality of evidence, though the authors concluded that expediting diagnosis was likely to be beneficial (Neal *et al*, 2015). Empirically examining the impact of timely diagnosis on outcomes such as stage at diagnosis and survival is complex because of the waiting time paradox, whereby patients with more advanced disease experience expedited diagnosis given their clinical condition and vice versa (Crawford, 2002). Nevertheless, studies accounting for the waiting time paradox have mostly identified that longer intervals to diagnosis are associated with poorer outcomes. Indeed, more recent analyses of diagnostic timeliness and stage at diagnosis of colorectal cancer found non-linear associations supporting the benefits of earlier diagnosis for clinical outcomes (Tørring *et al*, 2017).

[1] Many thanks to Deborah Furness, Head of Enquiry Services at UCL Library, for her assistance in identifying the author of this editorial

Evidence indicating the benefits of earlier diagnosis of cancer for patient-reported outcomes is more convincing. Prolonged intervals to diagnosis have been associated with worse experience of subsequent care, poorer patient satisfaction, and reduced quality of life among individuals diagnosed with cancer (Robinson *et al*, 2012; Grooss *et al*, 2016; Mendonca *et al*, 2016; Dahl *et al*, 2017). Faster diagnostic resolution is also likely to benefit patients with new symptoms who are investigated but not diagnosed with cancer (Brocken *et al*, 2012). Moreover, detecting cancer at an earlier than later stage is likely to be more cost-effective given the increasing costs of managing advanced disease (Smith & Hillner, 2011; Blumen *et al*, 2016).

1.2 How can diagnostic timeliness be conceptualised?

While the importance of earlier diagnosis of cancer has long been recognised, significant theoretical contributions to aid the understanding of diagnostic timeliness have mostly surfaced in the last decade. This section considers several theoretical models that provide the conceptual framing of the events and intervals that lead to a cancer diagnosis.

The National Awareness and Early Diagnosis Initiative (NAEDI) launched in 2008 proposed a logic model (referred to as 'The NAEDI hypothesis' in the literature) to guide emerging early diagnosis policies and research in England (Richards, 2009a). It proposed mechanisms by which earlier presentation could contribute to better survival and mortality, and provided a framework within which to identify priorities for targeting delays in diagnosis. The model was subsequently updated in 2015 to better reflect the complex nature of the diagnostic process for cancer patients from symptom onset to diagnosis and treatment (Hiom, 2015) (Figure 1.1).

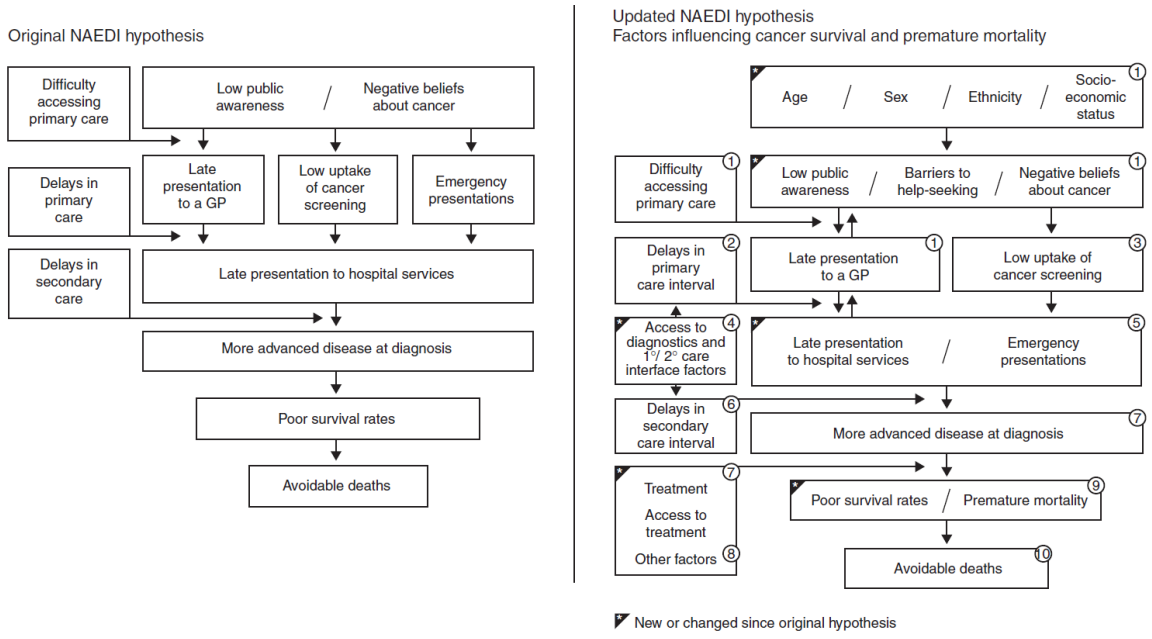


Figure 1.1 Original and updated NAEDI hypothesis (Hiom, 2015)

In addition to the 'NAEDI hypothesis', subsequent research in early diagnosis has been greatly influenced by the Pathways to Treatment model, which is based on the Andersen Model of Total Patient Delay (Figure 1.2) (Andersen *et al*, 1995; Walter *et al*, 2012; Scott *et al*, 2013). This model considers the series of events from the detection of a bodily change to the diagnosis of cancer, and the start of cancer treatment. It illustrates the iterative nature of the processes involved, and the resulting intervals in-between events leading to diagnosis and treatment of cancer in symptomatic patients (Walter *et al*, 2012). The subsequently published Aarhus statement brought international consensus on the definition of time intervals relating to cancer diagnosis (Figure 1.3) (Weller *et al*, 2012). Together, the above theoretical advancements published in the last 10 years have established key measures of diagnostic timeliness in symptomatic cancer patients (Richards, 2009a; Walter *et al*, 2012; Weller *et al*, 2012; Scott *et al*, 2013; Hiom, 2015).

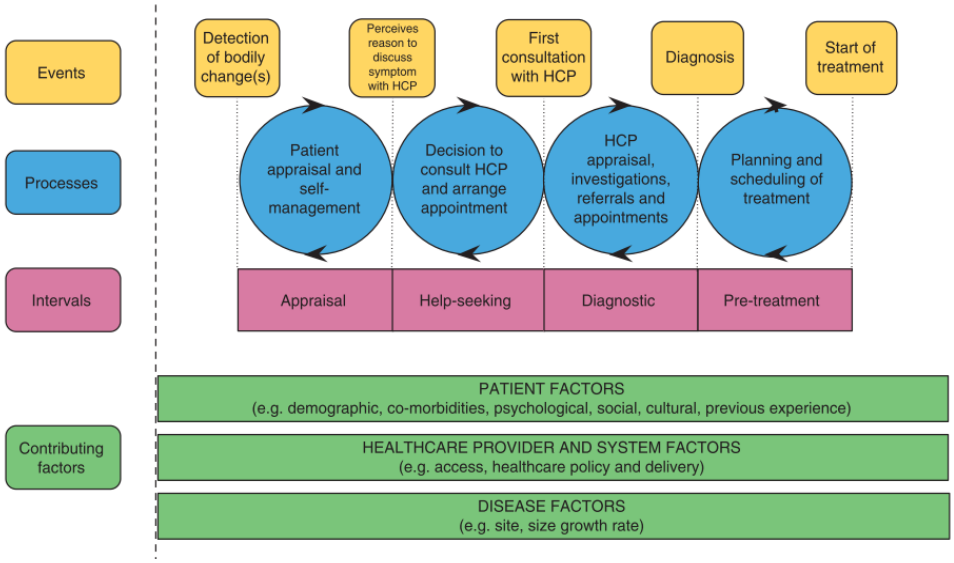


Figure 1.2 Walter's model of Pathways to Treatment (Walter et al, 2012)

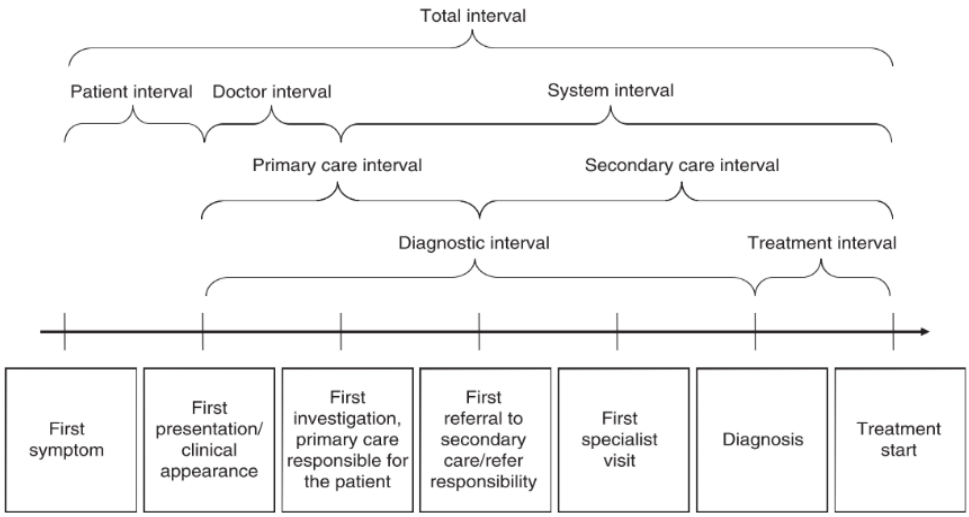


Figure 1.3 Aarhus statement pathway model (Weller et al, 2012)

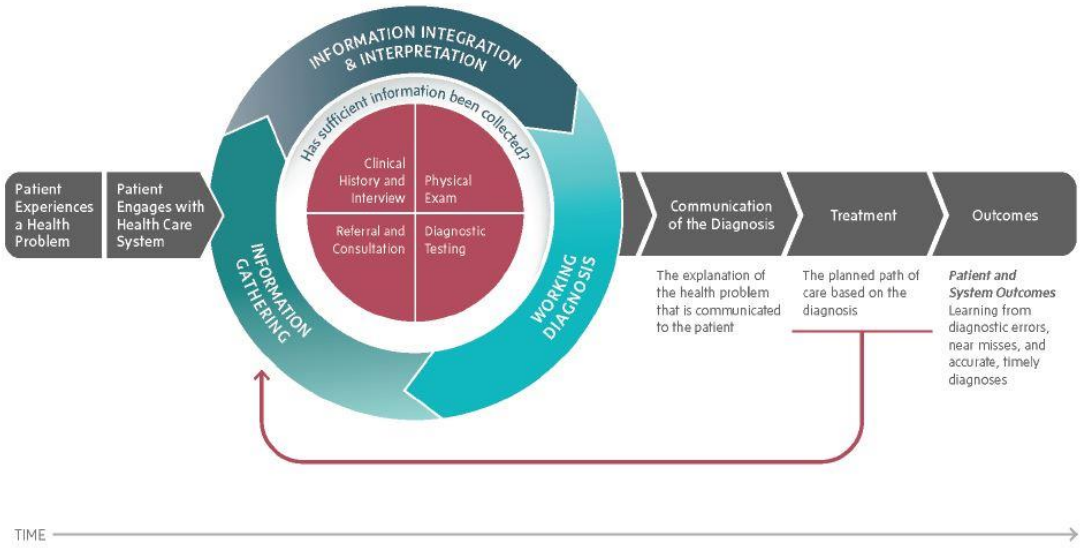


Figure 1.4 Conceptualisation of the diagnostic process by the US Institute of Medicine Committee (National Academies of Medicine, 2015)
The central sphere indicates the iterative nature of reaching diagnostic resolution, comparable to the cyclical processes conceptualised in the Pathways to Treatment model (Figure 1.2).

Other disciplines such as patient safety and quality improvement also offer relevant frameworks and concepts by which early diagnosis may be conceptualised (Lyratzopoulos *et al*, 2014; National Academies of Medicine, 2015; Singh & Sittig, 2015; Liberman & Newman-Toker, 2018). Critical among such examples is the US National Academy of Medicine 'Improving Diagnosis in Healthcare' framework, which conceptualised 'delayed' diagnosis as one possible diagnostic error arising from the iterative and complex processes required for reaching a diagnosis (Figure 1.4) (National Academies of Medicine, 2015). While the Institute of Medicine framework has no specific disease focus, cancer is often used as a disease model for quality improvement and therefore lends itself well to the field of early cancer diagnosis.

Additionally, research in early diagnosis is increasingly conducted with an awareness of the potential for overdiagnosis. The overdiagnosis of cancer may be defined as the diagnosis of a tumour that would have otherwise not have caused symptoms or premature death (Welch & Black, 2010; Esserman *et al*, 2014; Biswas *et al*, 2015). The tensions between early, and over-, diagnosis are touched upon briefly as part of this thesis (see Chapter 5).

1.3 Achieving early diagnosis: where are we now?

1.3.1 Early diagnosis activities in England

In concert with the above conceptual realisations, the unabated rise of cancer-related morbidity and mortality has led to considerable investment in early diagnosis interventions globally (World Health Organization, 2002; Calanzani *et al*, 2017). Early diagnosis of cancer has been especially prioritised in England, where cancer survival has been shown to be significantly poorer than countries with comparable health systems (Abdel-Rahman *et al*, 2009; Richards, 2009b). Two of the earliest examples of national-level early diagnosis policies targeting symptomatic individuals in England are the Be Clear on Cancer campaigns and the introduction of fast-track 2-week-wait referral pathways.

1.3.1.1 Be Clear on Cancer: a public health education campaign

The nationwide 'Be Clear on Cancer' campaigns were launched in 2012, representing what has been described as the greatest national-level investment in early diagnosis so far (Moffat *et al*, 2015; Public Health England, 2016a). The campaigns aimed to raise awareness of the likely symptoms of cancer to promote timely help-seeking, thereby shortening the average patient's time to presentation (Cancer Research UK, 2014). Such public health campaigns relate to a long-standing tradition of using social marketing as a vehicle for health promotion (Public Health England, 2017a; Wellcome Collection, 2017), and similar messages continue to be promoted across the world by governments, public health agencies, cancer charities, and patient advocacy groups (see Figure 1.5) (Scottish Government, 2012; Cancer Australia, 2013; World Wide Breast Cancer, 2016; CDC, 2017).

1.3.1.2 The '2-Week-Wait' system: an expedited clinical pathway for suspected cancer

Secondly, fast-track 2-week-wait referral pathways from primary care to specialist services for patients with suspected cancer were first introduced in 1999 in England, accompanied by guidelines to support their use (Department of Health, 2000). An updated version of the guidelines were published by the National Institute for Health and Care Excellence (NICE) in 2005, and subsequently updated more recently in 2015 (NICE, 2005, 2015). The pathways aimed to minimise delays to cancer diagnosis and thereby contribute to a reduction in cancer mortality (Jones *et al*, 2001). Elements of the above system have since been adopted in Scotland, Denmark, Australia, the Netherlands, and other countries (Prades *et al*, 2011; Probst *et al*, 2012; Healthcare Improvement Scotland, 2014; Cancer Council Austalia, 2017; Helsper *et al*, 2017).

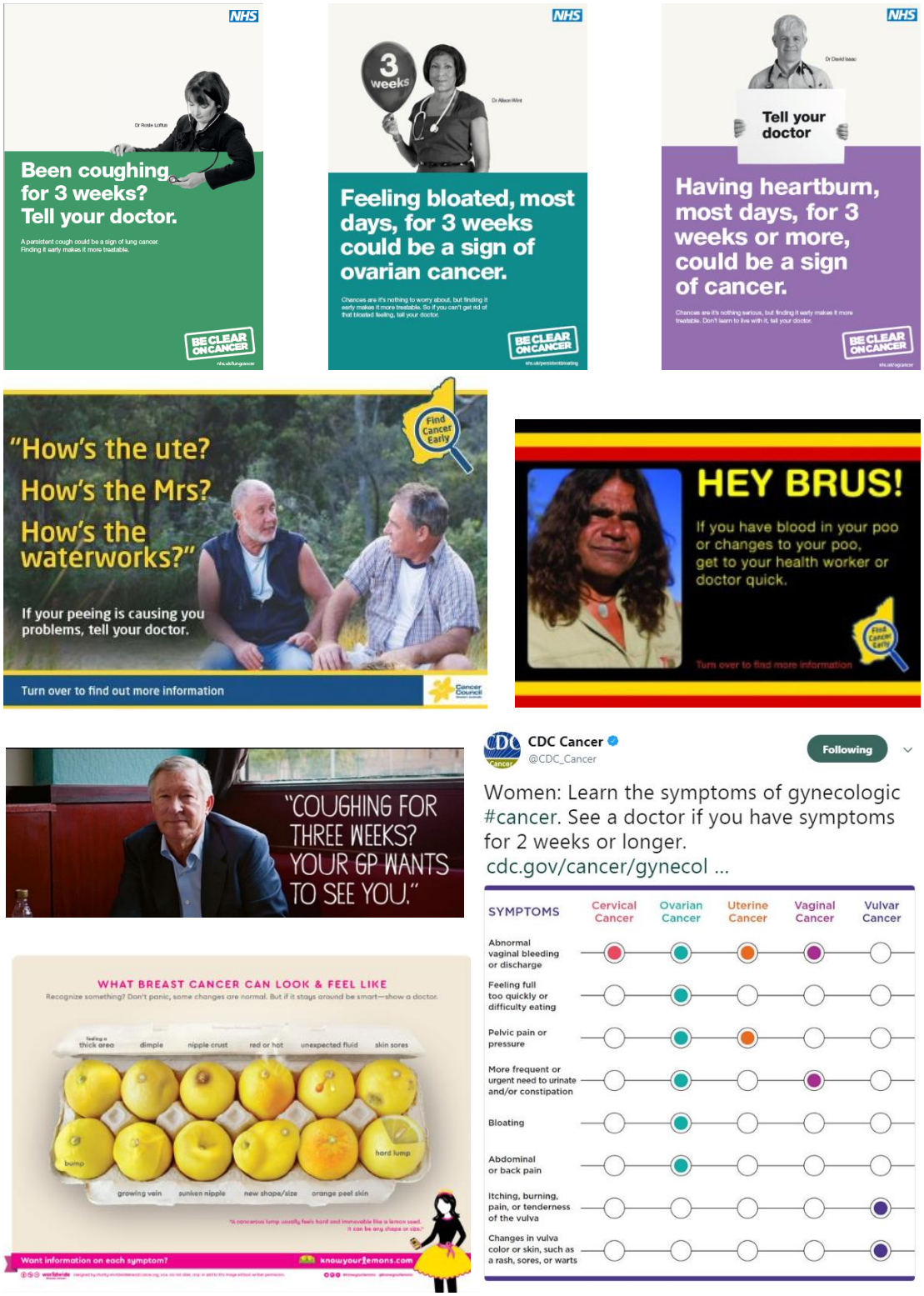


Figure 1.5 Examples of cancer symptom awareness campaigns
Clockwise from top: Be Clear on Cancer campaigns in England; Cancer Australia campaigns; US Centers for Disease Control and Prevention (CDC) promoting awareness of gynaecological cancer symptoms and prompt help-seeking; World Wide Breast Cancer "Know your lemons" campaign; and the Scottish Government's "Detect Cancer Early" campaign (Scottish Government, 2012; Cancer Australia, 2013; World Wide Breast Cancer, 2016; CDC, 2017).

In the face of continuing national and international interest in early diagnosis interventions such as the above, it is important to note that both the Be Clear on Cancer campaigns and 2-week-wait referral system represent complex interventions, characterised by multiple components and interacting factors that influence and moderate their effectiveness (Campbell *et al*, 2007; Craig *et al*, 2008; Calanzani *et al*, 2017). Yet, the theoretical understanding underpinning their design, implementation, and evaluation remains critically under-developed. In the absence of an explicit logic model, the design (and evaluation) of these early diagnosis interventions reflect several assumptions, a great many of which relate to the epidemiology of presenting symptoms.

1.4 Why are presenting symptoms important for early diagnosis?

The presenting symptoms of cancer are critical components of early diagnosis interventions. Public health education campaigns such as Be Clear on Cancer aim to raise awareness of particular symptoms likely to be associated with cancer, while urgent referrals for suspected cancer are mandated for specific symptomatic presentations.

Current epidemiological evidence supports substantial variation in intervals to diagnosis by cancer site, which suggests that symptoms at presentation are key determinants of diagnostic timeliness (Allgar & Neal, 2005; Baughan *et al*, 2009; Hansen *et al*, 2011; Keeble *et al*, 2014; Lyratzopoulos *et al*, 2015; Helsper *et al*, 2017). Indeed, drawing upon the theoretical frameworks summarised in Section 1.2, the nature of symptoms that cancer patients experience before diagnosis are likely to influence their recognition and appraisal as symptoms, and therefore affect the time to presentation (conceptualised as the patient interval). Post-presentation, similar appraisal processes by healthcare professionals will influence clinical decision-making and therefore the time to referral to specialist assessment (the primary care interval).

Evidence regarding the presenting symptoms of cancer and associated intervals before diagnosis could therefore contribute to the theoretical basis of early diagnosis interventions. This is particularly important as the effects of such interventions on outcomes such as stage at diagnosis or cancer survival remains uncertain (Ironmonger *et al*, 2014; Moffat *et al*, 2015; Jensen *et al*, 2016; Emery *et al*, 2017). At present however, evidence in this regard remains somewhat limited, as will be shown in Chapter 3: Literature review.

1.5 Chapter summary

Early diagnosis is a key aspect of cancer control strategies, which may amplify the effectiveness of more conventional approaches such as prevention and screening. While the potential benefits of early diagnosis for clinical outcomes are yet to be demonstrated for all cancer sites, long-standing beliefs in the value of expediting diagnosis have prevailed, supported by a more

robust body of evidence on the association between expedited diagnosis and patient-reported outcomes.

Several theoretical models have contributed to the conceptualisation of the diagnostic pathway as a series of intervals which are often the target of complex early diagnosis interventions. These interventions are often predicated on the nature of presenting symptoms that individuals experience before a cancer diagnosis, given the likely role of presenting symptoms on time to presentation and referral.

The growing burden of cancer across the world means that such interventions are increasingly commonplace, as exemplified in England, and are likely to continue to be used as an element of cancer control. Crucially however, the theoretical underpinning of such early diagnosis interventions is limited. Examining the epidemiology of presenting symptoms of cancer could provide insight into the implicit logic models behind early diagnosis interventions, and contribute to their effectiveness.

Chapter 2: Thesis aims and objectives

This chapter sets out the aims and objectives of the thesis, and maps the contribution of each subsequent chapter.

2 Thesis aims and objectives

This thesis aimed to use epidemiological approaches to examine the presenting symptoms of cancer patients, in order to contribute to the theoretical underpinning of early diagnosis interventions. Specifically, the objectives of this thesis were to:

1. Review currently available epidemiological evidence on the presenting symptoms of different cancer sites and how these relate to diagnostic timeliness (Chapter 3);
2. Code and taxonomise free-text information on symptom status of cancer patients as captured by an audit of cancer diagnosis in primary care in England (Chapter 4 & 5);
3. Examine the presenting symptoms of patients diagnosed with specific cancers, namely the 'symptom signature' of a given cancer (Chapter 6);
4. Examine the distribution of cancer diagnoses among cancer patients distinguished by specific symptoms or symptom status, namely the 'cancer signature' of a given symptom or group of symptoms (Chapter 7–8);
5. Investigate differences in measures of diagnostic timeliness among cancer patients with different symptoms by examining variation in the patient interval, primary care interval, and number of consultations before referral (Chapter 6–8);
6. Examine presenting cancer symptoms and their associations with recorded stage at diagnosis (Chapter 9).

The resulting findings and associated implications for the design, implementation, and evaluation of public health and healthcare interventions have been considered in the Discussion chapter (Chapter 10).

2.1 My contributions to the research in this thesis

I developed the aims of this thesis and the design of the empirical studies with input from my supervisors Professor Georgios Lyratzopoulos and Dr Christian von Wagner, and also with input from Professor Greg Rubin (Newcastle University, formerly the University of Durham Medical School until mid-2017), who led the design of the National Audit of Cancer Diagnosis in Primary Care (NACDPC) upon which the empirical analyses were based.

Statistical advice was provided by Dr Gary Abel (Exeter University) who also acted as external supervisor, and Dr Sean McPhail (Public Health England) who was involved in the aggregation of the audit data prior to the start of my PhD.

I am responsible for the coding of free-text symptom information collected by the audit in addition to all statistical analyses presented in this thesis.

Chapter 3: The epidemiology of presenting symptoms in cancer patients – a literature review

This chapter presents the findings of a literature review which aimed to examine the spectrum of symptoms at presentation for patients subsequently diagnosed with different cancers, and investigate associations between different presenting symptoms and diagnostic timeliness.

Aspects of this chapter have been the subject of a peer-reviewed publication in Neoplasia[1].

[1] Koo MM, Hamilton W, Walter F, Rubin G, Lyratzopoulos G (2017) **Symptom signatures and diagnostic timeliness in cancer patients: a review of current evidence.** Neoplasia 20 (2) 165–174, http://dx.doi.org/10.1016/j.neo.2017.11.005 (see Appendix 3.1)

3 Literature review

3.1 Outline and rationale of this chapter

As described in Chapter 1 (Introduction), the importance of early diagnosis of cancer has long been recognised, but the epidemiology of symptomatic presentation remains a relatively new area of research. In this chapter, I firstly consider the methodological challenges in capturing data on symptoms at presentation and compare different principal approaches (Section 3.2), and subsequently use these considerations to inform the scope and methods of a review of literature on presenting symptoms of cancer and diagnostic timeliness (Section 3.3–3.6).

3.2 Defining presenting symptoms before diagnosis

3.2.1 Challenges in measuring symptoms

Capturing information on symptoms is challenging for several reasons. Firstly, by their very nature symptoms cannot be objectively observed, and therefore information is reliant on subjective personal experience. Additionally, there are many different dimensions that distinguish a symptom, for example, 'abdominal pain' encompasses a range of presentations that vary greatly in nature, intensity, duration, and temporal evolution. Further, symptoms may have overlapping definitions, for example abdominal bloating (uncomfortable sensation of fullness) and distension (visible increase in abdominal girth) (Bankhead *et al*, 2005; Hamilton *et al*, 2009b); or have different connotations in lay versus professional language, for example 'change in bowel habit' is a term often used by clinicians to denote a clinical suspicion of colorectal cancer beyond the presence of constipation or diarrhoea alone (Hamilton *et al*, 2005b).

Secondly, the processes of symptom appraisal (detection of bodily changes, interpretation of these changes, and responses to interpretation) are influenced by a range of factors including the nature of symptoms and the context within which they are detected e.g. in the presence of pre-existing comorbidities; socio-cultural factors such as educational attainment, and health literacy (including awareness of likely cancer symptoms and awareness of the healthcare system); and psycho-social, emotional, and attitudinal aspects such as normalisation, fatalism, stoicism, optimism, and embarrassment (Powe & Finnie, 2003; Smith *et al*, 2005; Brindle *et al*, 2012; Emery *et al*, 2013; Niksic *et al*, 2015; Whitaker *et al*, 2015b; Dobson *et al*, 2018). The above challenges have consequences for how information on symptoms (and associated time intervals) may be collected for epidemiological analysis, as described in the next section.

3.2.2 How may presenting symptoms and intervals before diagnosis be studied in cancer patient populations?

As referred to in Chapter 1 (Introduction), measuring the patient interval and the primary care interval rely on judgements about the date of first symptom and date of first presentation or

clinical appearance (Weller *et al*, 2012). Two principal study designs have been used to obtain such information and examine presenting symptoms and diagnostic intervals: collecting self-reported information from patients, and extracting information from patients' health records (Weller *et al*, 2012; Keeble *et al*, 2014). The following section discusses the strengths and limitations of each approach, the different types of bias that may be expected as a result of the study design, and implications for the study of presenting symptoms and time to diagnosis.

3.2.2.1 The patient as the source of symptom information

Information on presenting symptoms can be directly elicited from patients through interviews (Burgess *et al*, 2006; Evans *et al*, 2014; Lim *et al*, 2014; Walter *et al*, 2014; McLachlan *et al*, 2015; Queenan *et al*, 2017) or questionnaires (Vine *et al*, 2003; Forbes *et al*, 2014; Howell *et al*, 2015). Such methods can provide valuable first-hand insights into cancer patients' symptomatic and diagnostic experience.

Patients may be prompted to identify their presenting symptoms from a pre-defined list (symptom recognition or 'closed' questions), or to describe them without any prompting (symptom recall or 'open' questions), which can affect the degree of recall inaccuracies or bias (Robb *et al*, 2009; Waller *et al*, 2009). Symptom information obtained via patient recall may lead to the underestimation of symptom frequencies, particularly for vague symptoms that are associated with a range of potential causes. In comparison, information obtained via recognition may lead to overestimation through the spurious affirmation of symptoms that are unlikely to be associated with the subsequently diagnosed cancer. Prompting patients to consider their symptom status in respect of calendar 'landmark' dates (such as public holidays or events and dates of personal significance) may be helpful, although evidence on how this affects accuracy remains limited (Mills *et al*, 2014).

Studies that use the patient as the source of information on presenting symptoms can also be distinguished by whether such information is collected before or after the diagnosis. While asking patients about their presenting symptoms after diagnosis is more convenient due to easier case identification and recruitment, it can lead to both recall and survivorship bias. Recall bias may, as described above, result in the under-reporting of vague or non-specific symptoms, while more distinctive 'alarm' symptoms (that are recognised as being strongly associated with cancer) may be over-reported and attributed to the subsequently diagnosed cancer. Survivorship bias on the other hand, result in the under-representation of cancer patients with poor prognosis, whose presenting symptoms could be different to that of the studied patients (Abel *et al*, 2016). Collecting information prospectively (before a diagnosis of cancer has been made) has the advantage of minimizing both these types of bias although it may be associated with practical difficulties (Walter *et al*, 2015, 2016b, 2016a).

3.2.2.2 Healthcare records as the source of symptom information

Alternatively, information on presenting symptoms recorded during healthcare encounters can be extracted from patients' medical records, most commonly from primary care (Hippisley-Cox *et al*, 2004; Blak *et al*, 2011; Herrett *et al*, 2015). Information is coded using structured diagnostic nomenclature systems such as Read codes, but may also be accompanied by unstructured free-text information (Nadkarni *et al*, 2011; National Academies of Medicine, 2015; Price *et al*, 2016; Helsper *et al*, 2017). Coded information is easier to extract and analyse, but can be less sensitive to the multi-dimensional nature of symptom experience and prone to misclassification bias, as commonly discussed in studies using databases (Koshiaris *et al*, 2018). In comparison, free-text information can provide richer insights into symptomatic presentation and related contextual details, but is often more difficult to access (due to greater information governance concerns about potentially sensitive or identifying information) and challenging to analyse (Koeling *et al*, 2011).

Importantly, using healthcare records to estimate the occurrence of symptoms before the diagnosis of cancer assumes that relevant symptoms are accurately communicated by the patient, elicited by the doctor, and recorded in the patient's record during the consultation. In many instances, these assumptions may not be met. For the patient, psychosocial barriers (Lim *et al*, 2014; Whitaker *et al*, 2015a), cancer fear (O'Mahony *et al*, 2013; McCutchan *et al*, 2015), and perceived or actual time pressures during the consultation may prevent complete disclosure (Forbes *et al*, 2013; Cromme *et al*, 2016), while doctors may not record all symptoms due to time constraints or perceived clinical importance (Price *et al*, 2016).

On the other hand, studies collecting symptom information from healthcare records are less prone to the risk of selection and recall bias compared to patient-reported data, as information is collected prospectively and prior to diagnosis for all patients.

3.2.3 Summary

Considering the strengths and limitations of the two methods of capturing information on presenting symptoms highlights the difficulties in measuring symptoms; comparisons have found inconsistences between patient-reported and doctor-reported information (Malats *et al*, 1995; Lynch *et al*, 2008; Pérez *et al*, 2008; Larsen *et al*, 2014; Leiva *et al*, 2017). Nonetheless, records-based approaches enable the examination of large and representative samples of patients, and are increasingly being used, particularly as they facilitate the study of patients with rarer cancers (Din *et al*, 2015; Jensen *et al*, 2015; Lyratzopoulos *et al*, 2015; Nadpara *et al*, 2015). Further, records-based studies are more comparable to analyses of the NACDPC data as planned in this thesis, which is based on information collected from primary care records. These considerations informed the scope of the following literature review, which aimed to examine

and synthesise current evidence on the presenting symptoms of cancer and associated diagnostic timeliness.

3.3 Literature review: Introduction

As examined in Section 3.2.1 of this chapter, measuring presenting symptoms (and diagnostic intervals) among cancer patients is challenging. Furthermore, the body of literature relating to the nature and frequency of presenting symptoms, and associations with diagnostic timeliness has not yet been synthesised. Therefore, I conducted a literature review aimed at examining the available evidence regarding the spectrum of presenting symptoms associated with different cancer sites, and also evidence about associations between presenting symptoms and intervals to diagnosis.

3.4 Literature review: Methods

3.4.1 Search strategy

The objective of the review was to identify studies describing the frequency and the spectrum of presenting symptoms in cancer patients. Searches were conducted in the MEDLINE database using the key words 'symptom', 'cancer', and 'diagnosis' and other synonyms for full text publications in English indexed through to April 2017. All retrieved articles were screened by title, abstract, and then articles that did not meet exclusion criteria were screened by full text. Reference lists of studies selected for full-text review were hand-searched, and I also consulted several experts in early diagnosis in order to identify further eligible studies[2]. All retrieved studies providing evidence regarding the symptom signature of a cancer site were additionally examined for information on associations between symptoms and diagnostic intervals.

3.4.2 Inclusion criteria

Section 3.2.2 of this chapter identified two approaches to measuring presenting symptoms and intervals to diagnosis among cancer patients: from the patient themselves (ideally where the information is collected prospectively and before diagnosis); or primary care records, to capture information on the symptoms of patients presenting to primary care. Therefore, I included studies where symptom information was collected prospectively as part of primary care records, or other approaches that involved prospective collection of symptom information.

3.4.3 Exclusion criteria

Studies describing self-reported symptoms captured retrospectively (after diagnosis) were excluded due to the high risk of bias. Studies on paediatric, teenager and young adult cancer

[2] Professor Georgios Lyratzopoulos (principal supervisor) and Professor Greg Rubin (collaborator), and Professors Fiona Walter and William Hamilton (co-authors of related publication)

patient populations, and studies based in low- and middle-income settings were excluded as they were deemed not comparable.

3.4.4 Risk of bias assessment

I assessed risk of bias within studies using a modified risk of bias assessment tool based on the REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) guidelines (Benchimol *et al*, 2015). The Methods, Results, and Discussion sections of the RECORD guidelines were reviewed and modified to suit the research question as necessary. The 'Title and Abstract', 'Introduction', and 'Other Information' (funding and accessibility of data) sections were not considered necessary for risk assessment and were excluded from considerations.

Additionally, features of the Quality Assessment of Diagnostic Accuracy Studies 2nd version (QUADAS-2) checklist were incorporated into the risk of bias assessment tool (Whiting, 2011). QUADAS-2 is a checklist designed to assess the quality of diagnostic accuracy studies. It was deemed as an appropriate source for developing the risk of bias tool for this literature review, as initial search results indicated that studies likely to be included in the review were most commonly studies examining the diagnostic accuracy (predictive value) of presenting symptoms in cancer patients. Signalling questions from the 'patient selection' and 'index test' domains of QUADAS-2 were mapped out onto the relevant RECORD sections. The other two domains of QUADAS-2 ('Reference Standard' and 'Flow and Timing') were not used as they were deemed not to be relevant to the context of the review.

The resulting composite tool largely followed the structure of the RECORD guidelines (see Appendix 3.2), assessing risk of bias across six dimensions:

- study setting
- study population
- data sources/measurement (symptoms)
- study size (external validity)
- data cleaning
- other sources of bias

Risk of bias was judged as 'low', 'high', or 'unclear', and the 'unclear' category was only used when there was insufficient data to make a judgement, as recommended in the QUADAS-2 checklist. This modified risk of bias tool was used to further exclude any studies that had three or more dimensions (i.e. half the number of dimensions) with "high" risk of bias.

3.5 Literature review: Results

3.5.1 Summary of findings

Retrieved studies: A total of 44 studies including information on presenting symptoms for 17 cancer sites (Appendix 3.3). The majority of retrieved studies were case-control or cohort

studies based in England examining the predictive value of symptoms: for such studies, the sample size and symptom frequency relevant to cases (and not controls) were extracted. Of the included studies, 20 (44%) had low risk of bias across all examined dimensions while 11 studies had high risk of bias across two dimensions (study population and symptom information) (Appendix 3.3).

Cancer site coverage: No high-quality population-level evidence on the frequency of presenting symptoms before diagnosis could be identified for the following 11 cancers: laryngeal cancer; melanoma; mesothelioma; oral cancer; penile cancer; sarcoma; small intestinal cancer; testicular cancer; thyroid cancer; vaginal cancer; and vulval cancer.

Of the included studies, nearly all studies focused on single cancer sites, with the exception of colorectal, oesophago-gastric, and renal tract cancers, which were treated as single entities respectively (Stephens *et al*, 2005; Hippisley-Cox & Coupland, 2011a, 2012c, Collins & Altman, 2012b, 2013b; Stapley *et al*, 2013). Most evidence related to colorectal cancer (9 studies), pancreatic cancer (6 studies), and lung cancer (6 studies); only a single publication was identified for five cancers (brain, cervical, endometrial, leukaemia, myeloma).

Evidence about associations between presenting symptoms and intervals: Four of the 43 studies included in the review contained evidence on associations between individual symptoms and intervals to diagnosis (Pruitt *et al*, 2013; Walter *et al*, 2015, 2016a, 2016b). One study was not included in the consideration of symptom signatures as there was uncertainty regarding exactly how symptom information was collected (described as a combination of recall and recognition) (Leiva *et al*, 2017). However, evidence pertaining to symptom-specific diagnostic timeliness was considered in Section 3.5.3.

3.5.2 The symptom signature of cancer sites

The findings of the literature review are presented based on the type of symptom signature, which takes symptom heterogeneity (the 'breadth' of the symptom signature) and the associated predictive value of symptoms into account (see Figure 3.1). Selected symptom signatures are presented in detail alongside discussions, with the remainder of symptom signatures presented in Appendix 3.4. Cancer sites (where first mentioned) are highlighted in bold print to aid navigation.

In this review, symptoms have been described as 'alarm' or 'non-alarm' based on known positive predictive value (PPV) for cancer. The NICE guidelines for referral of suspected cancer in England use a PPV threshold of 3% or greater to classify symptoms as mandating fast-track 2-week-wait referral or specialist investigation; this has been used as a reference point (NICE, 2015).

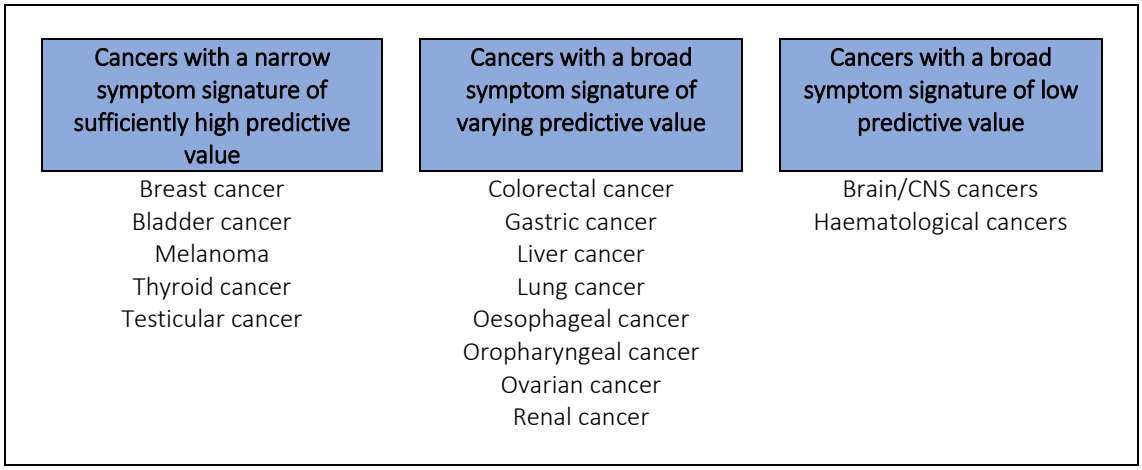


Figure 3.1 Taxonomy of cancer site-specific symptom signatures
Based on nature and frequency of presenting symptoms and their associated predictive value for malignancy at presentation.

3.5.2.1 Cancers with a narrow symptom signature

This category includes several cancers where the majority of patients present with one symptom with a strong association with a given cancer (typically these are alarm symptoms with relatively high predictive value for cancer). For example, the majority of women diagnosed with **breast cancer** initially present with a breast lump, which is associated with a relatively high predictive value for cancer (Table 3.1) (Walker *et al*, 2014; Redaniel *et al*, 2015). Similarly, most **bladder cancer** patients present with macroscopic (visible, frank) haematuria (Table 3.2) (Shephard *et al*, 2012; Price *et al*, 2014, 2016).

Other cancers that can be considered in this 'narrow symptom signature' category include **thyroid cancer, melanoma, testicular cancer, penile cancer, vaginal cancer,** and **vulval cancer**, although relevant epidemiological evidence to support this could not be identified.

Table 3.1 Population-based estimates of the frequencies of presenting symptoms among breast cancer patients (Walker et al, 2014; Redaniel et al, 2015)

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Walker et al., 2014	Primary care, CPRD data (Read coded)	2000–09	3166	40+ years	Breast lump 44.1% Breast pain 2.4% Nipple retraction 1.0% Nipple discharge 1.0%
Redaniel et al., 2015[1]	Primary care, CPRD data (Read coded)	1998–09	8544	15+ years	Breast lump 93.5% Breast pain 4.6% Nipple distortion 1.5% Nipple eczema 0.2% Breast skin changes 0.2% Bloody nipple 0.01%

1 All symptom frequencies calculated manually based on the number of breast cancer patients who had presented with a breast symptom, excluding those who were diagnosed following disclosure of family history (i.e. in the absence of any symptoms)

Table 3.2 Population-based estimates of the frequencies of presenting symptoms among bladder cancer patients (Shephard *et al*, 2012; Price *et al*, 2014, 2016)

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Shephard et al., 2012	Primary care, CPRD data (Read coded)	2000–09	4915	40+ years	Visible haematuria 53% Invisible haematuria 2.6% Dysuria 9% Abdominal pain 7% Constipation 6% Urinary tract infection 17%
Price et al., 2014[1]	Primary care, CPRD data (Read coded & uncoded data)	2000–09	4915	40+ years	Macroscopic haematuria 64.0% Microscopic haematuria 6.4%
Price et al., 2016[2]	Primary care, CPRD data (Read coded & uncoded data)	2000–09	4935	40+ years	Visible haematuria 63.8% Abdominal pain 12.2% Jaundice 0.4%

1 same study population as Shephard et al., 2012 & Price et al., 2016 but used uncoded data to examine haematuria only
2 majority of patients derived from same study population as Shephard et al., 2012 & Price et al., 2014; uncoded data was used (as in Price et al., 2014) to examine several purposefully selected symptoms

3.5.2.2 Cancers with a broad symptom signature

This category includes cancer sites characterized by a broad symptom signature that encompasses several common symptoms as opposed to a single dominant one. For some cancers, this includes certain alarm symptoms (e.g. **colorectal, lung, pancreatic, oesophago-gastric,** and **ovarian cancers**) while for other cancers, presenting symptoms are chiefly non-specific (e.g. **haematological malignancies,** and **brain and CNS cancers**).

Many common cancers have broad symptom signatures consisting of multiple symptoms, of which one or two are alarm symptoms. For example, nine studies report rectal bleeding as a common presenting symptom of **colorectal cancer**, although estimates vary substantially (16–60%) (Table 3.3) (Hamilton *et al*, 2005b, 2009a; Stapley *et al*, 2006; Collins & Altman, 2012a; Hippisley-Cox & Coupland, 2012a; Esteva *et al*, 2013; Pruitt *et al*, 2013; Renzi *et al*, 2016; Walter *et al*, 2016a). Other comparably common presenting symptoms among colorectal cancer patients include abdominal pain, diarrhoea, and constipation, which are associated with much lower predictive value and greater diagnostic difficulty.

Table 3.3 Population-based estimates of the frequencies of presenting symptoms among colorectal cancer patients (Hamilton *et al*, 2005b, 2009a; Stapley *et al*, 2006; Collins & Altman, 2012a; Hippisley-Cox & Coupland, 2012a; Pruitt *et al*, 2013; Redaniel *et al*, 2015; Renzi *et al*, 2016; Walter *et al*, 2016a)

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Hamilton et al., 2005a	Primary care, data from 21 general practices in Exeter	1998–02	349	40+ years	Rectal bleeding 42.4% Abdominal pain 42.4% Diarrhoea 37.8% Constipation 26.1% Weight loss 26.9%

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Stapley et al., 2006[1]	Primary care, data from 21 general practices in Exeter	1998–02	349	40+ years	Rectal bleeding 39% Abdominal pain 38% Diarrhoea 32% Constipation 23% Weight loss 23%
Hamilton et al., 2009a	Primary care, THIN data (Read coded)	2001–06	5477	30+ years	Abdominal pain 29.7% Constipation 27.0% Diarrhoea 18.0% Rectal bleeding 15.6% Change in bowel habit 11.2% Weight loss 10.2%[2]
Hippisley-Cox & Coupland, 2012a	Primary care, QResearch data (Read coded)	2000–10	2603	30–84 years	Rectal bleeding 32.3% Abdominal pain 32.5% Appetite loss 1.8% Weight loss 4.1% Change in bowel habit 1.5%[3]
Collins & Altman, 2012a[4]	Primary care, THIN data (Read coded)	2000–08	3712	30–84 years	Rectal bleeding 36.7% Abdominal pain 32.9% Appetite loss 1.2% Weight loss 5.8% Change in bowel habit 2.4%[3]
Pruitt et al., 2013[5]	SEER-Medicare data (ICD-9 codes)	1998–05	9669	65+ years	Abdominal pain 19.2% Abnormal stool 4.9% Anaemia 31.8% Anorexia or unexplained weight loss 5.2% Constipation 5.0% Fatigue 15.9% Other GI symptoms, other bowel changes 3.2% Rectal bleeding or rectal/GI haemorrhage 30.0% Vomiting or nausea 2.4%
Redaniel et al., 2015	Primary care, CPRD data (Read coded)	1998–09	5912	15+ years	Abdominal pain 28% Anorexia 0.8% Constipation 11.5% Diarrhoea 14.9% Fatigue 4.6% Weight loss 3.5%
Walter et al., 2016a[6]	Primary & secondary care data; self-reported symptoms before diagnosis	2010–12	152	40+ years	Change in bowel habit 61% Rectal bleeding 60% Indigestion, reflux, or persistent abdominal pain 28% Fatigue or tiredness 38% Feeling different “in yourself” 25% Loss of appetite 20% Back pain 8% Weight loss 16% Acute GI illness 2% Perianal pain or discomfort 2% Wind or flatulence 1% Bloating 1% Rectal mucus/discharge 1% Non-abdominal pain 1%

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Renzi et al., 2016[7]	Primary care, CPRD data (Read coded)	2005–06	1606	25+ years	Abdominal pain 31.1% Rectal bleeding 22.7% Anaemia 19.6% Diarrhoea 15.5% Change in bowel habit 12.8% Constipation 12.6% Fatigue 4.5% Weight loss 4.4%

GI: gastro-intestinal
1 same study population as Hamilton et al., 2005b but included presenting symptoms from 1 year before diagnosis only
2 calculated by combining frequencies of two weight loss categories (5–10% weight loss, ≥10% weight loss)
3 frequency of change in bowel habit was calculable among male patients only
4 all symptom frequencies calculated manually by combining symptom frequencies reported separately for men and women diagnosed with colorectal cancer
5 all symptom frequencies calculated manually by combining symptom frequencies reported separately for colon and rectal cancer patients; frequencies >2% listed only for space
6 all symptom frequencies calculated manually by combining symptom frequencies reported separately for colorectal cancer patients who experienced a symptom as a "first symptom" and "subsequent symptom"
7 all symptom frequencies calculated manually by combining symptom frequencies (that occurred 1 year before diagnosis) reported among colon and rectal cancer patients

Lung cancer also has a broad symptom signature with symptoms of varying predictive value: while it includes haemoptysis, a highly predictive symptom of malignancy (Hamilton *et al*, 2005a), evidence from six studies suggests that this is a relatively rare presenting symptom, reported in 23% of patients subsequently diagnosed with lung cancer at the most (Appendix 3.4) (Hamilton *et al*, 2005a; Hippisley-Cox & Coupland, 2011b; Ades *et al*, 2014; Nadpara *et al*, 2015; Redaniel *et al*, 2015; Walter *et al*, 2015).

Six studies describing the frequencies of the presenting symptoms of **pancreatic cancer** were identified (Appendix 3.4) (Hippisley-Cox & Coupland, 2012b; Stapley *et al*, 2012; Collins & Altman, 2013a; Keane *et al*, 2014; Price *et al*, 2016; Walter *et al*, 2016b). Jaundice has a high predictive value for pancreatic cancer but reported frequencies range from 12–43% among patients (Hippisley-Cox & Coupland, 2012b; Stapley *et al*, 2012; Collins & Altman, 2013a; Keane *et al*, 2014; Price *et al*, 2016; Walter *et al*, 2016b). The most common presenting symptom among pancreatic cancer patients is abdominal pain (reported range: 40–57% of cases), while other upper gastro-intestinal symptoms such as indigestion and nausea and vomiting are also common – and given their frequency among primary care consultees, these symptoms have low predictive values. The identified studies also reported frequencies of back pain and non-localized symptoms such as weight loss, and lethargy, fatigue, or malaise among considerable proportions of patients, indicating that the symptomatic picture of pancreatic cancer is usually a combination of vague and intermittent symptoms associated with considerable diagnostic difficulty (Hippisley-Cox & Coupland, 2012b; Stapley *et al*, 2012; Collins & Altman, 2013a; Keane *et al*, 2014; Price *et al*, 2016; Walter *et al*, 2016b).

Current data on the symptom signatures of **oesophageal** and **gastric cancers** are limited to those that consider these two sites in combination (Appendix 3.4) (Stephens *et al*, 2005; Hippisley-Cox & Coupland, 2011a; Collins & Altman, 2012b; Stapley *et al*, 2013). While dysphagia is the most common presenting symptom across this cancer patient population (an alarm symptom), one in two patients present with a broad spectrum of other symptoms, including abdominal pain, epigastric pain, reflux, dyspepsia, and systemic symptoms such as nausea or vomiting, loss of appetite, and weight loss (Stephens *et al*, 2005; Hippisley-Cox & Coupland, 2011a; Collins & Altman, 2012b; Stapley *et al*, 2013).

Likewise, **ovarian cancer** has a symptom signature encompassing a broad spectrum of abdominal symptoms (Appendix 3.4) (Ryerson *et al*, 2007; Hamilton, 2009a; Collins & Altman, 2012c; Hippisley-Cox & Coupland, 2012d; Lim *et al*, 2016). Abdominal distension has a reasonable predictive value for cancer but it is only reported by 6–38% of patients before diagnosis, while other common presenting symptoms such as abdominal bloating and abdominal pain have much lower predictive values (Hamilton *et al*, 2009b). Additionally, a wide variety of other symptoms with low predictive values have been described as presenting symptoms of ovarian cancer, such as bloating, vaginal bleeding, constipation, and systemic symptoms (Ryerson *et al*, 2007; Hamilton, 2009a; Collins & Altman, 2012c; Hippisley-Cox & Coupland, 2012d; Lim *et al*, 2016).

Some cancers have broad symptom signatures comprised almost entirely of symptoms with low predictive value. Published results from four studies indicate that haematological cancers **(leukaemia, lymphoma,** and **multiple myeloma)** have such symptom signatures, comprised of vague or non-localizing symptoms such as fatigue and weight loss (Appendix 3.4) (Howell *et al*, 2015; Shephard *et al*, 2015b, 2015a, 2016).

Although a proportion of patients with **brain cancer** are diagnosed after an acute event such as a seizure or neurological deficit, most patients are thought to initially experience non-specific symptoms, with very low predictive value (Table 3.4) (Hamilton & Kernick, 2007).

Table 3.4 Population-based estimates of the frequencies of presenting symptoms among brain or CNS cancer patients
(Hamilton & Kernick, 2007)

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Hamilton et al., 2007	Primary care, CPRD data	1988–06	3505	18+ years	Headache 10.2% Motor loss 8.8% New onset seizure 4.4% Confusion 3.1% Weakness 2.7% Memory loss 1.1% Visual disorder 1.0%

3.5.3 Symptoms and diagnostic intervals

To date, there has been limited examination of individual cancer symptoms and time to diagnosis. The majority of available evidence is based on the analysis of health records, and symptoms are often aggregated into broad categories for analysis. For example, patients with alarm symptoms across a range of cancers have been shown to experience shorter diagnostic intervals (time from symptomatic presentation to diagnosis) compared to those with non-alarm symptoms (Jensen *et al*, 2014; Neal *et al*, 2014; Din *et al*, 2015; Redaniel *et al*, 2015), and similar trends have been noted for the primary care interval among lung cancer patients (Guldbrandt *et al*, 2015). Other groupings of presenting symptoms have been used among specific cancer patient populations, such as lump versus no lump among breast or sarcoma patients (Dyrop *et al*, 2016; Webber *et al*, 2017).

Available evidence on individual symptoms and diagnostic timeliness was limited to five studies on three cancers (colorectal, lung, pancreatic), and chiefly derived from study designs that combined prospectively collected patient information with primary and secondary care records (Pruitt *et al*, 2013; Walter *et al*, 2015, 2016b, 2016a; Leiva *et al*, 2017).

The SYMPTOM studies collected symptom information prospectively from patients before diagnosis, and subsequently combined with information from primary and secondary care data. Investigators identified several symptoms associated with a shorter interval (e.g. chest or shoulder pain in lung cancer patients) while others were associated with longer time to diagnosis (e.g. weight loss in pancreatic cancer) (Walter *et al*, 2015, 2016b, 2016a).

The multi-centre DECCIRE study used a comparable design to collect information on the diagnostic process for 795 colorectal cancer patients in Spain (Esteva *et al*, 2007). Symptom information was elicited from patients shortly after diagnosis by a combination of recall and recognition, and corroborated with medical records from which pre-diagnostic intervals were estimated (Leiva *et al*, 2017). Of the symptoms examined by the DECCIRE study, vomiting was associated with a median time to diagnosis of 115 days (IQR: 931–171 days) while change in bowel habit was associated with a median time to diagnosis of 168 days (IQR: 88–288 days); however there was substantial variation in the relative ranks of symptoms (in respect of the reported respective median diagnostic interval) depending on the method of data collection (patient interview, hospital records, primary care records) (Leiva *et al*, 2017).

A US study using linked SEER-Medicare claims data described the length of time from symptomatic presentation to diagnosis (the diagnostic interval) by presenting symptom among colon and rectal cancer patients (Pruitt *et al*, 2013). The authors reported variation in time to diagnosis between symptoms, and distinct patterns of variation in interval length by symptom among colon cancer patients versus rectal cancer patients. For example, the median (IQR) diagnostic interval associated with rectal bleeding was 19 (7–48) days among colon cancer

patients and 21 (7–46) days among rectal cancer patients, while for nausea or vomiting, the median (IQR) diagnostic interval was 80 (20–206) days among colon cancer patients and 123 (30–253) days among rectal cancer patients. However the significance of these differences was not examined formally, either in univariate or multivariate analyses (Pruitt *et al,* 2013).

For other cancers, no high-quality studies regarding the association between symptoms and diagnostic intervals could be identified, although studies excluded from the review support variation in diagnostic timeliness by symptoms by cancer site. For example, studies reporting variation in diagnostic timeliness by symptom groups (bleeding versus all other, pain versus no pain) have been noted for cervical (Kaku *et al,* 2008); ovarian (Vine *et al,* 2003; Ouasmani *et al,* 2016); gastric (Stephens *et al,* 2005); sarcoma (George & Grimer, 2012); and testicular cancer (Öztürk *et al,* 2015) although external validity was often limited.

3.6 Literature review: Discussion

3.6.1 Main findings

This review identified a total of 44 population-based studies describing information on the symptom signatures of 17 common and rarer cancers. No high-quality evidence could be identified for 11 cancers, which included thyroid cancer, oral cancer, and mesothelioma. Information regarding the nature and frequency of symptoms before cancer diagnosis was typically available as part of investigations of the predictive value of a symptom for cancer.

Only four studies describing variation in diagnostic intervals by presenting symptom were included; these were all cancer site-specific investigations and described variation in diagnostic timeliness.

3.6.2 Comparison with existing literature

Previous discussions in early diagnosis research have described cancers as being 'harder' or 'easier' to suspect; this variation in diagnostic difficulty is widely regarded as being influenced by the symptom signature of the cancer (Lyratzopoulos *et al,* 2014). The findings of this review corroborate this hypothesis, demonstrate that the nature and frequencies of presenting symptoms varies substantially by cancer site, and provide an up-to-date summary of existing population-level evidence. It further indicates gaps in evidence, particularly regarding the association between symptoms and diagnostic timeliness.

The methods of this literature review were informed by the characteristics of rapid reviews and scoping reviews, which are increasingly utilised in health literature for evidence synthesis (Pham *et al,* 2014; Reynen *et al,* 2017). Nevertheless, the review also shared features of a systematic literature review: consultation with experts in the field and the use of a composite risk of bias assessment tool is likely to have minimised the omission of relevant high-quality evidence. More systematic methodologies are unlikely to have led to markedly different results and

conclusions given the substantial evidence gap regarding presenting symptoms of cancer and associated diagnostic timeliness.

3.6.3 Strengths and limitations

The scope and exclusion criteria of the literature review were informed by *a priori* consideration of two approaches to how symptom information may be captured. This led to the exclusion of several studies containing information on the presenting symptoms of cancer where symptom information had been collected retrospectively from the patient (many of these studies were qualitative). This may have led to the under-appreciation of the 'true' frequencies of presenting symptoms in cancer patients in the review presented in this chapter, as the frequency of presenting symptoms derived through patient self-reports is often higher compared to records-based methodologies (Rasmussen *et al*, 2014; Lim *et al*, 2016; Leiva *et al*, 2017).

Regardless of the exclusion, there was substantial variation in reported symptom frequencies across records-based studies reflecting the heterogeneity in how symptom information was reported, extracted, and collated. Nonetheless, the value and novelty of this literature review lies in the synthesis of epidemiological evidence on presenting symptoms, and its focus on studies that are likely to provide a closer comparison to data captured by the NACDPC, as examined in the subsequent chapters of this thesis.

Risk of bias in each of the included studies was assessed through a modified tool based on well-known and reputable RECORD and QUADAS-2 guidelines. Development of a modified tool is likely to have provided a better estimate of risk of bias compared to if I had used one of the existing guidelines alone. Given that the tool was developed by myself however, it may be limited in reliability or perspective. Additionally, assessment was conducted with no other reviewers; additional reviewers may have enhanced the validity of risk of bias assessments.

3.6.4 Implications and conclusions

Many of the included studies investigated patients with pre-specified symptoms (identified *a priori*), either from relevant literature or clinical guidelines. This means that rarer or less-specific symptoms might not have been captured, and/or underestimated. Examining all presenting symptoms of a cohort of cancer patients without prior restrictions can be valuable; a comparable data-driven method has resulted in the identification of thrombocytosis as a sign of cancer (Bailey *et al*, 2017).

Importantly, some presenting symptoms such as jaundice are thought to represent advanced disease (Walter *et al*, 2016b). In these patients, diagnostic difficulty could be minimal, but expediting their diagnosis may not necessarily lead to favourable clinical outcomes or alter prognosis. This association is explored empirically in Chapter 9 (Cancer alarm symptoms and stage at diagnosis).

3.6.5 Chapter summary

Measuring symptoms experienced before a cancer diagnosis at population level is challenging; the literature review presented in this chapter indicates a gap in evidence regarding presenting symptoms, and particularly with regards to associated diagnostic timeliness.

Investigating the presenting symptoms of cancer and how these are associated with diagnostic intervals will further the understanding of mechanisms that influence the diagnostic pathway at patient, healthcare professional, and system levels and could guide the development and implementation of public health and healthcare interventions promoting early diagnosis. The subsequent chapters aim to contribute to this evidence base.

Chapter 4: Data & Methods

This chapter provides an overview of the NACDPC initiative and data, describes how the symptom information was coded as part of this PhD, and explains aspects of data preparation that was needed before empirical analysis. Specific methods used for empirical analyses are described in each respective chapter separately (Chapters 5 to 9).

4 Data & methods

4.1 Overview of the NACDPC

4.1.1 Data provenance

As indicated in Chapter 3, there is a substantial gap in epidemiological evidence regarding the presenting symptoms of cancer patients, and associations with intervals before diagnosis. Clinical audits of primary care records can provide uniquely comprehensive insights into the diagnostic pathway of cancer patients (Munck *et al*, 2003; Baughan *et al*, 2009; Hansen *et al*, 2011; Rubin *et al*, 2011), but until the start of this PhD, this approach had not been used to examine the presenting symptoms of cancer and associations with diagnostic timeliness.

The empirical chapters of this thesis are based on data collected as part of the first English National Audit of Cancer Diagnosis in Primary Care (NACDPC) conducted between April 2009 and April 2010 (Rubin *et al*, 2011). The NACDPC was an initiative sponsored by the NHS "Cancer Action Team" that formed part of the broader National Awareness and Early Diagnosis Initiative (NAEDI) (see Chapter 1), and involved close collaboration with the Royal College of General Practitioners (RCGP) and the (then) National Cancer Intelligence Network (NCIN).

4.1.2 The contribution of the NACDPC to early diagnosis research thus far

Data collected through the NACDPC has been summarised in a report published by the UK's Royal College of General Practice (RCGP) (Rubin *et al*, 2011), and further examined as part of earlier diagnosis research, resulting in seven additional peer-reviewed publications (Lyratzopoulos *et al*, 2013a, 2015; Keeble *et al*, 2014; Hughes *et al*, 2015; Rubin *et al*, 2015; Tørring *et al*, 2017; Ozawa *et al*, 2018)[1]. This section briefly considers several notable reflections arising from these studies that are relevant to this thesis.

Two of the above publications have scrutinised the external validity of data collected through the NACDPC (Rubin *et al*, 2011; Lyratzopoulos *et al*, 2013a). Audited patients were found to be largely comparable to the contemporaneous incident cancer patient population in England, with regards to age groups, sex, and cancer site case-mix (Rubin *et al*, 2011). Further, a comparison of participating general practices with non-participating practices indicated very minor differences in organisational characteristics and measures of clinical care quality (Lyratzopoulos *et al*, 2013a). These findings support the representativeness of the NACDPC data.

The feasibility of examining the presenting symptoms of the NACDPC population has been demonstrated by several publications, albeit for a small number of cancers (Hughes *et al*, 2015;

[1] Of these publications, three were published during the duration of this PhD (2015–18) (Hughes *et al*, 2015; Tørring *et al*, 2017; Ozawa *et al*, 2018).

Ozawa *et al*, 2018), or otherwise as a confounding factor (Lyratzopoulos *et al*, 2013a; Rubin *et al*, 2015), with limited examination as an exposure of interest for diagnostic timeliness. As mentioned in Chapter 1 (Introduction), substantial variation in the patient and primary care interval has been described using the NACDPC data (Keeble *et al*, 2014; Lyratzopoulos *et al*, 2015); this has been hypothesised to be attributable to different symptomatic presentations. The empirical enquiries presented in this thesis therefore represent novel and original analyses while building on previously established research.

4.1.3 Data collection

The NACDPC initiative collected information on the diagnostic pathway of incident cancer patients in primary care (Rubin *et al*, 2011). Auditors (mainly general practitioners (GPs) and other health professionals) in participating primary care practices completed a questionnaire for incident cases of cancer during the audit period, using information from primary care records, also including hospital correspondence. Guidance notes were provided within the Excel spreadsheet proforma to aid completion of the questionnaire. A financial incentive of £500 per practice was available for participating practices. Screen-detected cancers, in-situ carcinomas, and non-melanotic carcinomas of the skin were excluded from the NACDPC data collection a priori.

A total of 1,170 general practices participated, representing around 14% of all practices in England, and provided information on approximately 18,879 cancer patients. Data were collected in pre-formatted Excel spreadsheets, and reviewed by the practice team before submission. Anonymous data on practice-level audits were then collated at the level of the former Cancer Networks, and stored by the NCIN under information governance arrangements as applicable to cancer registries. By design, it was not possible to link the NACDPC data to information in cancer registration or any other routine data records. After assessment of the NACDPC data as anonymous and non-disclosive, the dataset was made available by the NCIN without research ethics requirements.

4.2 Measuring presenting symptoms

4.2.1 Available information in the NACDPC

The audit collected free-text information on presenting symptoms in response to the question "*what was the main presenting symptom*?" for each patient.

I manually coded the data based on a 'bottom up' (or 'data driven') approach informed by principles of natural language processing (NLP) (Nadkarni *et al*, 2011). Other approaches to analysing the free-text information such as content analyses were not considered as the audit questionnaire was not considered to be a valid qualitative data gathering technique (Vaismoradi *et al*, 2013). Further, machine-based natural language processing was deemed unsuitable given

the relatively small size of the corpus of text, and the nature of the free-text (which included spelling mistakes, inconsistencies in grammar and punctuation, and non-standard word order) which would have required manual coding into the synonymous entities (Savkov *et al*, 2016).

Categories were generated with no *a priori* restrictions or definitions, and assignment of symptom categories was initially based on inspection of the free-text field only, with no knowledge of the cancer diagnosis or demographic details such as age and sex. Rather than assigning records to pre-determined short symptom lists, this approach enabled greater appreciation of the symptom spectrum, in particular by allowing a more refined understanding of rarer symptoms, and rarer variants of common symptoms. The following sections describe symptom coding and validation in greater detail.

4.2.2 Symptom coding

Firstly, I extracted the patient pseudo-ID number and free-text symptom field entries from a Stata dataset into Microsoft Excel 2013. The records were sorted in alphabetical order in order to easily code commonly recurring categories with identical symptom entries e.g. 'breast lump' or 'rectal bleeding'. A further coding wave grouped phenotypic expressions of the same symptom construct, including spelling or grammatical errors (e.g. 'haematuria', 'hematuria', 'bleeding per urethra'…) into single symptom entities. The remaining patients with multiple symptoms, rarer symptoms or more complex symptom status entries were coded individually, by assigning them to already generated categories or creating new additional symptom categories as required, with advice from GL and GPR[2]. This generated over 150 initial symptom categories.

After this initial approach, several logic rules were applied to the data to enable the fine-tuning of categorisation, based on further information in the patient record such as cancer type (see Appendix 4.1 for rules and justification). For example, an unspecified "bleed" symptom field entry in a woman subsequently diagnosed with cervical or endometrial cancer was assumed to represent vaginal bleeding, while in colorectal cancer patients it was assumed to mean rectal bleeding.

4.2.3 Symptom definitions and validation

Following the above, coding verification was done by GL and GPR for the common symptom categories with adjudication by consensus between GL, GPR, and myself, informed by the International Classification of Primary Care, Second edition (WHO, 2016). Where necessary, patients were re-classified (removal or addition of symptom categories) or new symptom categories were created, or the record was deemed invalid due to ambiguity. A total of 145

[2] GL: Professor Georgios Lyratzopoulos (principal supervisor); GPR: Professor Greg Rubin (collaborator)

symptom categories were created (see Appendix 4.2 for full list of symptom categories and definitions).

4.3 Measuring diagnostic timeliness

The NACDPC collected information on several measures of diagnostic timeliness, of which this thesis has focused on the patient interval, the primary care interval, and the number of pre-referral consultations. These variables were calculated by the NCIN after aggregation and made available as part of the audit dataset, therefore the following definitions are taken from the original audit report (Rubin *et al*, 2011).

The patient interval was defined as the number of days from first symptom to first primary care consultation. This was calculated based on the estimated date that the patient first developed their presenting symptom (first symptom) and date of first notification to any health care professional working within the Primary Health Care Team about a symptom or sign which was probably due to the cancer (first consultation), as recorded in the health record of the patient. For all analyses presented in this thesis, individuals with a patient interval of greater than 730 days were excluded, given that symptoms noted before this time are unlikely to be associated with the diagnosed cancer (Biswas *et al*, 2015).

The primary care interval was defined as the number of days from first primary care consultation to referral to specialist care. This was calculated based on date of first consultation and the date that the referral letter was sent, or if not available, the date the proforma was completed, or letter written (date of referral). The number of pre-referral consultations has been shown to be highly correlated with the primary care interval, and was therefore included as a measure of diagnostic timeliness in this thesis where appropriate (Lyratzopoulos *et al*, 2012). This was calculated based on the number of times the patient attended surgery before the date of referral.

4.4 Other variables of interest

The NACDPC collected information on patient characteristics such as age, sex, and ethnicity, and tumour characteristics such as cancer site and stage and were available as categorical variables in the NACDPC dataset.

The cancer diagnosis of each audited patient was determined as one of 27 specified cancers or as an 'Other' cancer based on information provided by auditors. The 'Other' cancer category comprised two groups: patients who had mostly been described as being diagnosed with an unspecified "Other [verbatim]" cancer (98%, 566/575 patients), or patients with a free-text entry describing a cancer that did not otherwise fit into the 27 specified cancer categories (2%, 9/575 patients). Based on the cancer incidence estimates in England in 2014, I estimated that

7% of cancers could be categorised as an 'Other' cancer based on the above definition (Office for National Statistics, 2016). The apparent under-recording of 'Other' cancers in the NACDPC may have been influenced by one of the 20 cancer networks placing greater emphasis on data collection for the four most common cancers (breast, bowel, lung, and prostate) across the general practices in its network.

Information on stage at diagnosis was provided in NACDPC based on information in the clinical records, and processed by the NCIN using an adaptation of the SEER LRD system (local, regional, or distant disease).

4.5 Data cleaning

The NACDPC dataset received from NCIN was cleaned before analysis to create the final sample population that could be analysed (see Figure 4.1 for sample derivation). This section describes patients that were excluded from analyses and other notable characteristics of the sample population.

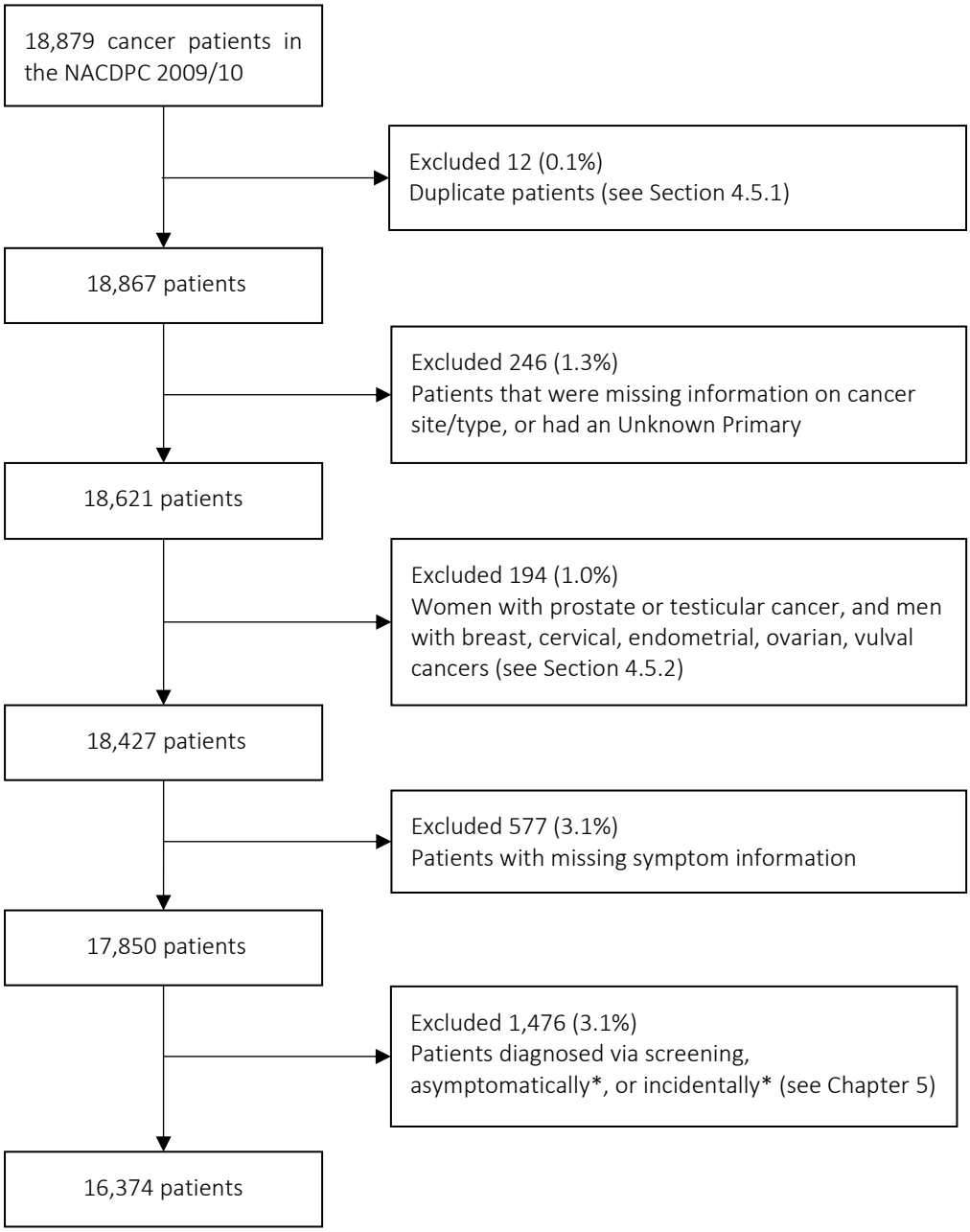


Figure 4.1 Flow chart indicating excluded populations common to all analyses included in this thesis.
*These patients have been examined separately in Chapter 5: Atypical diagnosis of cancer

4.5.1 Duplicate patient entries

There were a small number of patients who had identical free-text entries describing presenting symptoms, where matching wording or typographic or grammatical errors were suggestive of duplication (n=42 cases). I used other available variables in the dataset to compare patient

records as part of de-duplication, which resulted in the exclusion of 12 records (in one case, there were three identical entries and so two were removed). For the remaining 19 patients, there were conflicts across one or more variables, which were suggestive of errors generated by autofill functions, overwritten fields, copying-pasting, or other similar mechanisms in Excel during the data entry process rather than 'genuine' duplication of patient details. These records were therefore taken at face value and were maintained in the dataset as unique patients.

4.5.2 Exclusion based on cancer diagnosis

For all analyses, patients that were missing information on cancer diagnosis, had "No information/ Unknown Primary" stated as their cancer diagnosis, or erroneous or implausible cancers (e.g. women who had prostate or testicular cancer, and men who had cervical, endometrial, ovarian, vulval cancers) were excluded. Genuine cases of men with breast cancer could not be reliably distinguished from women with breast cancer that had been mistakenly recorded as being men and so these patients were excluded from all analyses.

4.5.3 Exclusion based on presenting symptoms

A proportion of patients (n=577; see Figure 4.1) were missing information on presenting symptoms (blank response to audit question regarding presenting symptoms), and were therefore excluded from all analyses.

Examination of the free-text information on presenting symptoms in the remainder of patients indicated that a small proportion of patients (n=78) were diagnosed through a population-based screening programme (for breast, cervical, colorectal cancers) in the absence of any symptoms; given that the audit had not intended to capture screen-detected cancer patients a priori, these patients were also excluded from all analyses.

For another group of patients, the free-text information on presenting symptoms indicated a range of atypical routes to cancer diagnosis. These patients were further examined in Chapter 5, and subsequently excluded from analyses focusing on the symptomatic cancer patient population in Chapter 6–9.

A proportion of patients had missing information on variables relating to diagnostic timeliness. Of the 16,374 patients who had complete information on cancer diagnosis and symptoms, 25% (4,148 patients) had missing patient interval, while 11% (1,869 patients) had missing primary care interval, and 14% (2,309) had missing information on number of pre-referral consultations. Some of this is likely to have been "appropriately" missing, given that if the patient was diagnosed incidentally during a routine healthcare appointment, or otherwise diagnosed in a place other than primary care, they will not have been expected to have had a valid patient or primary care interval (see Appendices 6.2, 7.2, and 8.1 corresponding to Chapter 6, 7, and 8 respectively). There was some overlap in missingness across the three variables; 891 patients (5.4% of the total patient population) had missing information for all three (see Figure 4.2).

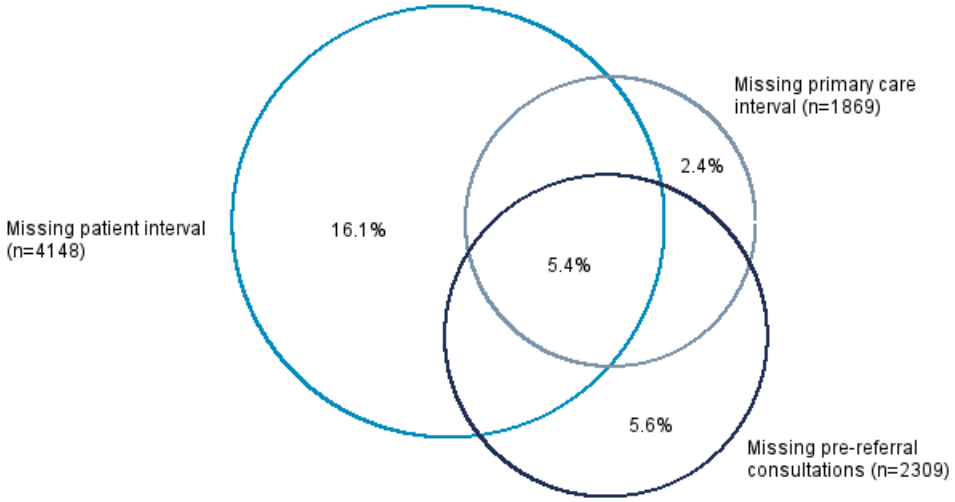


Figure 4.2 Venn diagram illustrating overlap of missingness for measures of diagnostic timeliness
Patient interval, primary care interval, and number of pre-referral consultations in the NACDPC patient population, n=16,374.

Additionally, a proportion of patients had missing information on age, sex, and ethnicity, and cancer stage. Missingness patterns of these variables are explored empirically in subsequent chapters individually, with further considerations of the possible effects on the findings and interpretation in section 4.6 below, and in Chapter 10 (Discussion).

4.5.4 Truncated symptom entries

There were a small number of patients for whom the free-text entries describing symptoms appeared to be truncated (n=24). This is likely to have occurred during data aggregation and generation of the dataset, which was implemented in Stata 11 (Rubin *et al*, 2011). Older versions of Stata (before Stata 13) limit the value of within-cell string variables to 244 characters. Having verified with NCIN that the truncated information was not retrievable, these patient records were retained, and coded based on existing information.

4.6 Discussion

The NACDPC dataset is a large and valid source of data that has already been used to examine aspects of diagnostic timeliness, which means it is ideally placed for investigating the aims and objectives of this thesis. This section reflects on the strengths and limitations of the NACDPC dataset.

Firstly, the audited patient population is a large and representative cohort of cancer patients, which supports the external validity of the analyses. Although audit participation was voluntary, the NACDPC data has been shown to be representative both at individual and general practice level (Rubin *et al*, 2011; Lyratzopoulos *et al*, 2013a).

Building on the arguments of Chapter 3 (Literature review), it is clear that collecting information on the presenting symptoms of cancer patients is challenging. While study designs collecting patient-reported data may have offered a more accurate picture of the symptomatic experience, the risk of recall and survival bias may have been higher relative to the records-based approach utilised in the NACDPC. Other prospective designs collecting patient-reported data before diagnosis are more resource intensive, and therefore are unlikely to have obtained a sample size comparable to that of the NACDPC patient population.

Information on the presenting symptoms of cancer patients as collected as part of the NACDPC is not only dependent on the accurate elicitation and recording of symptoms in the primary care record as discussed in Chapter 3, but could be further prone to biases associated with the process of data extraction for the audit. On the other hand, manual case note review by healthcare professionals may have contributed to higher quality data collection than examining routinely collected primary care data, as auditors were able to access both coded and free-text information in the health record of each patient and may have also been the attending physician. The face validity of the NACDPC and information on symptoms is supported by the publication of several peer-reviewed studies that used symptom information from the NACDPC independent to this thesis (Lyratzopoulos *et al*, 2013a; Hughes *et al*, 2015; Rubin *et al*, 2015; Ozawa *et al*, 2018).

The free-text information on presenting symptoms was coded using a 'data driven' approach, without any a priori symptom classification system and with minimal aggregation of symptom categories. This allows for better appreciation of the granular nature of symptom information, particularly in the cases of rare symptoms, or rarer variants of common symptoms. This also enables flexible aggregation of symptoms dictated or necessitated by clinical, epidemiological, and statistical considerations, as demonstrated in subsequent chapters.

The length of the patient and primary care interval was estimated for each patient based on the estimated dates of symptom onset, first relevant presentation, and first referral to specialist services. Determination of these dates based on information in health records could have introduced bias given that pinpointing date of onset may be harder for certain (vague and non-specific, or intermittent) symptoms; this could have contributed to the levels of missingness associated for each of the intervals. However, the levels of missingness described in the NACDPC population are comparable to that of other studies that have examined symptoms and intervals as reviewed in Chapter 3 (Pruitt *et al*, 2013; Walter *et al*, 2015, 2016b, 2016a; Leiva *et al*, 2017). Further, given the poor availability of information on the patient interval in other routine data sources, this is a relative strength of the NACDPC dataset.

Validation of information on other variables of interest (such as stage at diagnosis) in the NACDPC dataset was not possible due to the study design. While noting that the availability of

staging data exceeds that of national levels of contemporaneous staging information, this has been further discussed and reflected upon in Chapter 9 where stage at diagnosis is the outcome of interest. The limitations outlined above apply to all studies included in this thesis; therefore the implications of these will be taken into consideration within each study chapter, and will also be subject to further scrutiny and discussion in Chapter 10 (Discussion).

In conclusion, at the start of this PhD in 2015, the NACDPC dataset represented an ideal source for examining the epidemiology of presenting symptoms among cancer patients. At submission in 2018, it remains the largest cancer patient population for which information on both presenting symptoms, and diagnostic intervals (particularly the patient interval) is available according to my knowledge. In addition to enabling the study of the association between individual symptoms and measures of diagnostic timeliness, the NACDPC dataset offers the opportunity to study potential associations adjusting for potential confounders including age, sex, cancer site and stage at diagnosis.

Chapter 5: Atypical diagnosis of cancer

Some patients are diagnosed with cancer in the absence of tumour-related symptoms, outside of population-based screening. Prior to exclusion from subsequently presented empirical chapters, this chapter investigates the characteristics of this atypically diagnosed cancer patient population.

Aspects of this chapter have been presented at the Preventing Overdiagosis 2017 conference in Quebec[1].

[1] Koo MM, Rubin G, Lyratzopoulos G (2017) **Common pathways to incidental diagnosis of cancer beyond screening: insights from a national audit of cancer patients in England.** Oral presentation (15 minutes)

5 Atypical diagnosis of cancer

5.1 Rationale for this chapter

During the process of coding the presenting symptoms of the NACDPC cancer patient population (see Section 4.5.3 of Chapter 4), I identified patients who did not present with symptoms relevant to their subsequently diagnosed cancer. I examined these patients before exclusion from analyses relating to symptomatic cancer patients, and noted that this heterogeneous patient group experienced a variety of other routes to diagnosis which are poorly described in the literature.

Measures of diagnostic timeliness (which are predicated on dates including that of symptom onset, first presentation, and referral relating to suspected cancer (Weller *et al*, 2012)) are largely inapplicable and less meaningful among cancer patients who have been diagnosed through an atypical route. Consequently, atypically diagnosed cancer patients are therefore (by virtue of how they were diagnosed) unlikely targets or beneficiaries of early diagnosis interventions. Further examination of these patients acknowledges the complexity of the diagnostic process of cancer however, and could nevertheless be worthy for other aspects of health service improvement, for example strategies to manage individuals at greater risk of cancer, together with the examination of prognostic, psychological and economic implications of atypically diagnosed cancers.

This chapter therefore aimed to examine the characteristics of this patient group prior to exclusion, and describe the circumstances preceding cancer diagnosis where further information was available.

5.2 Introduction

Most patients with cancer are diagnosed following presentation to primary care with symptoms caused by the malignancy, while a proportion are diagnosed asymptomatically through screening programmes (Elliss-Brookes *et al*, 2012; Jensen *et al*, 2014). Anecdotal evidence suggests that cancer may be diagnosed incidentally, while cancer may also be detected through the surveillance of high-risk states (e.g. pre-malignant conditions, or family history of cancer) (Dove-Edwin, 2005; Ong *et al*, 2014; Cufari *et al*, 2016; Ramsay, 2017).

However, epidemiological literature describing the frequency and context of such diagnoses is limited. Incidental diagnosis of important disease, including cancer, can result from the use of imaging technologies (predominantly in secondary care settings) but may also result from blood or urine analysis and physical examination (Lumbreras *et al*, 2010; Avilés-Izquierdo *et al*, 2016; Kroczek *et al*, 2017).

The incidental diagnosis of cancer may represent a fortuitous finding of an otherwise invasive tumour at an earlier stage, potentially resulting in clinical benefit for the patient. However, it may also represent potential overdiagnosis, whereby the detected cancer would not have become symptomatic in the patient's lifetime (Welch & Black, 2010). Thus far, concerns about cancer overdiagnosis have largely been focused on screen-detected cancers among asymptomatic individuals (such as breast and lung cancer screening), but there is increasing awareness of the challenging balance between earlier diagnosis and overdiagnosis in contexts other than screening (Biswas *et al*, 2015; Jenniskens *et al*, 2017; Nicholson, 2017). Empirical evidence regarding the frequency and characteristics of incidentally diagnosed cancer patients could be useful before detailed consideration of potential overdiagnosis in subsequent epidemiological studies.

Against this background, I aimed to examine the frequency of atypically diagnosed cancer in a representative cohort of patients; describe the characteristics of patients diagnosed in this way; and elucidate common pathways and mechanisms resulting in such a diagnosis.

5.3 Methods

5.3.1 Study population

The analysis sample comprised 14,082 cancer patients included in the NACDPC with complete and valid information on age (among patients aged fifteen years or older), sex, and symptom status (see Figure 5.1 for flow chart of sample derivation).

Patients determined to have been detected via national screening programmes were excluded given that the audit had been designed to exclude these patients *a priori* (n=78). Patients with missing information in response to the audit question on presenting symptoms were also excluded (n=577; see Section 4.5.3 of Chapter 4, page 63).

Patients diagnosed with prostate cancer were excluded from analyses, given the difficulties in reliably interpreting and determining their symptom status (see below). Similarly, patients diagnosed with undefined 'other' cancers were also excluded.

5.3.2 Atypically diagnosed cancer: case definition

Informed by previous literature, I defined the atypical diagnosis of cancer as the diagnosis of cancer in an individual who was either asymptomatic before diagnosis (and not participating in population-based screening programmes), or whose presenting symptoms could not plausibly be related to their subsequent diagnosis (Davies *et al*, 2010; Kocher *et al*, 2016).

Cancer patients who were diagnosed following participation in population-based screening programmes were not included, as the audit was designed a priori to exclude these patients, and therefore they represented anomalies in the final resulting NACDPC dataset (see Section 4.1.3). However, those for whom information collected in the audit indicated diagnosis

following surveillance were included in order to provide a more complete picture of patients excluded from the subsequently presented empirical analyses (Chapters 6–9).

The nature of cancer diagnosis (atypical or typical), was ascertained by examination of the free-text information on presenting symptoms for each patient, which could also include information on circumstances surrounding the diagnosis. Cases were operationally defined as having had an atypical diagnosis of cancer if information in the free-text field describing the presenting symptoms indicated:

1. That the diagnosis was incidental, described as such explicitly (verbatim) or implicitly (as suggested by contextual details);
2. That the patient had no symptoms before diagnosis, with further contextual information (e.g. surveillance of a different cancer);
3. That the patient had no symptoms before diagnosis, without further contextual information; or
4. Use of a test without any information on symptoms prompting its use.

For the subgroups of patients with further relevant information on the context of the cancer diagnosis (i.e. 1 and 2 in the above list), I was able to examine cases at a greater level of resolution and identify common scenarios preceding atypical diagnosis (n=422, 50% of all atypical diagnoses), as described in the next section.

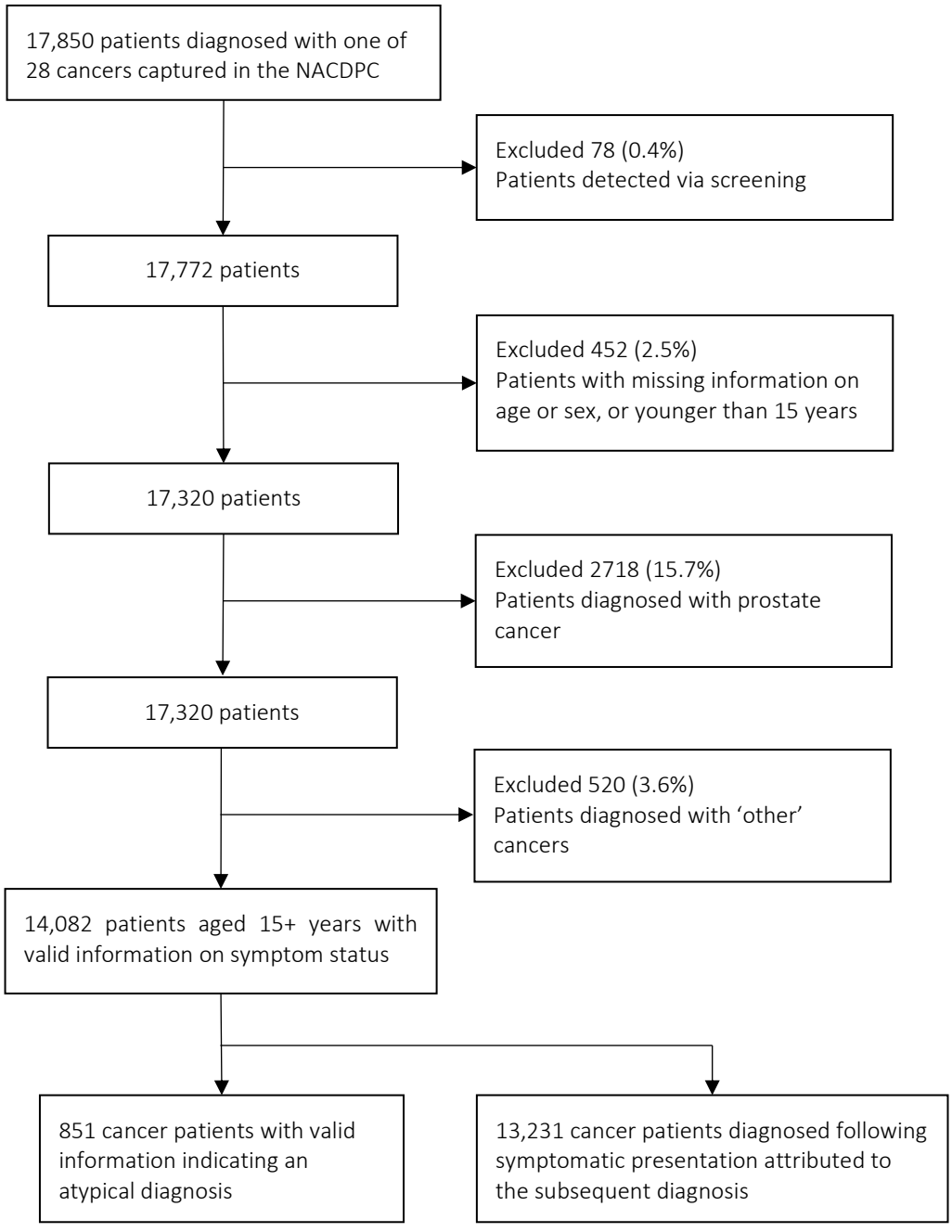


Figure 5.1 Flow diagram of sample derivation for the study population of this chapter

5.3.3 Statistical analyses

Firstly, I described the demographic and clinical characteristics of atypically diagnosed cancer patients. Logistic regression was used to calculate crude and adjusted odds ratios of atypical diagnosis (versus 'typical symptomatic diagnosis') by sex, age group, and cancer site. Colorectal cancer was used as the reference category for cancer type, as it was the most common non-gender specific cancer in the study population. I also examined the relative frequencies of cancer sites among the atypically diagnosed patient population.

Subsequently, I described common clinical scenarios leading to an atypical cancer diagnosis based on a subgroup of patients that had further relevant information detailing the circumstances prior to diagnosis available (n=422, 50% of all atypical diagnoses). I synthesised this additional information narratively.

5.3.4 Supplementary analyses

5.3.4.1 Including ethnicity as a covariate

In order to explore potential confounding of the association between sex, age, and cancer site and odds of atypical diagnosis by ethnicity, I repeated the analysis adjusting for ethnic group among the patients for whom ethnicity information was available (among 12,446/14,082 (88%) of the population). Findings are presented in Appendix 5.1.

5.4 Results

5.4.1 Frequency and characteristics of atypically diagnosed cancer

I identified atypical cancer diagnosis in a total of 851/14,082 (6%) patients aged 15+ years, who had been diagnosed with one of 26 cancer sites (other than prostate cancer). Men were more likely to be diagnosed atypically than women (8% of men versus 5% of women, adjusted OR (95% CI) for women: 0.8 (0.7–1.0)) (see Table 5.1). The odds of an atypical versus typical diagnosis with cancer increased with age, although there was no evidence for a difference between older age groups (joint Wald test p-value for overall variation between age categories <0.001). Sensitivity analyses examining ethnicity produced comparable findings (Appendix 5.1).

Atypical diagnosis was seen in a third (33%) of leukaemia patients, almost a quarter (24%) of all liver cancer patients, and more than a tenth of renal cancer (17%), myeloma (16%), and thyroid cancer (14%) patients. In contrast, less than 1% of patients with testicular, cervical, or brain cancers were diagnosed atypically. Crude and adjusted odds ratios indicated substantial variation in the odds of atypical diagnosis between cancer sites (see Figure 5.2 and Table 5.1).

Table 5.1 Characteristics of patients with an atypical cancer diagnosis versus cancer patients with a typical cancer diagnosis, with crude/adjusted odds ratios

	Total	Atypically diagnosed cancer		Crude (n=14082)	Adjusted[2] (n=14082)
	N	n	% (95% CI)	OR (95% CI)	OR (95% CI)
Total	14,082	851	6% (6–6%)	-	-
Sex					
Men	5983	454	8% (7–8%)	Ref.	Ref.
Women	8099	397	5% (4–5%)	**0.6 (0.6–0.7)**	**0.8 (0.7–1.0)**
Joint Wald test P-value	-	-	-	**<0.001**	**0.012**
Age group					
15–49 years	2089	53	3% (2–3%)	**0.4 (0.3–0.5)**	**0.4 (0.3–0.5)**
50–59 years	2080	107	5% (4–6%)	**0.8 (0.6–1.0)**	0.8 (0.6–1.0)
60–69 years	3264	215	7% (6–7%)	Ref.	Ref.
70–79 years	3739	268	7% (6–8%)	1.1 (0.9–1.3)	1.1 (0.9–1.3)
80+ years	2910	208	7% (6–8%)	1.1 (0.9–1.3)	1.1 (0.9–1.4)
Joint Wald test P-value	-	-	-	**<0.001**	**<0.001**
Cancer site					
Leukaemia	511	169	33% (29–37%)	**10.3 (7.9–13.5)**	**11.4 (8.7–14.9)**
Liver	116	28	24% (17–33%)	**6.7 (4.2–10.6)**	**6.6 (4.1–10.6)**
Renal	373	65	17% (14–22%)	**4.4 (3.2–6.1)**	**4.6 (3.3–6.4)**
Myeloma	241	39	16% (12–21%)	**4.0 (2.7–6.0)**	**4.0 (2.7–5.9)**
Thyroid	113	16	14% (9–22%)	**3.4 (2.0–6.0)**	**5.1 (2.9–9.2)**
Gallbladder	68	5	7% (3–16%)	1.7 (0.7–4.2)	1.8 (0.7–4.5)
Bladder	869	58	7% (5–9%)	**1.5 (1.1–2.1)**	**1.4 (1.0–2.0)**
Lung	1913	118	6% (5–7%)	**1.4 (1.1–1.8)**	**1.4 (1.0–1.8)**
Lymphoma	704	40	6% (4–8%)	1.3 (0.9–1.8)	**1.5 (1.0–2.1)**
Vulval	73	4	5% (2–13%)	1.2 (0.4–3.4)	1.4 (0.5–4.0)
Mesothelioma	75	4	5% (2–13%)	1.2 (0.4–3.3)	1.1 (0.4–3.0)
Colorectal	2431	111	5% (4–5%)	Ref.	Ref.
Melanoma	839	35	4% (3–6%)	0.9 (0.6–1.3)	1.1 (0.8–1.7)
Stomach	303	11	4% (2–6%)	0.8 (0.4–1.5)	0.8 (0.4–1.5)
Ovarian	398	14	4% (2–6%)	0.8 (0.4–1.3)	0.9 (0.5–1.7)
Breast	2717	89	3% (3–4%)	**0.7 (0.5–0.9)**	0.9 (0.7–1.3)
Pancreatic	374	11	3% (2–5%)	0.6 (0.3–1.2)	0.6 (0.3–1.2)
Endometrial	413	11	3% (1–5%)	0.6 (0.3–1.1)	0.7 (0.4–1.3)
Laryngeal	121	3	2% (1–7%)	0.5 (0.2–1.7)	0.5 (0.2–1.7)
Oropharyngeal	213	5	2% (1–5%)	0.5 (0.2–1.2)	0.6 (0.2–1.4)
Small Intestine	53	1	2% (–10%)	0.4 (0.1–2.9)	0.4 (0.1–3.1)
Sarcoma	107	2	2% (1–7%)	0.4 (0.1–1.6)	0.5 (0.1–2.0)
Oesophageal	567	8	1% (1–3%)	**0.3 (0.1–0.6)**	**0.3 (0.1–0.6)**
Brain	215	2	1% (–3%)	**0.2 (0.05–0.8)**	**0.2 (0.1–0.9)**
Cervical	126	1	1% (–4%)	0.2 (0.02–1.2)	0.3 (0.03–1.8)
Testicular	149	1	1% (–4%)	0.1 (0.02–1.0)	0.3 (0.04–2.0)
Joint Wald test P-value	-	-	-	**<0.001**	**<0.001**

[1] Adjusted for sex, age group, and cancer site. For sex-specific cancers (breast, cervical, endometrial, ovarian, and testicular cancer), the odds ratio refers to the comparison with a patient of the same sex with the reference cancer (colorectal cancer).

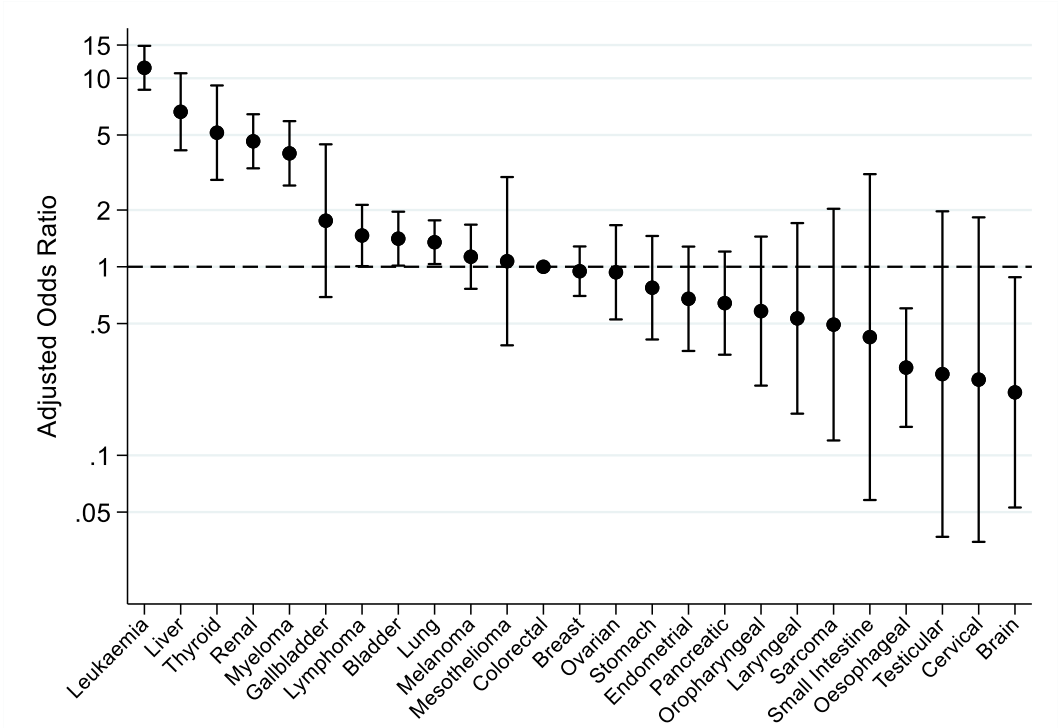


Figure 5.2 Adjusted odds ratios for atypical versus typical diagnosis by cancer site

5.4.2 Cancer site case-mix of atypically diagnosed cancer

Among the 851 atypically diagnosed patients, a fifth (20%, 17–23%) were diagnosed with leukaemia, while other common cancer sites included lung (14%), colorectal (13%), and breast cancers (10%) (see Figure 5.3). There were 11 other cancer sites represented among the atypically diagnosed cancer patient population with 10 or more patients each.

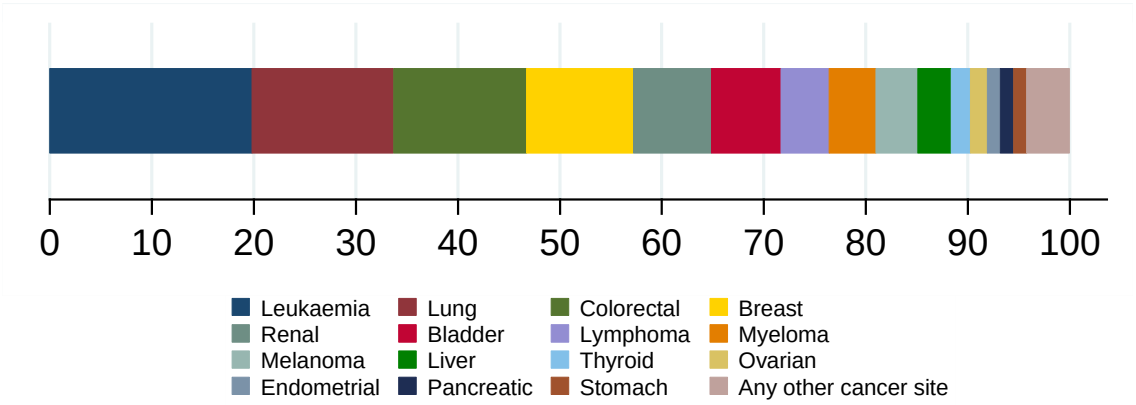


Figure 5.3 Common cancer sites among the atypically diagnosed cancer patient population
Cancers with 10 or more cases specified only; see Appendix 5.2 for exact frequencies.

5.4.3 Clinical circumstances leading to atypically diagnosed cancer

Among a subgroup of cancer patients with more detailed information, I identified three common clinical scenarios frequently preceding the atypical diagnosis of cancer (n=422, 50% of all such patients) (see Figure 5.4).

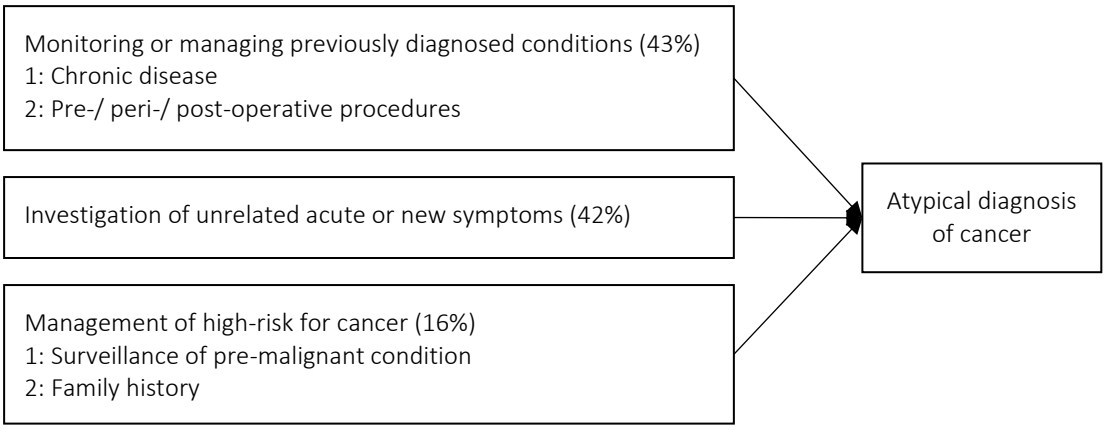


Figure 5.4 Routes to atypically diagnosed cancer
Based on cancer patients with sufficient information, n=422.

5.4.3.1 Monitoring or managing previously diagnosed conditions

43% of patients in the subgroup were diagnosed with cancer following a clinical encounter for a known condition. This included routine blood or urine testing, as part of chronic disease (or related risk factor) monitoring, revealing abnormalities that led to the diagnosis of unsuspected cancer (e.g. microscopic haematuria on dipstick urine testing), or procedures relating to elective surgery for unrelated indications, such as blood or imaging investigations. For some patients in this group, cancer diagnosis signified the detection of a second primary malignancy detected as part of follow up investigation for a pre-existing cancer.

5.4.3.2 Investigation of an unrelated acute or new symptom

Some patients were serendipitously diagnosed with cancer during a healthcare encounter due to unrelated new symptoms (e.g. irregular moles, lumps, or lymph nodes) or as a result of investigations of presenting symptoms unlikely to be related to the subsequent cancer diagnosis (e.g. an MRI scan investigating back pain identifying ovarian cancer).

5.4.3.3 Management of high-risk patients

Some patients were diagnosed with cancer during the management of high-risk for malignancy. Examples include detection of colorectal cancer in the context of colonoscopic surveillance of previously diagnosed polyps, and oesophageal cancer detection during endoscopic monitoring of Barrett's oesophagus. In a few patients, cancer was detected following investigations for known familial risk of cancer (e.g. breast cancer in a close relative).

5.5 Discussion

5.5.1 Main findings

Atypical diagnosis of cancer was observed in more than 1 in 20 cancer patients in the study population. Older patients and men, and patients with leukaemia, liver, renal, multiple myeloma, and thyroid cancer were more likely to be diagnosed in this way compared to diagnosis with relevant symptoms. Three common clinical scenarios that preceded an atypical diagnosis of cancer were: 1) management of prior/chronic conditions; 2) new/acute conditions unrelated to cancer; and 3) assessment of known high-risk status.

5.5.2 Comparison with prior evidence

Literature examining atypical routes to cancer diagnosis is limited. A few hospital-based studies have examined clinical scenarios surrounding incidental or asymptomatic diagnosis but relate to individual cancer sites (Avilés-Izquierdo *et al*, 2016; Cufari *et al*, 2016; Kocher *et al*, 2016). Some evidence of relevance can be gleaned from studies reporting incidental findings of cancer or other diseases detected in the context of research, but estimates of 'clinically important' findings vary substantially depending on imaging field (whole body, or specific organ) and modality, and research participants may not be representative of the general population (Booth *et al*, 2010; The Royal College of Radiologists, 2011). In comparison with these previous studies, I examined the incidental diagnosis of cancer (alongside other atypical diagnoses of cancer) in a population-based incident cohort comprising patients with one of 26 common and rarer cancers, which enabled comparison of relative probability for atypical diagnosis between cancer sites.

5.5.3 Strengths and limitations

Information on the circumstances surrounding diagnosis is not routinely recorded as part of cancer registration, healthcare records, or other administrative databases. Against this background, and given the paucity of prior relevant evidence, a strength of this analysis is that it provides evidence about these less well documented pathways to cancer diagnosis among a large and representative cohort of incident cancer patients.

Correspondingly however, interpretation of the findings should be mindful of how atypical diagnoses were defined. The aims and objectives of this thesis are relevant to cancer patients who presented with (plausibly tumour-related) symptoms before diagnosis. This chapter intended to describe all cancer patients who did not fit into this category, and therefore by definition, describes a very heterogeneous population including patients diagnosed following surveillance (for whom a cancer diagnosis is not unexpected), as well as those diagnosed serendipitously.

Further, the audit question from which symptom status was derived was originally designed to collect free-text information on presenting symptoms. Therefore, the findings are reliant on accurate and complete extraction of information on symptoms from health records by auditors.

For instance, patients for whom the response to the audit question mentioned an investigation alongside further contextual information (indicating that it did not relate to symptoms of the subsequently diagnosed cancer) were assumed to have been atypically diagnosed. I assumed that atypical diagnosis had also occurred in patients for whom investigation was mentioned *without* further information about the circumstances prompting its use. Among this latter group of patients, it is possible that investigations may have been triggered by symptomatic presentation which was not noted in the patient's primary care record at point of presentation, or otherwise not extracted for the audit. Therefore, assuming these patients were asymptomatic may have led to the overestimation of the overall frequency of atypically diagnosed cancer.

Nevertheless, for half of all atypically diagnosed patients, case definition was ascertained by additional information regarding the circumstances leading to the diagnosis of cancer. If I had considered only these patients, I may have underestimated the true frequency of atypically diagnosed cancer.

5.5.4 Implications

The findings elucidate atypically diagnosed cancer, which occurred in a substantial minority of cancer patients captured in the NACDPC.

Older men were more likely to be diagnosed with cancer in an atypical route, possibly reflecting greater morbidity burden leading to more frequent investigations in this patient group (Higashi *et al*, 2007).

Notable proportions of patients with thyroid and renal cancer, and melanoma were diagnosed in the absence of tumour-related symptoms. This is consistent with prior evidence indicating potential overdiagnosis of these cancers, although I was unable to directly examine this research question given the cross-sectional nature of the NACDPC data (Welch & Black, 2010; Moynihan *et al*, 2012; Weyers, 2012; Ahn *et al*, 2014). Comparison of tumour characteristics and longer-term clinical outcomes between patients diagnosed incidentally and those diagnosed symptomatically should be addressed by further research.

A proportion of patients were diagnosed with cancer following monitoring of high-risk for malignancy. For these patients, the diagnosis of cancer is an anticipated possible outcome of surveillance regimens. However, for other patients who were diagnosed incidentally, the diagnosis of cancer is likely to be unexpected. Evidence from cancer screening suggests that unexpected findings may be associated with anxiety and distress in some patients; additional

support may be beneficial for patients who are incidentally diagnosed with cancer (Davies *et al*, 2017; Gibson *et al*, 2017).

The relative frequencies of clinical scenarios that preceded cancer diagnosis are likely to be affected by physician and health system level factors such as approaches to chronic disease monitoring, incentives and thresholds for investigation, and elective surgery rates (Pollack *et al*, 2017). Furthermore, the increasing prevalence of multi-morbidities in the context of an aging population, and increasing cancer survivorship (with accompanied increase incidence of second primary cancer diagnoses) suggest that increasing proportions of cancer patients may be diagnosed in this way (Murphy *et al*, 2017).

5.6 Chapter summary

Not all cancer patients are diagnosed following presentation with symptoms caused by the tumour or malignancy. This study provides evidence about the frequency and nature of previously under-appreciated circumstances preceding cancer diagnosis. The findings indicate that the frequency of atypical diagnosis (including incidental findings of cancer) is substantial, and that it varies by demographic characteristics and cancer site. Having characterised this group prior to exclusion, the following chapters of this thesis focus on patients who present with symptoms related to their subsequently diagnosed cancer.

Chapter 6: The symptom signature of breast cancer and associated diagnostic intervals

This chapter examines the nature and frequencies of presenting symptoms (namely the symptom signature) of women who were subsequently diagnosed with breast cancer, and variation in the patient and primary care interval by symptom.

Aspects of this chapter have been the subject of a peer-reviewed publication in Cancer Epidemiology[1].

[1] Koo MM, Wagner C von, Abel G, McPhail S, Rubin G, Lyratzopoulos G (2017) **Typical and atypical symptoms in women with breast cancer: Evidence of variation in diagnostic intervals from a national audit of cancer diagnosis.** *Cancer Epidemiology* 48 140–146, http://dx.doi.org/10.1016/j.canep.2017.04.010 (see Appendix 6.1)

6 The symptom signature of breast cancer and associated diagnostic intervals

6.1 Rationale of this chapter

This chapter serves as an exemplar of examining the nature and frequencies of presenting symptoms (namely the symptom signature) associated with a cancer site, and how different symptoms may be related to diagnostic timeliness.

I chose to focus on breast cancer patients for several reasons. Breast cancer is a common cancer diagnosed in around 150 women a day in the UK; therefore understanding rarer symptoms and how these relate to diagnostic timeliness remains relevant for the earlier diagnosis of a considerable number of patients (Cancer Research UK, 2017). Secondly, the literature review in Chapter 3 characterised the symptom signature of breast cancer as narrow. For such cancers, there is a need to determine the nature and frequency of rarer atypical symptoms which are less well documented in the literature (Walker *et al*, 2014; Redaniel *et al*, 2015). Finally, several qualitative or otherwise small scale studies indicated that compared to help-seeking for breast lump, help-seeking for non-lump symptoms is often delayed, but this had not yet been demonstrated among women diagnosed with breast cancer in a large and representative sample and was therefore deemed worthy of examination (Burgess *et al*, 1998; Ramirez *et al*, 1999; Webber *et al*, 2017).

6.2 Introduction

The findings of Chapter 3 (Literature review) identified breast lump as a symptom with relatively high predictive value for malignancy, and the most common presenting symptom among women with breast cancer (Walker *et al*, 2014; Redaniel *et al*, 2015). Indeed, public health education campaigns have long focused on raising awareness of breast lump in the general population (Roberts *et al*, 1984; Janz *et al*, 1989). Although women with breast cancer typically experience short diagnostic intervals compared to other cancer patients, some women continue to experience long diagnostic intervals (Baughan *et al*, 2009; Hansen *et al*, 2011; Lyratzopoulos *et al*, 2013b; Neal *et al*, 2014; Redaniel *et al*, 2015). This is concerning as longer intervals to diagnosis have been shown to be associated with lower five-year survival in women with breast cancer (Richards *et al*, 1999). Additionally, inequalities in stage at diagnosis and survival of breast cancer patients have been linked to variation in the length of the patient interval (Lyratzopoulos & Abel, 2013; Rutherford *et al*, 2013; Marcu *et al*, 2016; Romanoff *et al*, 2017).

Prior literature exploring reasons for delayed help-seeking for symptoms of the breast suggests that non-lump breast symptoms may often be attributed to other non-malignant causes such

as hormonal changes, trauma, or breastfeeding (Ramirez *et al*, 1999; O'Mahony *et al*, 2013; Khakbazan *et al*, 2014). While this provides an explanation of why some women may experience long intervals to presentation, the literature reviewed in Chapter 3 indicated that there is a lack of evidence quantifying diagnostic timeliness in large representative samples of women with breast cancer. Furthermore, existing studies examining the presenting symptoms of breast cancer often dichotomised symptoms based on the presence or absence of breast lump, greatly limiting the appreciation of the large spectrum of presenting symptoms within the heterogeneous 'non-lump' breast symptoms category (Burgess *et al*, 1998; Friedman *et al*, 2006; Redondo *et al*, 2009; Innos *et al*, 2013; Poum *et al*, 2014).

Following the above considerations, I aimed to describe the diverse range of presenting symptoms in a large representative sample of women with breast cancer included in the NACDPC, and to examine associations between different symptomatic presentations and the length of diagnostic intervals.

6.3 Methods

6.3.1 Study population

The analysis sample comprised 2,316 breast cancer patients included in the NACDPC with complete and valid information on age, ethnicity, and presenting symptoms (see Figure 6.1for sample derivation). As described in Chapter 4 (Data & methods), individuals diagnosed incidentally or asymptomatically were excluded from the analysis.

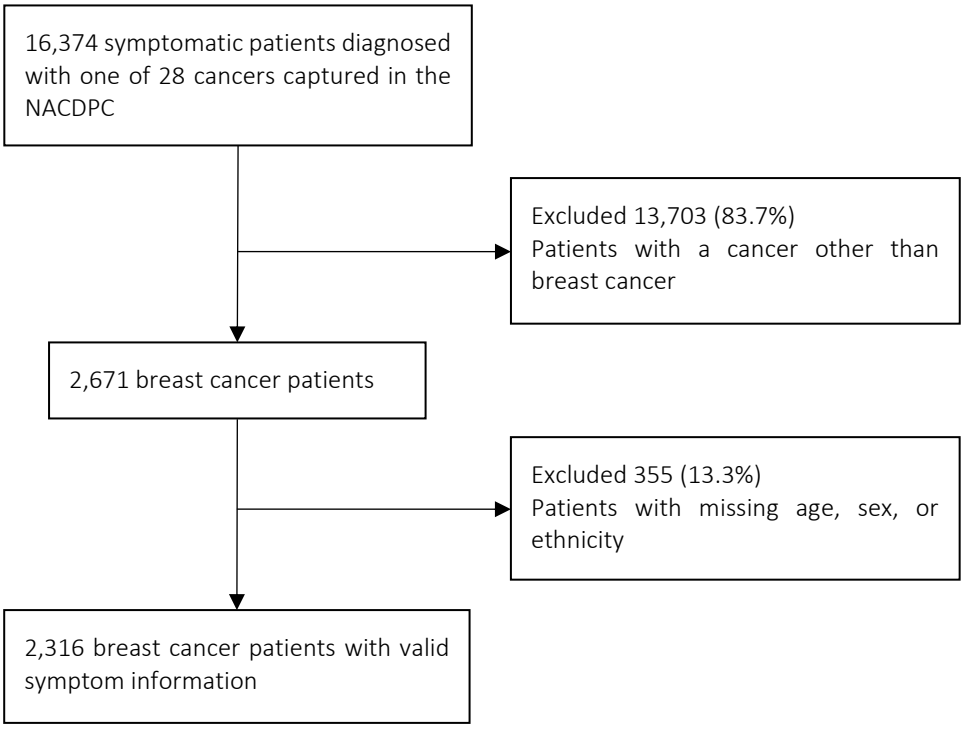


Figure 6.1 Flow diagram indicating sample derivation for this chapter
See Section 4.5 in Chapter 4 for further information.

6.3.2 Outcomes of interest

The patient interval and primary care interval were measured in days in the NACDPC and treated as continuous variables. As indicated in Chapter 4 (Data & methods), the number of pre-referral consultations was also examined as a measure of diagnostic timeliness (Lyratzopoulos *et al*, 2013b). In this analysis, it was parameterised as a binary outcome (1 pre-referral consultation versus 2 or more pre-referral consultations) as the vast majority of patients (90%) had one consultation.

Among these women, 1,883 (81%), 2,201 (95%), and 2,002 (86%) had complete information on the patient interval, the primary care interval, and the number of pre-referral consultations respectively. Women with missing interval or pre-referral consultation data were more likely to be older (70 years or over) and less likely to have presented in general practice, without evidence for variation by ethnicity, symptom category, or number of symptoms (Appendix 6.2). Given the above, complete case analysis was deemed appropriate.

6.3.3 Statistical analyses

Firstly, I described the frequencies of recorded presenting symptoms using exact confidence intervals, and the distribution of the patient and primary care intervals for each symptom among women with complete interval values. The mean, median and selected centile interval values were summarised and reported, together with the proportion of women that experienced 2 or more pre-referral consultations by symptom category. Additionally, the proportion of women with interval values exceeding 90 days was estimated, given prior evidence of poorer survival among women experiencing diagnostic intervals of 3 months or longer (Richards *et al*, 1999).

Following this detailed analysis by symptom, a degree of aggregation into broader symptom groups was necessary for the examination of diagnostic timeliness, given the very small number of cases with some symptoms. Therefore, the presenting symptoms of breast cancer were categorised into three main symptom types: a) *breast lump*, b) *non-lump breast symptoms* (including breast pain, breast skin or shape abnormalities and nipple abnormalities), and c) *non-breast symptoms* (including fatigue, breathlessness, axillary symptoms, neck lump, and back pain) (see Figure 6.2). Women could belong to more than one symptom group if they had multiple symptoms; from the resulting seven combinations of the three symptom groups, I focused on the four largest combinations for further analysis ('lump', 'lump and non-lump', 'non-lump', and 'non-breast' symptom groups).

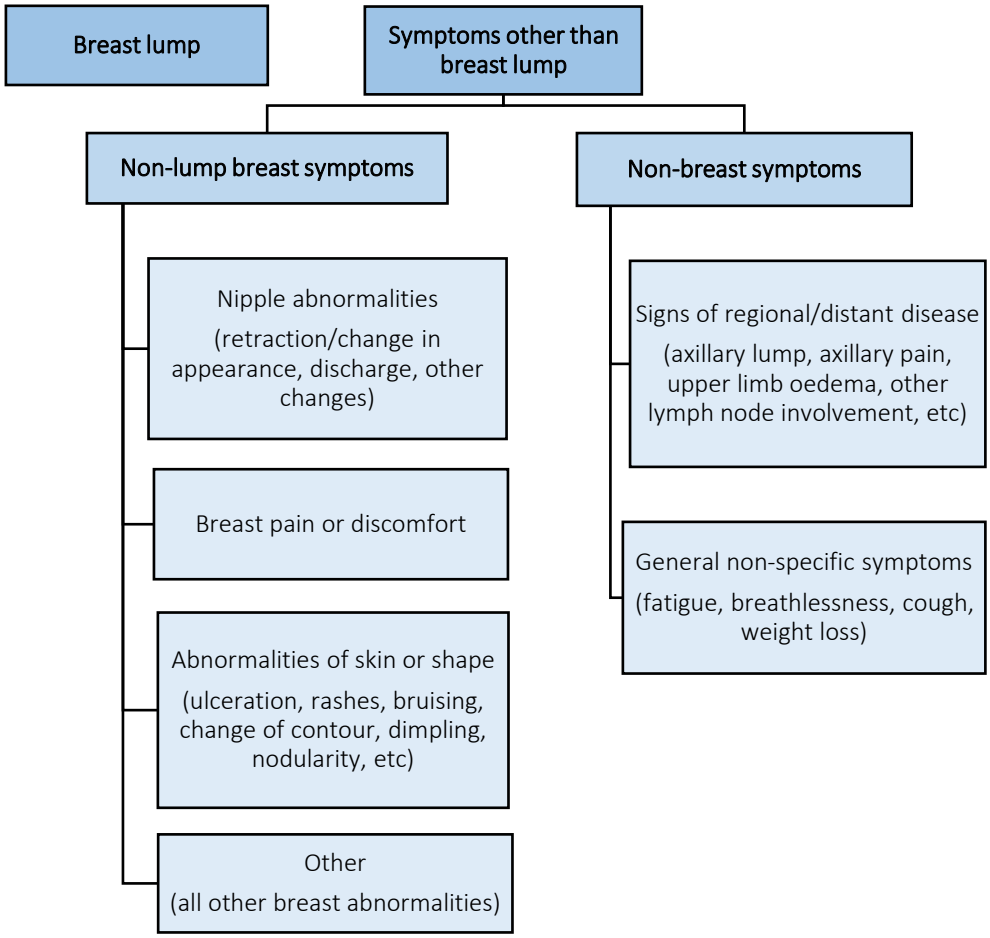


Figure 6.2 Taxonomy of presenting symptoms among breast cancer patients in the NACDPC

Kruskal-Wallis and Chi-squared tests were used to compare the length of diagnostic intervals and the number of pre-referral consultations by symptom combination, respectively. Subsequently, a regression framework was used to examine the association between symptom group and length of the patient and primary care intervals and pre-referral consultations, adjusting for age (parameterised as <50 years, 50–69 years, 70+ years) and ethnicity (white, non-white) given prior evidence suggesting likely associations between these socio-demographic variables with diagnostic intervals (Burgess *et al*, 2006; Marlow *et al*, 2014).

Preliminary examination of the data had indicated that the distribution of the patient and primary care interval values were positively skewed (skew and kurtosis test $p<0.0001$), likely reflecting skewed residuals which would contravene the assumptions of normal distribution required for linear regression. Therefore, quantile regression was used to examine variation in the length of the patient and primary care intervals. This is an established non-parametric method in analysing skewed data (as is often the case for diagnostic interval data), allowing relationships between outcome and predictor variables to be estimated at any point (quantile)

of the distribution of the interval values, where the direction and magnitude of association may differ to that of the mean (Jensen *et al*, 2014, 2015).

A continuity correction and log-transformation was applied to both variables before regression (in other words, 0.5 days was added to all intervals to allow zero intervals to be transformed), and significance testing was based on bootstrapping. Model coefficients should be interpreted as the relative difference in interval length between a given variable category compared to its reference, at specified centiles of the distribution.

For examining variation in the number of pre-referral consultations, a logistic regression model was used as this variable had been parameterised as a binary outcome; coefficients have been presented as the corresponding odds ratios.

6.4 Results

6.4.1 The symptom signature of breast cancer

A total of 2,316/2,783 (83%) of symptomatic women with breast cancer were included in the analysis. Among them, 2,543 symptoms were recorded, averaging 1.1 symptoms per woman.

Considering individual symptoms, a total of 56 distinct presenting symptoms were reported in the study population (Table 5.1), in 95 unique combinations. Breast lump was the most common symptom, recorded in about four-fifths of all women (83%). The next most commonly reported presenting symptoms were nipple abnormalities (7%), breast pain (6%), and breast skin abnormalities (2%).

Considering the patient interval, overall, 164 women (9% of those with patient interval values) waited longer than 90 days before seeking help. Among the larger non-lump breast symptoms, more than one in five women with breast ulceration (50%), nipple abnormalities (23%) and breast infection or inflammation (21%) had patient intervals of more than 90 days (Table 5.1).

In contrast to the substantial proportion of women with patient intervals longer than 3 months (9%, as above), only 2% of women had recorded primary care interval values of 90 days or longer. This small group of women tended to have symptoms such as non-specific breast abnormalities, back pain, musculoskeletal pain, chest pain, and fatigue or weakness.

Table 6.1 Frequencies of the 23 most common symptoms[1] among 2,316 women with breast cancer and measures of diagnostic timeliness

	Symptom signature and frequency (n=2,316)		Pre-presentation[2] (n=1,883)		Post-presentation[2]		
Symptom	No of women	% relative frequency (95% CI)	Patient Interval Median (IQR) 90th	% Patient Interval > 90 days (95% CI)	Primary Care Interval Median (IQR) 90th (n=2,201)	% Primary Care Interval > 90 days (95% CI)	% 2+ pre-referral consultations (n=2,002)
Breast lump	1922	83.0% (81.4–84.5%)	7 (1–27) 75	8% (7–9%)	0 (0–0) 3	1% (1–2%)	6%
Nipple abnormalities	158	6.8% (5.9–7.9%)	17 (2–71) 275	23% (17–31%)	0 (0–1) 7	1% (0.4–5%)	12%
Breast pain	149	6.4% (5.5–7.5%)	10 (3–41) 96	12% (8–19%)	0 (0–3) 34	3% (1–7%)	20%
Breast skin abnormalities	46	2.0% (1.5–2.6%)	13 (1–30) 129	10% (4–24%)	0 (0–1) 3	2% (0.4–12%)	8%
Axillary lump	27	1.2% (0.8–1.7%)	2.5 (0–12) 15	0% (0–15%)	0 (0–14) 34	4% (1–18%)	36%
Breast ulceration	25	1.1% (0.7–1.6%)	122 (0–276) 594	56% (27–81%)	0 (0–1) 1	0% (0–15%)	7%
Back pain	24	1.0% (0.7–1.5%)	9.5 (1–51) 107.5	10% (3–30%)	21 (0–105) 145	26% (13–46%)	65%
Breast contour abnormalities	17	0.7% (0.5–1.2%)	5 (4–18) 184	15% (4–42%)	0 (0–1) 3	0% (0–20%)	7%
Breast infection or inflammation	15	0.6% (0.4–1.1%)	2.5 (0–30) 366	21% (8–48%)	9 (0–23) 37	7% (1–31%)	60%
Breast swelling	14	0.6% (0.4–1.0%)	3.5 (0–14) †	10% (2–40%)	0 (0–3.5) 8	0% (0–24%)	15%
Musculoskeletal pain	14	0.6% (0.4–1.0%)	0.5 (0–22) †	10% (2–40%)	54 (0–187.5) 399	25% (9–53%)	75%
Breathlessness	11	0.5% (0.3–0.8%)	5 (0–35.5) †	0% (0–49%)	1 (0–10.5) †	0% (0–32%)	57%
Breast rash	10	0.4% (0.2–0.8%)	0 (0–16) †	0% (0–39%)	0 (0–7) †	0% (0–32%)	20%
Neck lump or lymph node abnormalities	9	0.4% (0.2–0.7%)	0 (0–10) †	0% (0–39%)	4.5 (0–19.5) †	0% (0–32%)	29%
Abdominal pain	8	0.3% (0.2–0.7%)	39 (18–62) †	17% (3–56%)	3 (2–6) †	0% (0–43%)	71%
Other breast abnormalities	8	0.3% (0.2–0.7%)	6 (0–8) †	0% (0–43%)	0 (0–98) †	33% (10–70%)	14%
Chest pain	8	0.3% (0.2–0.7%)	18 (10–43) †	0% (0–32%)	24 (9.5–83) †	25% (7–59%)	75%
Fatigue or weakness	7	0.3% (0.1–0.6%)	10.5 (1.5–33) †	0% (0–49%)	2 (0–27) †	14% (3–51%)	29%
Weight Loss	6	0.3% (0.1–0.6%)	56 (51–61) †	0% (0–66%)	18 (11–22) †	0% (0–43%)	60%
Cough	6	0.3% (0.1–0.6%)	5.5 (0–11) †	0% (0–66%)	13.5 (6.5–38) †	0% (0–49%)	60%
Axillary pain	5	0.2% (0.1–0.5%)	15 (0–126) †	33% (6–79%)	5 (1–8) †	0% (0–43%)	40%
Breast bruising	5	0.2% (0.1–0.5%)	7 (7–14) †	0% (0–43%)	0 (0–8) †	0% (0–43%)	40%
Oedema of upper limb	5	0.2% (0.1–0.5%)	76 (19–133) †	50% (10–91%)	0.5 (0–1) †	0% (0–49%)	0%
Total	2316	-	7 (1–28) 80	9% (8–10%)	0 (0–1) 7	2% (1–2%)	10%

†90th centile patient interval and primary care interval values not shown for symptoms where there were <10 patients with non-missing values.

1 Symptoms with a relative frequency of 0.2% or more are presented; see Appendix 6.3 for the full list of 56 symptoms. Symptom frequencies do not add up to 100% as some women had more than one symptom.

2 19%, 5%, and 14% of all observations had missing information on the patient interval, the primary care interval, and the number of pre-referral consultations respectively. For exact proportion by symptom please see Appendix 6.3.

6.4.2 Symptom group characteristics

The vast majority (99%) of women had one of four symptom combinations: 'lump only' (76%); 'non-lump only' (11%); 'both lump and non-lump' (6%); and 'non-breast symptoms' (5%) (Figure 6.3). There was no difference in frequency of symptom combination by age group of ethnicity (Table 6.2).

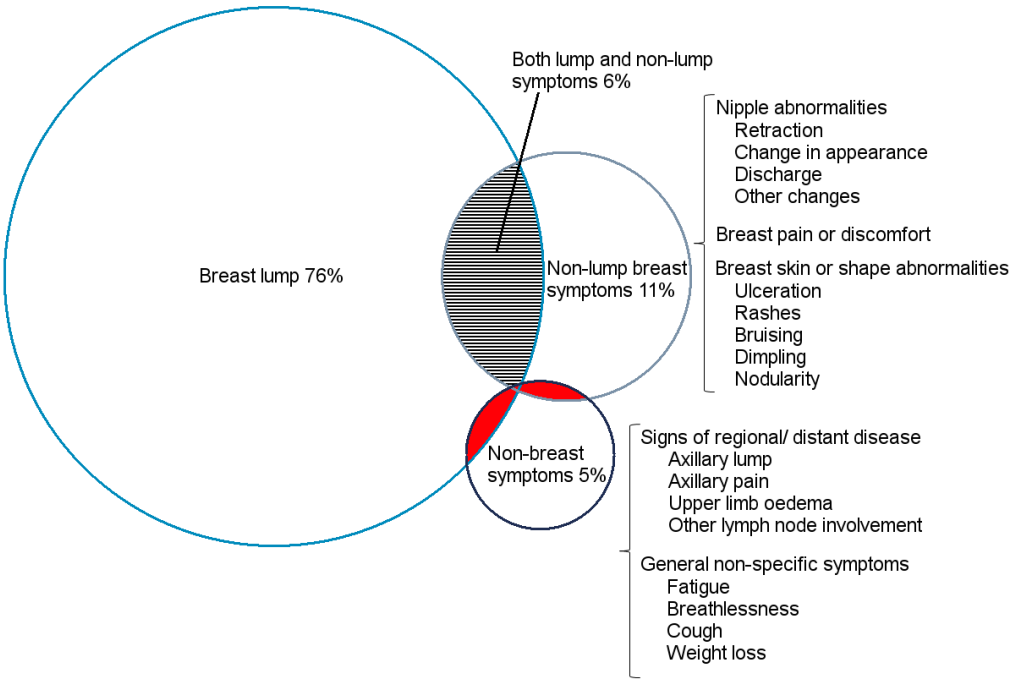
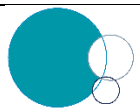
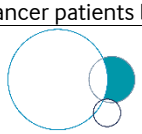
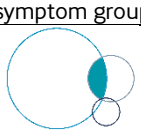
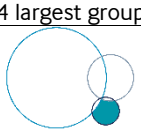


Figure 6.3 Venn diagram depicting the four largest symptom groups in 2,316 breast cancer patients
The three shaded groups in red were not investigated due to small numbers: lump and non-breast symptoms (n=12), non-lump breast symptoms and non-breast symptoms (n=7), and lump, non-lump breast symptoms, and non-breast symptoms (n=1).

Table 6.2 Characteristics of breast cancer patients by symptom group (4 largest groups shown)

	 Breast lump only	 Non-lump only	 Both lump & non-lump	 Non-breast only	Total*	χ^2 p-value[1]
Total	1,770 (76%)	262 (11%)	139 (6%)	125 (5%)	2,316	-
Age group						
<50yrs	515 (81%)	62 (10%)	34 (5%)	22 (3%)	637	
50–69yrs	586 (75%)	99 (13%)	43 (6%)	49 (6%)	781	
70+yrs	669 (74%)	101 (11%)	62 (7%)	54 (6%)	898	0.063
Ethnicity						
Non-white	1,640 (76%)	248 (12%)	129 (6%)	115 (5%)	2,150	
White	130 (78%)	14 (8%)	10 (6%)	10 (6%)	166	0.905

1 includes 20 patients that were not investigated further due to small numbers: lump and non-breast symptoms (n=12), non-lump breast symptoms and non-breast symptoms (n=7), and lump, non-lump breast symptoms, and non-breast symptoms (n=1)

6.4.3 Variation in measures of diagnostic timeliness by symptom group

On average, the patient interval was substantially longer than the primary care interval (median 7 versus 0 days, and 90th centile 80 versus 7 days, respectively, see Figure 6.4 and Table 6.3). Quantile plots of the patient and primary care interval values by symptom group indicated that while the majority of breast cancer patients experienced relatively short patient and primary care intervals, some women experienced much longer intervals (Figure 6.4). Among the women who experienced longer intervals, there was visible variation by symptom group. As the variation in interval length between symptom groups was concentrated at the long right tail of the distribution, I examined differences at the 90th centile in addition to the median.

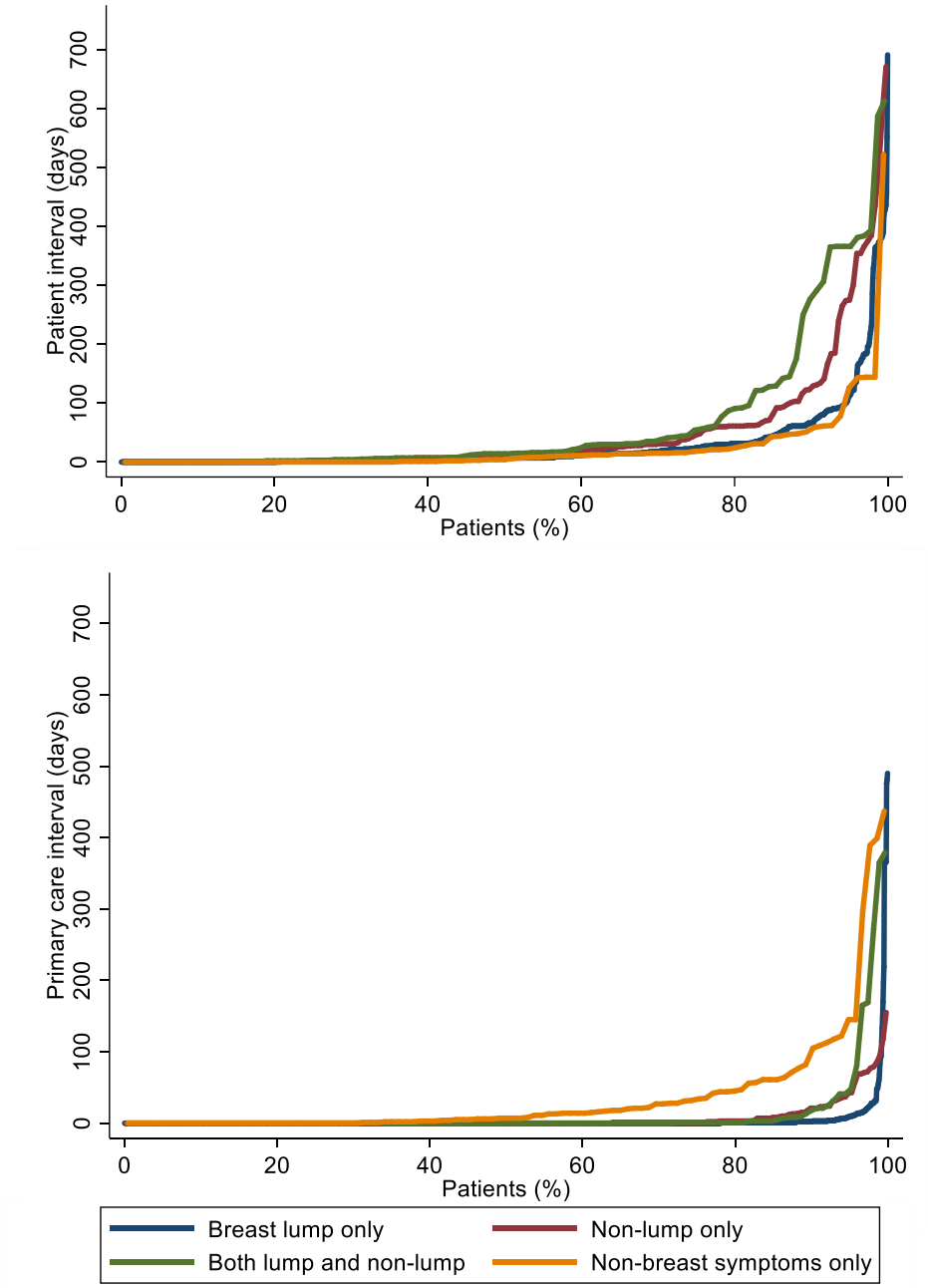


Figure 6.4 Quantile plot distribution of the patient (left) and primary care (right) intervals by symptom group
Data relate to the four largest presenting symptom groups.

6.4.3.1 Patient interval

There was strong evidence for variation in the patient interval by symptom group (p<0.001). Women with 'lump only' symptoms had median (90th centile) patient interval values of 7 (66) days. In contrast, those with 'non-lump only' or 'both lump and non-lump' symptoms had median (90th centile) intervals of 12 (126) days and 14 (276) days, respectively, while women with 'non-breast symptoms' had shorter intervals (of 4 (59) days) (see Table 6.3).

Observed patterns of variation in the patient interval by symptom group remained largely unchanged after adjusting for age group and ethnicity. Compared to women who presented with 'breast lump only', women in the 'non-lump only' group had patient intervals that were 1.6-fold (p=0.05) to 2.3-fold (p=0.003) longer at different centiles, while women with 'lump and non-lump' symptoms had patient intervals that were 1.9-fold (p=0.01) to 3.5-fold (p=0.001) longer at different centiles (see Table 6.3). There was no evidence for variation in the length of the patient interval by age or ethnicity at any of the quantile points examined.

6.4.3.2 Primary care interval

Observed primary care interval values also varied by symptom group (3): women presenting with 'lump only' had the shortest median (90th centile) intervals (0 (2) days), while those with 'non-breast' symptoms had the longest intervals (7 (105) days), respectively. Concordant patterns of variation by symptom group were apparent when examining the proportion of women with 2 or more pre-referral consultations (see Table 6.3).

Adjusting for differences in age group and ethnicity, symptom groups other than the 'lump only' group had longer intervals to referral (see Table 6.3). Women with 'non-breast' symptoms had particularly long primary care intervals compared to those with 'breast lump only' (15-fold greater at the median, p<0.001). Women with 'non-lump only' and 'lump and non-lump' symptoms also had longer time to referral after adjusting for age group and ethnicity, but this was only seen in the upper quantiles, in other words affecting a smaller number of women.

There was no evidence to support variation in the primary care interval by ethnicity. Younger women (aged <50 years) experienced longer time to referral compared to the reference age group (aged 50–69 years) at the 75th centile and above, but there was no evidence for this at the median.

Table 6.3 Descriptive statistics of the patient and primary care intervals, and relative differences in length of interval at different centiles among symptomatic women with breast cancer

Symptom group	Median (IQR) 90th	Kruskal-Wallis p-value	% women >90 days (95% CI)	Q(0.25) (95% CI)	Q(0.50) (95% CI)	Q(0.75) (95% CI)	Q(0.90) (95% CI)	Joint Wald test p-value
Patient interval (n=1,878)†								
All women	7 (1–28) 80	-	9% (8–10%)	-	-	-	-	
Breast lump only	7 (1–24) 66		7% (6–9%)	Ref.	Ref.	Ref.	Ref.	
Non-lump only	12 (2–46) 126	**<0.001**	15% (11–20%)	1.7 (0.7–3.8)	1.6 (1.0–2.6)	**1.9 (1.2–2.9)**	**2.1 (1.4–3.2)**	**<0.001**
Lump and non-lump	14 (3–54) 276		20% (14–29%)	**2.3 (1.1–4.8)**	**1.9 (1.2–3.2)**	1.9 (0.9–4.0)	**3.5 (1.6–7.3)**	
Non-breast symptoms	4 (0–18) 59		6% (2–12%)	**0.3 (0.2–0.5)**	0.7 (0.3–1.7)	0.8 (0.5–1.2)	0.9 (0.5–1.6)	
White	7 (1–28) 80	0.509	9% (8–10%)	Ref.	Ref.	Ref.	Ref.	0.779
Non-white	6 (0–30) 78		8% (5–14%)	1.0 (0.4–2.8)	0.9 (0.5–1.5)	1.2 (0.8–1.7)	1.0 (0.6–1.5)	
<50 years	7 (1–27) 66		7% (5–10%)	1.0 (0.7–1.5)	1.2 (0.9–1.5)	1.1 (0.8–1.5)	1.0 (0.7–1.4)	
50–69 years	7 (1–25) 72	0.148	8% (6–10%)	Ref.	Ref.	Ref.	Ref.	0.421
70+ years	7 (1–31) 92		11% (9–13%)	1.0 (0.8–1.2)	1.2 (0.9–1.5)	1.3 (1.0–1.8)	1.4 (1.0–2.1)	
Primary care interval (n=2,194)†								
All women	0 (0–1) 7	-	2% (1–2%)	-	-	-	-	
Breast lump only	0 (0–0) 2		1% (1–2%)	Ref.	Ref.	Ref.	Ref.	
Non-lump only	0 (0–1) 21	**<0.001**	1% (0.4–4%)	1.0 (1.0–1.0)	1.0 (1.0–1.0)	**3.0 (1.8–5.0)**	**5.4 (2.7–10.9)**	**<0.001**
Lump and non-lump	0 (0–1) 18		4% (2–8%)	1.0 (1.0–1.0)	1.0 (1.0–1.0)	3.0 (1.0–9.3)	**3.2 (1.1–9.6)**	
Non-breast symptoms	7 (0–34) 105		10% (6–17%)	1.0 (0.5–2.0)	**15.0 (8.1–27.6)**	**57.0 (36.3–89.6)**	**41.0 (23.8–70.7)**	
White	0 (0–1) 7	0.620	2% (1–2%)	Ref.	Ref.	Ref.	Ref.	0.712
Non-white	0 (0–0) 10		1% (0.3–5%)	1.0 (1.0–1.0)	1.0 (1.0–1.0)	1.0 (0.7–1.4)	1.4 (0.6–3.3)	
<50 years	0 (0–1) 15		3% (2–5%)	1.0 (1.0–1.0)	1.0 (1.0–1.0)	**3.0 (1.5–6.0)**	**3.8 (2.0–7.1)**	
50–69 years	0 (0–0) 4	**0.016**	1% (1–2%)	Ref.	Ref.	Ref.	Ref.	**<0.001**
70+ years	0 (0–1) 3		1% (1–2%)	1.0 (1.0–1.0)	1.0 (1.0–1.0)	1.0 (1.0–1.0)	0.6 (0.3–1.1)	

†19% and 5% of women had missing information on the patient interval and the primary care interval respectively.
NB the smallest 2 symptom groups were excluded from the model due to small numbers and the 'lump and non-breast' symptom group was included in the model but is not reported here.

Table 6.4 Adjusted odds ratios of 2+ versus 1 pre-referral consultation by symptom group, age group, and ethnicity

Symptom group	% 2+ pre-referral consultations	X^2 p-value	Adjusted odds ratio	Joint Wald test p-value
All women	10%	-	-	-
Breast lump only	5%	<0.001	Ref.	<0.001
Non-lump only	17%		4.2 (2.7–6.4)	
Lump and non-lump	15%		3.8 (2.2–6.5)	
Non-breast symptoms	54%		28.1 (17.6–45.0)	
White	10%	0.434	Ref.	0.7198
Non-white	12%		1.1 (0.6–2.0)	
<50 years	14%	<0.001	2.3 (1.6–3.4)	<0.001
50–69 years	9%		Ref.	
70+ years	7%		0.8 (0.5–1.1)	

NB the smallest 2 symptom groups were excluded from the model due to small numbers and the 'lump and non-breast' symptom group was included in the model but is not reported here.

6.5 Discussion

6.5.1 Main findings

About 1 in 6 women with breast cancer presented without a breast lump, instead experiencing a wide spectrum of symptoms before seeking help. The length of the patient and the primary care intervals varied by symptom group, particularly in the upper centiles of the distribution. Women in the 'non-lump only' and 'both lump and non-lump' symptom groups had longer median patient intervals compared to those with 'breast lump only'. Similar associations were also seen post-presentation, although in general primary care intervals were appreciably shorter than patient intervals.

6.5.2 Comparison with prior evidence

The present analysis substantially amplifies previous findings in this field, providing evidence of quantifiable differences in diagnostic timeliness by the symptoms of breast cancer (Ramirez *et al*, 1999; Webber *et al*, 2017). Regarding the symptom signature of breast cancer, a previous study using Read-coded electronic primary care data reported similar proportions of non-lump breast symptoms to those observed here (Redaniel *et al*, 2015), but I described a substantially wider range of presenting symptoms in greater detail than the broader categorisations used previously.

6.5.3 Strengths and limitations

This is the first and largest study to examine associations between presenting symptom categories and the length of the patient and the primary care interval in a nationwide representative sample of women subsequently diagnosed with breast cancer.

There are several limitations that should be acknowledged. As patient records were examined retrospectively (and in the knowledge of the patient's diagnosis), non-specific, particularly non-breast, symptoms may have been under-captured by the audit as discussed in Chapter 3

(Literature review). There were missing outcome data regarding intervals and number of consultations for a minority of women, in proportions comparable to previous studies of diagnostic timeliness (Hansen *et al*, 2011; Walter *et al*, 2016a, 2016b; Leiva *et al*, 2017). Women who did not first present in primary care and were older were more likely to have missing data but were otherwise similar across other characteristics of interest.

The length of patient intervals by symptom may vary by socio-economic status, but I was unable to examine variation in diagnostic timeliness by level of deprivation, or other patient-level characteristics such as health literacy or history of screening participation (Marcu *et al*, 2016, 2017). Although I initially aimed to analyse associations between individual symptoms of breast cancer with diagnostic timeliness, sample size limitations (particularly for non-breast symptoms) meant that I ended up analysing these associations using aggregate symptom groups and symptom combinations.

6.5.4 Implications

This study provides detailed evidence about the symptom signature of breast cancer and the frequencies and diagnostic intervals associated with different symptoms which could inform the design of public health campaigns. The findings indicate that in general, help-seeking and referral for specialist investigation is prompt among symptomatic women diagnosed with breast cancer. Nevertheless, a small proportion of the study population experienced significantly longer intervals to help-seeking and referral, and among these women, those with non-lump symptoms were more prone to longer intervals than those with breast lump alone.

This suggests that awareness interventions should continue to encompass (if not emphasise) the likely importance of 'non-lump' breast symptoms to promote the earlier diagnosis of breast cancer (Campbell *et al*, 2015; Kaushal *et al*, 2016; Public Health England, 2016a; World Wide Breast Cancer, 2016). Relatedly however, and as discussed in Chapter 3 (Literature review), the predictive value of symptoms for a given cancer is also important in the context of designing symptom awareness campaigns (in addition to considering the symptom signature and associated diagnostic timeliness). Currently, there is little relevant evidence beyond that for breast lump, but some non-lump breast symptoms (such as nipple eczema or breast ulceration) may have equal or greater positive predictive values for breast cancer (Dalberg *et al*, 2008; Huggenberger & Andersen, 2015).

Women in the 'both lump and non-lump' group had longer patient intervals compared to those with 'breast lump only' group. This finding is somewhat puzzling given that breast lump, which is associated with shorter intervals, is present in both groups. This may reflect a higher tendency to normalise a lump in the breast in the presence of other non-lump breast symptoms (Marcu *et al*, 2016). Previous research has shown that among women with prolonged patient intervals (12 weeks or longer), some had initially experienced non-lump breast symptoms and then had

subsequently developed a lump by the time of (delayed) presentation (Burgess *et al*, 1998). While the sequence of experienced symptoms could not be ascertained among this study population, further research regarding symptom progression and possible associations with time to help-seeking could be valuable, for example patient interviews, discrete choice experiment, or vignette studies examining the psychological tendencies of interpreting symptom combinations depending on order of experience.

The majority of women had much shorter intervals post-presentation than pre-presentation (1 in 2 women with breast cancer in the NACDPC had a primary care interval of 0 days) and there was no evidence for variation in the median primary care interval by symptom group. The small minority of women who presented with 'non-breast symptoms' (e.g. back pain or breathlessness) however had substantially longer primary care intervals compared to those with breast lump or non-lump breast symptoms. Shortening diagnostic intervals in those women will improve patient experience (Mendonca *et al*, 2016), although it may not lead to better clinical outcomes given that distant symptoms are often signs of late stage (metastatic) disease. I consider the broader question about associations between different presenting symptoms of cancer and stage at diagnosis in Chapter 9.

Identifying these women is also likely to be challenging, due to the low predictive values of these symptoms for breast cancer. New diagnostic services for non-specific symptoms such as the Danish three-legged strategy' and the Multi-disciplinary Diagnostic Centres (MDCs) piloted as part of the Accelerate, Coordinate, Evaluate (ACE) programme in England may be of particular value in this regard, as they aim to expedite the investigation and diagnostic resolution for patients with serious non-specific symptoms that could indicate cancer (Vedsted & Olesen, 2015; Fuller *et al*, 2016; Forster *et al*, 2018).

6.6 Chapter summary

This study provides a detailed description of the symptom signature at presentation among women subsequently diagnosed with breast cancer, and confirms an association between non-lump presenting symptoms of the breast and prolonged diagnostic intervals (particularly with regard to the patient interval). It demonstrates the value of researching presenting symptoms and associated diagnostic intervals in cancer patients. Nevertheless, taking a cancer-specific approach is somewhat limited given that cancer patients often initially present with symptoms that do not immediately betray the tumour site. Indeed, Chapter 3 (Literature review) indicated that the vast majority of cancers have broad symptom signatures. Thus, the next two chapters will examine a broad group of symptoms whereby it may be possible to stretch the impact of a single early diagnosis intervention to beyond a single cancer.

Chapter 7: Abdominal symptoms and time to presentation

This chapter focuses on cancer patients who presented with abdominal symptoms before diagnosis and their time to presentation, and considers the implications of the findings for public health awareness campaigns that promote timely help-seeking.

Aspects of this chapter have been the subject of a peer-reviewed publication in the Journal of Public Health[1].

[1] Koo MM, Wagner C von, Abel G, McPhail S, Rubin G, Lyratzopoulos G (2018) **The nature and frequency of abdominal symptoms in cancer patients and their associations with time to help-seeking: evidence from a national audit of cancer diagnosis.** *Journal of Public Health*, https://doi.org/10.1093/pubmed/fdx188 (see Appendix 7.1)

7 Abdominal symptoms and time to presentation

7.1 Rationale of this chapter

Following on from Chapter 6, this chapter focuses on cancer patients identified and grouped by their symptoms at presentation (rather than diagnosed cancer), and examines the range of subsequently diagnosed cancer sites (namely the cancer signature), and variation in diagnostic timeliness between patients with different symptoms.

The decision to focus on abdominal symptoms was influenced by the prominence of abdominal symptoms in early diagnosis. Firstly, a Be Clear on Cancer campaign focusing on a range of abdominal symptoms was piloted by Public Health England in February to March 2017 in the East Midlands and West Midlands (Public Health England, 2017b). Secondly, abdominal symptoms have been the focus of novel multi-disciplinary centres designed to expedite the investigation and diagnosis of certain abdominal symptoms. This chapter focuses on abdominal symptoms and time to presentation in the context of public health campaigns, while Chapter 8 will expand on the implications of abdominal symptoms and time to referral for early diagnosis interventions post-presentation.

7.2 Introduction

As referred to in Chapter 1 (Introduction), population-wide awareness campaigns about possible cancer symptoms are commonly organised as part of early diagnosis activities (Danish National Board of Health, 2005; Austoker *et al*, 2009; NHS Scotland, 2012; Cancer Australia, 2016; Public Health England, 2016a; CDC, 2017). To date, symptom awareness campaigns have tended to implicitly focus on cancer sites, by targeting 'alarm' symptoms explicitly associated with specific cancers, such as 'blood in poo' and colorectal cancer (NHS Scotland, 2012; Cancer Australia, 2013; Public Health England, 2016b).

There is however growing interest in campaigns targeting symptoms relating to a body area or system rather than a cancer site, as this provides an opportunity to promote the earlier presentation of rare and less common cancers. In England, an abdominal symptoms campaign was recently piloted at regional level focusing on a range of symptoms (diarrhoea, bloating, abdominal discomfort, constipation, nausea, and blood in poo) (Public Health England, 2017b). Abdominal symptoms include both alarm and non-alarm symptoms of cancer, and could represent a number of cancer sites including common cancers with particularly poor outcomes (such as colorectal cancer) and rarer cancers that would be otherwise not be included in population-wide education campaigns due to concerns regarding cost-effectiveness.

Examining the length of the patient interval (time from symptom onset to presentation) associated with different abdominal symptoms could contribute to the emerging evidence base

supporting the design of symptom awareness campaigns (Moffat *et al*, 2015; Power & Wardle, 2015; Wagland *et al*, 2016). As described in Chapter 1 (Introduction), symptom-specific patient intervals may be interpreted as measures of relative need for such interventions (Lyratzopoulos, 2014). Alongside considerations of other factors such as the predictive value of a symptom for cancer, and the prevalence of different symptoms in the general population, this symptom-specific evidence could support the prioritisation of certain symptoms over others in the focus of awareness campaigns.

Further, evidence regarding the anticipated cancer site case-mix of a particular symptom could help guide the direction of evaluation strategies. Estimating the impact of a symptom awareness campaign has been shown to be challenging due to the diffuse and broad-reaching nature of campaigns; such difficulties are likely to be exacerbated by symptom-based approaches that target more than one cancer site (Ironmonger *et al*, 2014; Moffat *et al*, 2015; Emery *et al*, 2017).

I therefore aimed to examine the frequency of abdominal symptoms at presentation in a representative population of incident cancer patients; describe the range of cancers associated with abdominal symptoms in an incident cohort; and investigate variation in the length of the patient interval by presenting abdominal symptom.

7.3 Methods

7.3.1 Study population

For this study, 15,956 NACDPC cancer patients with complete and valid information on age group (among patients aged fifteen years or older), sex, and presenting symptoms were included (see Figure 7.1 for flow chart of sample derivation). As described in Chapter 4 (Data & methods), individuals diagnosed incidentally or asymptomatically and those with cancer sites categorised as "No information" and "Unknown Primary" were excluded from the analysis.

7.3.2 Abdominal symptom definition

I selected a total of 18 symptom constructs from those that were coded in the NACDPC data based on the abdominal symptoms described by the 2015 NICE guidelines for suspected cancer referral (NICE, 2015). The symptom constructs were further aggregated into eight abdominal symptom groups, with clinical advice from GL and GPR[2] (see Table 7.1). Abdominal symptom constructs within each symptom group were examined in further analyses (see Section 7.3.5.1). Hereafter, these eight symptom groups are simply referred to as abdominal 'symptoms'.

[2] GL: Professor Georgios Lyratzopoulos (principal supervisor); GPR: Professor Greg Rubin (collaborator)

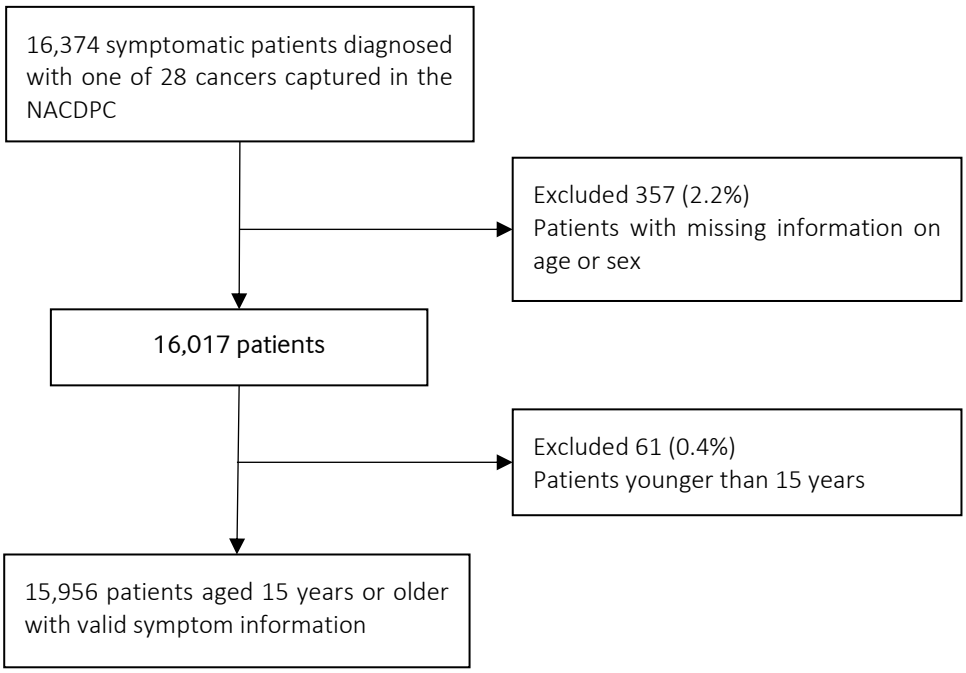


Figure 7.1 Flow diagram of sample derivation for the study population of this chapter

Table 7.1 Abdominal symptom definitions, based on NICE 2015 guidelines

Symptom	Symptom constructs
Abdominal Pain	Abdominal pain[1]
	Epigastric pain
	Right Iliac Fossa (RIF) pain
	Suprapubic pain
	Loin pain & renal colic
Change in Bowel Habit	Constipation
	Change in bowel habit
	Diarrhoea
Dyspepsia	Dyspepsia and related epigastric symptoms[2]
Dysphagia	Odynophagia
	Dysphagia
Reflux	Reflux
Bloating or Distension	Abdominal bloating
	Abdominal distension
	Ascites
Nausea or Vomiting	Vomiting
	Nausea
Rectal bleeding	Rectal bleeding[3]

1 Abdominal pain that was not otherwise specified, excluding acute abdominal pain
2 includes dyspepsia, indigestion, waterbrash, gastritis, burping, belching, "GI upset", and "upper GI symptoms".
3 includes blood in stool and rectal bleeding, excludes acute rectal bleeding

7.3.3 Variables of interest

For the purposes of examining the cancer signature of abdominal symptoms and in order to aid interpretation, cancer sites were grouped into three categories: 1) abdominal or adjacent organ cancers (cancers arising in the intra-abdominal organs, together with oesophageal and prostate cancer); 2) solid tumour malignancies other than the above (i.e. cancers of non-abdominal organs) hereafter called 'other solid tumour' cancers; or 3) haematological cancers (lymphoma, leukaemia, and myeloma).

The interval of interest in this analysis was the patient interval (time from symptom onset to first presentation). Information on the patient interval was highly complete among the study population (79% of 2,253 cancer patients with a single abdominal symptom, see Section 7.3.4 below for justification of sample size) and so I conducted complete case analysis treating the patient interval as a continuous variable. Patients missing information on the patient interval were more likely to have first presented in places other than general practice, and to have presented with nausea or vomiting, without evidence for variation in missing patient interval by age or sex (see Appendix 7.2).

7.3.4 Statistical analyses

Firstly, I estimated the frequencies (and associated exact confidence intervals) of abdominal symptoms in the studied cancer patient population. I then described the cancer site case-mix (cancer signature) of abdominal symptoms, namely the range and relative frequencies (proportions) of different cancer sites subsequently diagnosed among cancer patients presenting with one or more abdominal symptoms. Symptom-specific cancer signatures have been examined as part of Chapter 8.

Subsequently, I examined variation in the patient interval by abdominal symptom. As public awareness campaigns target individual symptoms rather than specific symptom combinations, these analyses were restricted to the majority of cancer patients with a single recorded presenting abdominal symptom (n=2,253, 62% of all patients reporting an abdominal symptom). Nevertheless, I examined patients with multiple abdominal symptoms by investigating common abdominal symptom combinations in further supplementary analyses (see Section 7.3.5.2 below).

The mean, median, interquartile range and 90th centiles of the patient interval were estimated for each abdominal symptom along with 95% confidence intervals using a bootstrap approach with 1000 replications. Kruskal-Wallis tests were used to examine variation in median interval length by abdominal symptom. The proportion of patients with each symptom that experienced a patient interval of 2 months (60 days) or more was also calculated to help to further contextualise the findings.

I then used generalised linear models (GLM) to examine the association between abdominal symptoms and the patient interval adjusted for age group (parameterised as <50 years, 50–69 years, 70+ years), sex (men, women), and ethnicity (white, non-white) given prior evidence supporting their associations with diagnostic timeliness (Lyratzopoulos *et al*, 2012). To account for skewed outcome data, a log link function was used, and significance testing was again based on bootstrapping (1000 replications). Variation in interval length across categorical variables was examined using joint Wald tests, with statistical significance at the 5% level. Due to 10% of patients with missing information on ethnicity and 18% of patients with missing information on the patient interval, the adjusted model was based on 1559 cancer patients (3% of patients had missing information on both).

7.3.5 Supplementary analyses

7.3.5.1 Abdominal symptom constructs

In order to assess the robustness and validity of how I had defined the abdominal symptoms, I examined the 18 symptom constructs within the eight abdominal symptoms. Findings are presented in Appendix 7.3.

7.3.5.2 Patients with multiple abdominal symptoms

I undertook supplementary analysis to examine patients who had presented with multiple abdominal symptoms. I examined the frequency of the four most common pairs of abdominal symptoms and their associated distributions of the observed patient interval, alongside patients who had presented with a single abdominal symptom for comparison. This was conducted in 3,438/3,661 (94%) patients who presented with one or more abdominal symptom. Findings are presented in Section 7.4.4 below.

7.4 Results

7.4.1 Frequency of presenting abdominal symptoms in cancer patients

Of a total of 15,956 patients with cancer, 3,661 (23%) presented with one or more abdominal symptoms. Abdominal pain was the most common abdominal symptom across the entire cohort of cancer patients (8%), followed by change in bowel habit (6%) and rectal bleeding (5%) (Table 7.2 below).

Table 7.2 Frequency of abdominal symptoms among symptomatic cancer patients (n=15,956)

Symptom	No. of patients	Percentage of symptomatic cancer patients (95% CI)
Abdominal pain	1268	7.9% (7.5–8.4%)
Change in bowel habit	1010	6.3% (6.0–6.7%)
Rectal bleeding	768	4.8% (4.5–5.2%)
Dysphagia	418	2.6% (2.4–2.9%)
Nausea or vomiting	261	1.6% (1.5–1.8%)
Dyspepsia	256	1.6% (1.4–1.8%)
Bloating or distension	250	1.6% (1.4–1.8%)
Reflux	71	0.4% (0.4–0.6%)
Any abdominal symptom	3661	22.9% (22.3–23.6%)

NB Number of patients (percentages) sum to more than 3661 (23%) as patients could have more than one abdominal symptom.

7.4.2 Cancer site case-mix of abdominal symptoms in cancer patients

Among the 3,661 cancer patients who presented with abdominal symptoms, the majority (89%, 3,244/3,661) were diagnosed with solid cancers of abdominal or adjacent organs (Figure 7.2). The most commonly diagnosed cancer site was colorectal cancer (47%), followed by oesophageal (13%), ovarian (7%), and pancreatic (6%) cancers (see Table 7.3 and Figure 7.2). A further 14 cancer sites were represented among the remainder of patients, including other solid tumours (8%) and haematological cancers (4%). The cancer signature of each individual abdominal symptom is examined further in the next chapter (Chapter 8: Abdominal symptoms and time to referral).

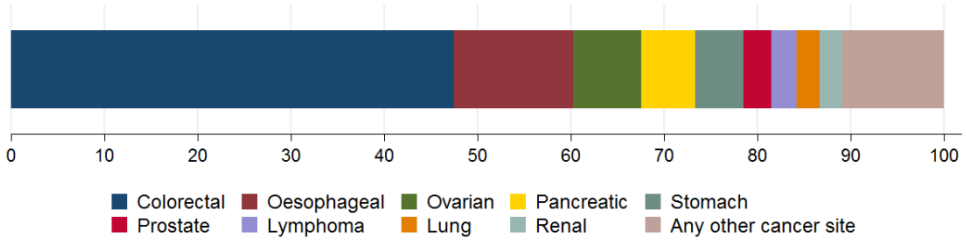


Figure 7.2 Cancer site case-mix of patients who presented with one or more abdominal symptom (n=3,661) Proportions of the nine most frequent cancers across all abdominal symptoms shown only; other cancer diagnoses represented as 'Any other cancer site' category. See Table 7.3 for exact proportions.

I also considered the relative importance of abdominal symptoms for each cancer site by calculating the proportion of patients with a given cancer who had presented with one or more abdominal symptoms. Unsurprisingly, over two-fifths (41%) of cancer patients diagnosed with an abdominal cancer had presented with abdominal symptoms, although this ranged from 84% of patients later diagnosed with oesophageal cancer to 5% of patients later diagnosed with prostate cancer (see Table 7.3 for full breakdown). Notably, abdominal symptoms were common at presentation among many rarer cancers such as pancreatic, ovarian, small intestinal, and gallbladder cancer. Patients with other solid tumours arising outside the abdominal region were much less likely to report abdominal symptoms (4%, n=279). In contrast, patients diagnosed with haematological cancers were more likely to report abdominal

symptoms at presentation (11%, n=138), almost two thirds of those being patients with lymphoma (Table 7.3).

Table 7.3 Cancer site case-mix of patients with one or more abdominal symptoms (n=3,661) and proportion of patients with a given cancer that had abdominal symptoms

Cancer	No. of patients	Percentage of patients with one or more abdominal symptoms subsequently diagnosed with a given cancer (95% CI)	Percentage of patients with a given cancer who had one or more abdominal symptoms
Abdominal or adjacent organ cancers†			
Colorectal	1737	47.4% (45.8–49.1%)	75%
Oesophageal	468	12.8% (11.7–13.9%)	84%
Ovarian	267	7.3% (6.5–8.2%)	70%
Pancreatic	214	5.8% (5.1–6.7%)	59%
Stomach	189	5.2% (4.5–5.9%)	65%
Prostate	110	3.0% (2.5–3.6%)	5%
Renal	89	2.4% (2.0–3.0%)	29%
Bladder	40	1.1% (0.8–1.5%)	5%
Liver	38	1.0% (0.8–1.4%)	44%
Small Intestine	36	1.0% (0.7–1.4%)	69%
Gallbladder	32	0.9% (0.6–1.2%)	51%
Endometrial	24	0.7% (0.4–1.0%)	6%
Sub-total	3244	88.6% (87.5–89.6%)	41%
Other solid tumours			
Lung	91	2.5% (2.0–3.0%)	5%
Oropharyngeal	20	0.5% (0.4–0.8%)	10%
Breast	14	0.4% (0.2–0.6%)	1%
Laryngeal	12	0.3% (0.2–0.6%)	10%
Brain	10	0.3% (0.1–0.5%)	5%
Cervical	10	0.3% (0.1–0.5%)	8%
Sarcoma[1]	10	0.3% (0.1–0.5%)	10%
Testicular	5	0.1% (0.1–0.3%)	3%
Melanoma	4	0.1% (0.04–0.3%)	0.5%
Mesothelioma	4	0.1% (0.04–0.3%)	6%
Thyroid	4	0.1% (0.04–0.3%)	4%
Sub-total	279[2]	7.6% (6.8–8.5%)[2]	4%[2]
Haematological cancers			
Lymphoma[1]	97	2.6% (2.2–3.2%)	15%
Leukaemia	25	0.7% (0.5–1.0%)	7%
Myeloma	16	0.4% (0.3–0.7%)	8%
Sub-total	138	3.8% (3.2–4.4%)	11%

†Defined as cancers arising in the intra-abdominal organs, together with oesophageal and prostate cancer (see page 102)
NB Ordered by frequency among patients with abdominal symptoms
1 It is likely that a proportion of sarcomas and lymphomas were intra-abdominal but information regarding their exact location was not available.
2 Includes 476 cases described as 'Other' cancers, of which 2.6% (n=95) presented with abdominal symptoms, and 69 vulval cancer patients, none of whom presented with abdominal symptoms.

7.4.3 Patient interval by presenting abdominal symptom

Among cancer patients with a single presenting abdominal symptom (n=2,253), there was strong evidence for variation in the patient interval (symptom-onset-to-presentation) by symptom (p<0.001, Figure 7.3 and Table 7.4).

Patients presenting with change in bowel habit or dysphagia had the longest patient intervals: one in two patients with either of these symptoms waited at least a month before presentation, while a quarter waited two months or longer (median (IQR) patient interval: 30 (4–73) days for change in bowel habit; and 30 (10–61) days for dysphagia). A considerable proportion (25–30%) of patients with bloating or distension, reflux, and rectal bleeding also waited for two months or longer before presentation. In contrast, patients presenting with abdominal pain or nausea/vomiting went to the doctor sooner on average (7 (0–28) days and 7 (0–23) days respectively). The variation in interval length by abdominal symptom persisted after adjusting for age group, sex, and ethnicity (Table 7.4).

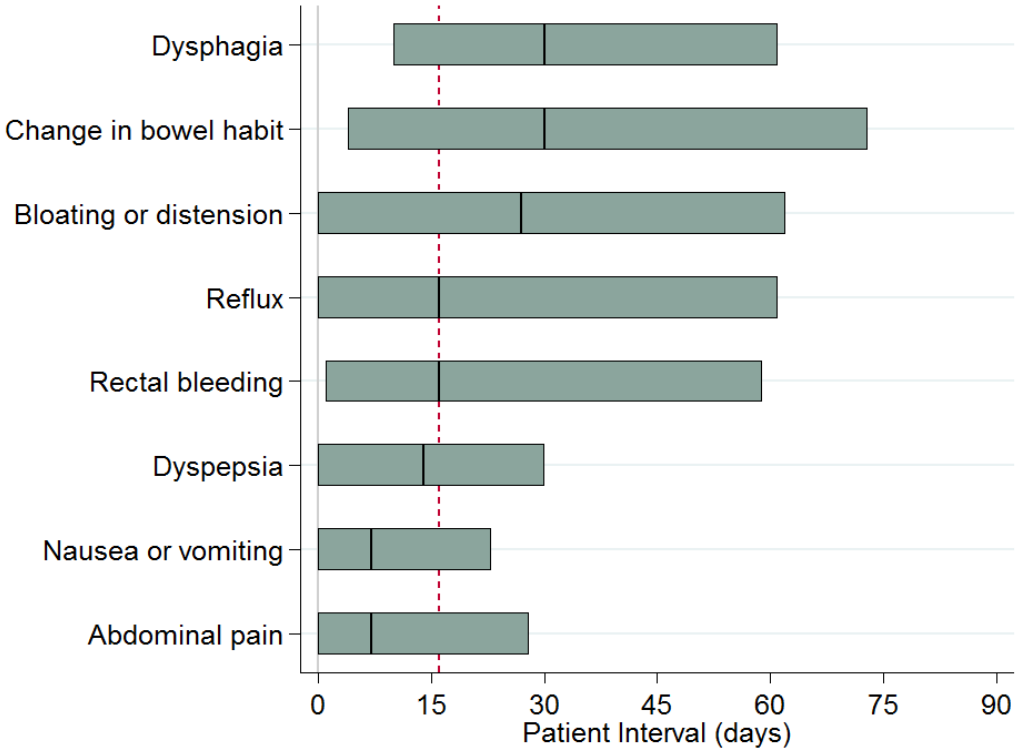


Figure 7.3 The length of the patient interval by presenting abdominal symptom
Among patients with a single abdominal symptom (n=1,783 as 21% had missing patient interval values). Bar length = IQR, vertical line = median value; the red dashed vertical line represents the median patient interval value (16 days) across all abdominal symptoms. Symptoms are ordered by median interval length.

Table 7.4 Summary statistics for the patient interval (measured in days), and adjusted GLM coefficient by abdominal symptom among patients with a single abdominal symptom

Symptom	N	Mean	25th	50th	75th	90th	% 60+ days	Adjusted GLM coefficient†[1]
Dysphagia	267	48	10	30	61	116	25%	**1.7 (1.3–2.4)**
Change in bowel habit	525	63	4	30	73	182	32%	**2.5 (1.9–3.4)**
Bloating or distension	96	48	0	27	62	118	30%	**2.1 (1.3–3.2)**
Reflux	29	41	0	16	61	128	27%	1.4 (0.7–2.9)
Rectal bleeding	495	55	1	16	59	136	25%	**2.2 (1.5–3.1)**
Dyspepsia	118	31	0	14	30	87	15%	1.1 (0.6–2.0)
Nausea or vomiting	53	49	0	7	23	183	21%	2.1 (0.6–6.7)
Abdominal pain	670	28	0	7	28	70	12%	Ref.
All abdominal symptoms	2253	47	1	16	54	122	23%	-

†31% of observations had missing information on the patient interval or ethnicity resulting in a final model size of n=1559.
1 Joint Wald test of variation in patient interval by symptom: p<0.001. The exponentiated coefficient represents the factor difference applicable to the geometric mean patient interval of the reference group. For example, a coefficient of 2 for a symptom would mean that on average, the patient interval was 2 times longer among patients with that symptom compared to patients with abdominal pain (reference group), adjusted for sex, age group, and ethnicity.

7.4.4 Supplementary analyses: patients with multiple abdominal symptoms

Further analyses examining patients with four most common abdominal symptom combinations were largely comparable to the main analyses findings in respect of associations with the patient interval, and are described below (Table 7.5 and Table 7.6).

Patients with the four most frequent abdominal symptom combinations were selected for further analyses, alongside eight abdominal symptoms as single symptoms (n=3,438, 94% of all patients with an abdominal symptom in the sample) (Table 7.7). The four symptom pairs were:

- Change in bowel habit (CIBH) and rectal bleeding;
- Abdominal pain and CIBH;
- Abdominal pain and nausea/vomiting; and
- Abdominal pain and bloating/distension.

Change in bowel habit and rectal bleeding was the most common symptom pair (seen in 0.9% of all symptomatic cancer patients).

Table 7.5 Frequency of the 12 most common abdominal symptom combinations

	Frequency of 12 most common abdominal symptom combinations among symptomatic cancer patients (n=15,956)	
Symptom combination[1]	No. of patients	% (95% CI)
Abdominal pain alone	952	6.0% (5.6–6.3%)
CIBH alone	694	4.3% (4.0%–4.7%)
Rectal bleeding alone	595	3.7% (3.4%–4.0%)
Dysphagia alone	355	2.2% (2.0%–2.5%)
Dyspepsia alone	168	1.1% (0.9–1.2%)
CIBH & Rectal bleeding	147	0.9% (0.8–1.1%)
Nausea or vomiting alone	140	0.9% (0.7–1.0%)
Bloating or distension alone	138	0.9% (0.7–1.0%)
Abdominal pain & CIBH	93	0.6% (0.5–0.7%)
Abdominal pain & Nausea or vomiting	64	0.4% (0.3–0.5%)
Abdominal pain & Bloating or distension	56	0.4% (0.3–0.5%)
Reflux alone	36	0.2% (0.2–0.3%)

CIBH: change in bowel habit. Symptom pairs are in bold print.
1 Symptom combinations of abdominal symptoms only

Patients who had one of the three symptom pairs including abdominal pain waited longer before seeking help (median patient interval: 14–19 days). Patients who presented with CIBH and rectal bleeding had a longer patient interval compared to those who had CIBH or rectal bleeding alone (median patient interval: 33 days versus 30 days and 17 days respectively).

Table 7.6 Summary statistics of the patient interval and proportion of patients that experienced intervals exceeding 60 days, by symptom combination

Symptom combination	N[1]	Mean	25th	50th[2]	75th	90th	% 60+ days
Abdominal pain alone	723	29	0	7	30	77	14%
CIBH alone	571	62	4	30	73	178	32%
Rectal bleeding alone	498	56	1	17	61	136	26%
Dysphagia alone	293	49	9	30	61	118	25%
Dyspepsia alone	125	41	0	17	38	92	19%
CIBH & Rectal bleeding	128	59	8	33	90	127	38%
Nausea or vomiting alone	100	43	2	14	33	99	17%
Bloating or distension alone	101	41	2	17	58	109	25%
Abdominal pain & CIBH	80	44	6	18	49	124	23%
Abdominal pain & Nausea or vomiting	51	24	1	14	35	67	14%
Abdominal pain & Bloating or distension	44	44	1	19	58	143	23%
Reflux alone	29	36	3	11	57	128	21%

CIBH: change in bowel habit. Symptom pairs are in bold print.
1 Number of patients (percentages) sum to 2,743 as 20% (n=695/3,438) of observations had missing information on the patient interval.
2 Kruskal-Wallis test of variation in median patient interval values across patients with one of 12 symptom combinations; p<0.001

7.5 Discussion

7.5.1 Main findings

Almost one in four cancer patients presented with abdominal symptoms before diagnosis. The majority of cancer patients who presented with abdominal symptoms were subsequently diagnosed with a range of common and rarer cancers of abdominal or adjacent organs, but a proportion of patients had tumours of other solid organ tumours, or haematological malignancies. The median patient interval ranged from 7 days for abdominal pain to 30 days for dysphagia. The observed differences in interval length by abdominal symptom remained when adjusted for age, sex, and ethnicity. Variation across patients with abdominal symptom combinations was comparable to that of patients with single abdominal symptoms alone.

7.5.2 Comparison with prior evidence

Colorectal, oesophageal, ovarian, and pancreatic cancers accounted for the majority of cancer patients that presented with one or more abdominal symptoms, consistent with previous evidence (Ebell *et al*, 2016; Walter *et al*, 2016a). There were also large proportions of patients diagnosed with rarer cancers such as stomach (65%), small intestinal (69%), and gallbladder cancers (51%) presenting with abdominal symptoms. These findings indicate that a generic abdominal symptom campaign may also help to expedite the diagnosis of these rarer abdominal cancers.

Comparable evidence on the association between the patient interval and abdominal symptoms is limited to two English studies on colorectal and pancreatic cancers respectively (Walter *et al*, 2016a, 2016b). Rectal bleeding and dyspepsia-like symptoms were associated with shorter time to presentation compared with other studied symptoms, in line with my findings (Walter *et al*, 2016a).

7.5.3 Strengths and limitations

Analyses were confined to eight abdominal symptoms based on those recommended for urgent referral in national clinical guidelines (NICE, 2015). This was a pragmatic decision that has face validity as symptom awareness campaigns are unlikely to include symptoms with a very low predictive value: there was a lot of overlap in symptom coverage with the Be Clear on Cancer abdominal symptoms campaign (see Figure 7.4).

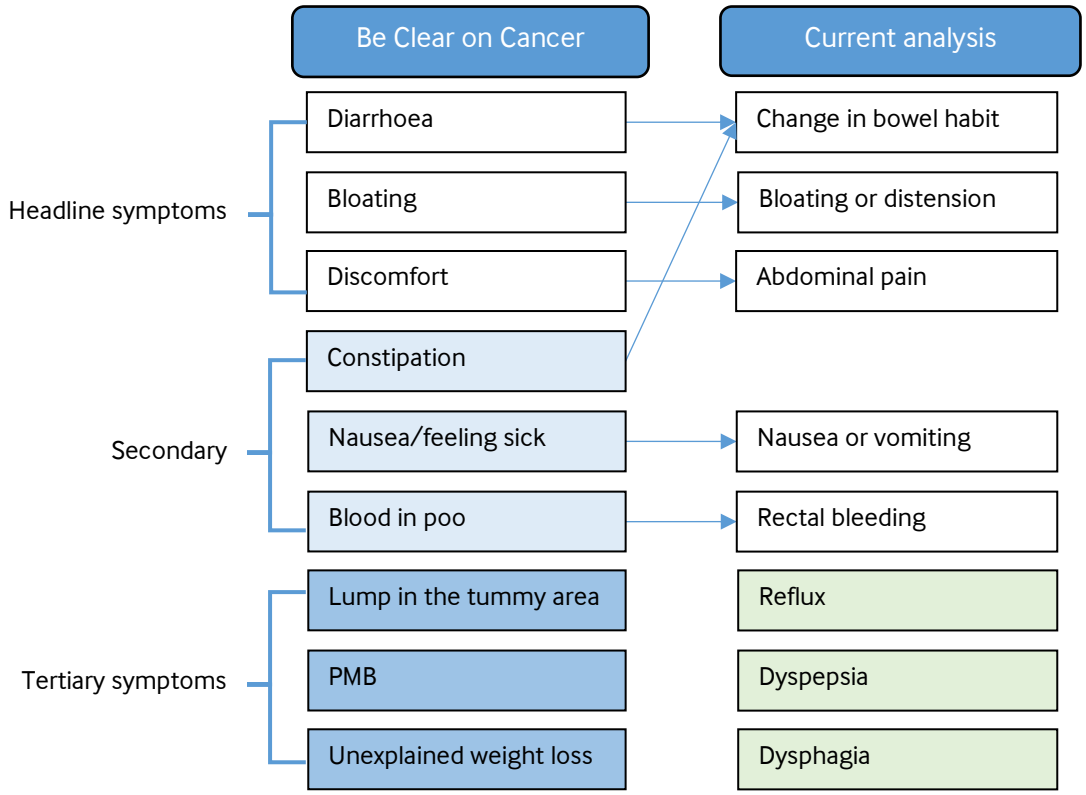


Figure 7.4 Comparison of Be Clear on Cancer (BCOC) campaign abdominal symptoms and the abdominal symptoms included in this chapter
NB BCOC symptoms were categorised as 'headline', 'secondary', or 'tertiary' symptoms based on phrasing and configuration of campaign materials. The 'tertiary' BCOC symptoms were not included in the current analysis, while three symptoms (reflux, dyspepsia, and dysphagia) were analysed but did not form part of the BCOC campaign.

Variation in patient interval was examined among patients with a single abdominal symptom (n=2,253, 62%) for ease of interpretation, again because campaign messages have thus far focused on single symptoms as opposed to synchronous symptom combinations. Nonetheless, supplementary analyses considering the four most frequent symptom combinations among 3,438/3,661 (94%) of cancer patients who presented with one or more abdominal symptom indicated concordant findings.

7.5.4 Implications

Abdominal symptoms appear to be common among incident cases of cancer, suggesting that symptom awareness campaigns focusing on abdominal symptoms could potentially contribute to the earlier diagnosis of a large range of both common and rarer cancers in the abdominal region.

Evidence regarding variation in the length of the patient interval associated with different symptoms could help to identify particular symptoms for prioritisation in campaigns. For example, one in two cancer patients with dysphagia waited almost a month before presenting. This finding is consistent with qualitative evidence indicating that dysphagia may initially be intermittent and mild, which can lead to it being normalised and explained away despite increasing interference with daily life (Lewis *et al*, 2017). Further, difficulty swallowing has been consistently shown to be one of the lesser well known symptoms of cancer among the general public (Robb *et al*, 2009; Cancer Research UK, 2016b). As dysphagia is also an established 'alarm' symptom for cancer, this finding argues for its further targeting by future campaigns (Stapley *et al*, 2013). In contrast, cancer patients with abdominal pain presented after a median interval of 7 days, and given its high prevalence and low predictive value, there may be little to be gained by raising its awareness among the general population (Hamilton *et al*, 2005b; Elnegaard *et al*, 2015).

Patients who presented with change in bowel habit (CIBH) and rectal bleeding had a longer patient interval compared to those who had CIBH or rectal bleeding alone (median (IQR) patient interval: 33 (8–90) days versus 30 (4–73) days and 17 (1–61) days respectively). Comparable differences in time to help-seeking have been noted among Danish colorectal patients (Pedersen *et al*, 2013); the added presence of CIBH may have contributed to greater normalisation and misattribution of rectal bleeding to non-malignant causes such as haemorrhoids.

The findings of this study relate to the significance of abdominal symptoms among patients subsequently diagnosed with cancer, but abdominal symptoms in primary care may represent other important diseases such as inflammatory bowel disease (Walter *et al*, 2016a; Stapley *et al*, 2017). Understanding the prevalence of abdominal symptoms among the general population, the predictive values of symptoms for cancer, and the potential diagnostic experiences of patients that seek help for such symptoms beyond the cancer context is likely to be an equally important consideration for early diagnosis interventions (Whitaker *et al*, 2014; Elnegaard *et al*, 2017).

Previous evaluations of Be Clear on Cancer campaigns have examined the increase in number of 2-week-wait referrals, the corresponding conversion rates to cancer cases, and diagnostic activity following campaign launches (Cancer Research UK, 2016a). For campaigns targeting

groups of symptoms, understanding the anticipated range of affected cancer sites could be helpful for directing the assessment of the campaign's impact.

Importantly, the study findings indicate that a small but important group of other solid tumour cancers are also diagnosed after presentation with an abdominal symptom; I examine symptom-specific cancer signatures in the next chapter (Chapter 8: Abdominal symptoms and time to referral). Raising awareness of cancer symptoms will not always identify those with early (more treatable) disease, and while expediting the diagnosis of cancer among symptomatic patients is likely to remain preferred by patients and relatives, better understanding of which presenting symptoms of cancer are associated with early versus late disease could be insightful. These realisations motivated further analysis as described in Chapter 9 (Cancer alarm symptoms and stage at diagnosis).

7.6 Chapter summary

Public health education campaigns targeting abdominal symptoms have the potential to reduce time to help-seeking across a range of common and rarer cancers. The timeliness of presentation associated with individual symptoms could inform the design of campaigns, in conjunction with other aspects of epidemiological evidence.

Building on the findings of Chapter 6 (The symptom signature of breast cancer and associated diagnostic intervals), the findings of this study support substantial variation in time to help-seeking by presenting symptom among patients subsequently diagnosed with cancer and demonstrate the opportunities for improving the design of public health education campaigns based on such evidence.

Chapter 8: Abdominal symptoms and time to referral

This chapter examines the spectrum of subsequently diagnosed cancer sites among cancer patients who present with a specific abdominal symptom, and associations between abdominal symptoms and time to referral in the context of post-presentational interventions promoting early diagnosis.

8 Abdominal symptoms and time to referral

8.1 Rationale of this chapter

In the previous chapter (Chapter 7: Abdominal symptoms and time to presentation), I found that cancer patients who had presented with abdominal symptoms had variable time to presentation, and were subsequently diagnosed with a broad range of cancers. In this chapter I examine this group of cancer patients post-presentation.

8.2 Introduction

Abdominal symptoms at presentation are common among patients subsequently diagnosed with cancer (Holtedahl *et al*, 2017). As described in Chapter 1, fast-track referral pathways for patients with suspected cancer supported by accompanying clinical guidelines have been developed to expedite investigation and diagnosis in individuals who present with abdominal symptoms of relatively high predictive value for cancer (Healthcare Improvement Scotland, 2014; NICE, 2015; Cancer Council Austalia, 2017). However, many abdominal symptoms are not associated with a single cancer site.

Fewer than 40% of all cancer patients are diagnosed through the 2-week-wait pathways in England (National Cancer Registration and Analysis Service, 2017). These realisations have led policy-makers to pilot novel diagnostic pathways and services to complement the 2-week wait referral system. For example, multi-disciplinary diagnostic services based in secondary care are currently being piloted for patients presenting with serious but vague abdominal symptoms (see Figure 8.1) (UCLH, 2015; Vedsted & Olesen, 2015; Cancer Research UK, 2016c; Fuller *et al*, 2016).

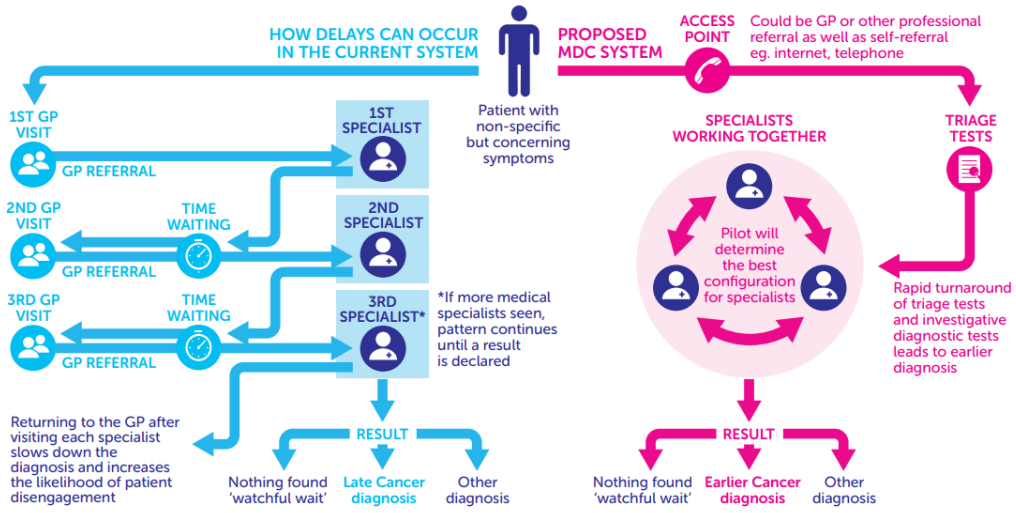


Figure 8.1 The hypothesised diagnostic pathway of cancer patients who present with non-specific symptoms (Cancer Research UK, 2016)

Symptomatic patients who initially present to primary care or emergency departments and fit referral criteria for these novel diagnostic services can be referred to a Multidisciplinary Diagnostic Centre (MDC) where a range of specialist investigations can be performed, to achieve faster diagnostic resolution (London Cancer, 2017). There is however limited evidence on the optimal sequence (cascade) of investigations that could be used in these patients. Understanding the range of different cancers that are associated with a given abdominal symptom could contribute to the configuration of these novel diagnostic pathways.

Relatedly, many existing early diagnosis interventions such as the 2-week-wait referral system implicitly link specific alarm symptoms with a given cancer site, for example rectal bleeding mandates referral to lower gastro-intestinal specialists for suspected colorectal cancer (NHS Digital, 2016). However, not all cancer patients presenting with a specific alarm symptom will be subsequently diagnosed with the cancer typically associated with it. This group, which includes a large proportion of patients with abdominal symptoms, experience complex and prolonged diagnostic pathways, as recently reported by a Danish study (Nielsen *et al*, 2018).

Against the above background, I aimed to examine the spectrum of subsequently diagnosed cancers among patients who had presented with abdominal symptoms, and variation in the primary care interval (time from symptomatic presentation to referral to a specialist) by presenting abdominal symptom.

8.3 Methods

8.3.1 Study population

As described in Chapter 7, the analysis sample comprised 15,956 cancer patients included in the NACDPC with complete and valid information on age group (among patients aged fifteen years or older), sex, and presenting symptoms, and excluded individuals diagnosed incidentally or asymptomatically and those with cancer sites categorised as "No information" and "Unknown Primary" (see Figure 7.1 on page 101 for flow chart of sample derivation).

8.3.2 Abdominal symptom definition and classification as 'alarm' or 'non-alarm' symptoms

Abdominal symptoms were defined as described previously in Chapter 7, informed by the abdominal symptoms described by the 2015 NICE guidelines for suspected cancer referral (NICE, 2015). The symptoms were further classified as 'alarm' or 'non-alarm' based on recommendations in the 2015 NICE guidelines: symptoms for which the guidelines suggested urgent investigation or referral via the 2-week-wait pathway (in the absence of other symptoms) were considered to constitute alarm symptoms, while those for which the guidelines recommended referral, or non-urgent investigations were considered non-alarm symptoms (NICE, 2015) (see Table 8.1).

Table 8.1 Abdominal symptom definitions based on NICE 2015 guidelines as listed in Table 7.1, with additional categorisation of symptoms as 'alarm' or 'non-alarm'

Symptom	Symptom constructs
Alarm	
Rectal bleeding	Rectal bleeding[3]
Dysphagia	Odynophagia
	Dysphagia
Change in Bowel Habit	Constipation
	Change in bowel habit
	Diarrhoea
Non-alarm	
Abdominal Pain	Abdominal pain[1]
	Epigastric pain
	Right Iliac Fossa (RIF) pain
	Suprapubic pain
	Loin pain & renal colic
Dyspepsia	Dyspepsia and related epigastric symptoms[2]
Reflux	Reflux
Bloating or Distension	Abdominal bloating
	Abdominal distension
	Ascites
Nausea or Vomiting	Vomiting
	Nausea

1 Abdominal pain that was not otherwise specified, excluding acute abdominal pain
2 includes dyspepsia, indigestion, waterbrash, gastritis, burping, belching, "GI upset", and "upper GI symptoms".
3 includes blood in stool and rectal bleeding, excludes acute rectal bleeding

8.3.3 Variables of interest

For the purposes of examining the cancer signature of each abdominal symptom, cancer sites were grouped into three categories as for Chapter 7, namely:

1) abdominal or adjacent organ cancers (cancers arising in the intra-abdominal organs, together with oesophageal and prostate cancer);
2) solid tumour malignancies other than the above (i.e. cancers of non-abdominal organs) hereafter called 'other solid tumours' cancers; or
3) haematological cancers (lymphoma, leukaemia, and myeloma).

The interval of interest in this analysis was the primary care interval (time from first presentation in primary care to referral to specialist services). Information on the primary care interval was highly complete among the study population (90% of 2,253 cancer patients with a single abdominal symptom, see Section 8.3.4 below for justification of sample size) and so I conducted complete case analysis treating the primary care interval as a continuous variable. Patients missing information on the primary care interval were more likely to have first

presented in places other than general practice, and to have presented with nausea or vomiting, without evidence for variation in missing primary care interval by age or sex (see Appendix 8.1).

8.3.4 Statistical analyses

In Chapter 7, I described the range and relative frequencies of cancer sites that were subsequently diagnosed among incident cancer patients presenting with abdominal symptoms, namely the cancer signature of abdominal symptoms. Here, I continued this line of enquiry by describing the corresponding cancer signature of each individual abdominal symptom (cancer site frequencies are presented in Appendix 8.2).

Subsequently, I examined variation in the primary care interval by abdominal symptom using the same methodology as in Chapter 7 for examining variation in the patient interval. As before, the main analysis was restricted to the majority of cancer patients with a single recorded presenting abdominal symptom (n=2,253, 62% of all patients reporting an abdominal symptom), and common abdominal symptom combinations were examined as part of supplementary analyses.

Firstly, the mean, median, interquartile range and 90th centiles of the primary care interval were estimated for each abdominal symptom. These were estimated along with 95% confidence intervals using a bootstrap approach with 1000 replications. A generalised linear model (GLM) was used to test variation in primary care interval by abdominal symptom, adjusted for age group (parameterised as <50 years, 50–69 years, 70+ years), sex (men, women), and ethnicity (white, non-white) as in the analyses presented in Chapter 7. To account for skewed outcome data, a log link function was used (which allows the covariates to be modelled on a linear additive scale), and significance testing was again based on bootstrapping. Variation in interval length across categorical variables was examined using joint Wald tests, with statistical significance at the 5% level.

Additionally, for patients who had presented with each of the three abdominal alarm symptoms, I compared the median length of primary care interval between cancer patients who were diagnosed with the 'typical' cancer associated with the symptom versus those who were diagnosed with a 'non-typical' cancer (for that symptom). 'Typical' cancers associated with each of the three abdominal alarm symptoms were operationally defined as the cancer sites accounting for at least 75% of all cancer cases presenting with the studied symptom, i.e.:

- **Colorectal cancer** being the typical cancer for *change in bowel habit*;
- **Colorectal cancer** being the typical cancer for *rectal bleeding*; and
- **Oesophago-gastric cancer** (treated as a single site) being the typical cancer for dysphagia.

I repeated the above analyses using the referral interval (see Section 8.3.5.2 below).

8.3.5 Supplementary analyses

8.3.5.1 Patients with multiple abdominal symptoms

Similar to the supplementary analyses conducted in Chapter 7, I examined patients who presented with one of four most common pairs of abdominal symptom in addition to the eight abdominal symptoms alone, their cancer signatures, and their associations with the length of the primary care interval. This was among 3,438/3,661 (94%) patients who presented with one or more abdominal symptom. Findings are summarised in Section 8.4.4 below and Appendix 8.3.

8.3.5.2 Variation in the referral interval by abdominal alarm symptom

I hypothesised that patients with the same symptomatic presentation but different subsequent cancer diagnosis could experience variation in diagnostic timeliness beyond the primary care interval. Therefore, I compared the median length of the referral interval (from first presentation in primary care to first consultation in secondary care) between cancer patients who were diagnosed with the 'typical' cancer(s) (defined as above), and those diagnosed with a 'non-typical' cancer. The Aarhus statement does not include a definition of the referral interval; I defined the referral interval as the number of days from first presentation in primary care to first secondary care consultation, thereby being an interval inclusive of the primary care interval (see Figure 8.2). Findings are summarised in Section 8.4.5 below and Appendix 8.4.

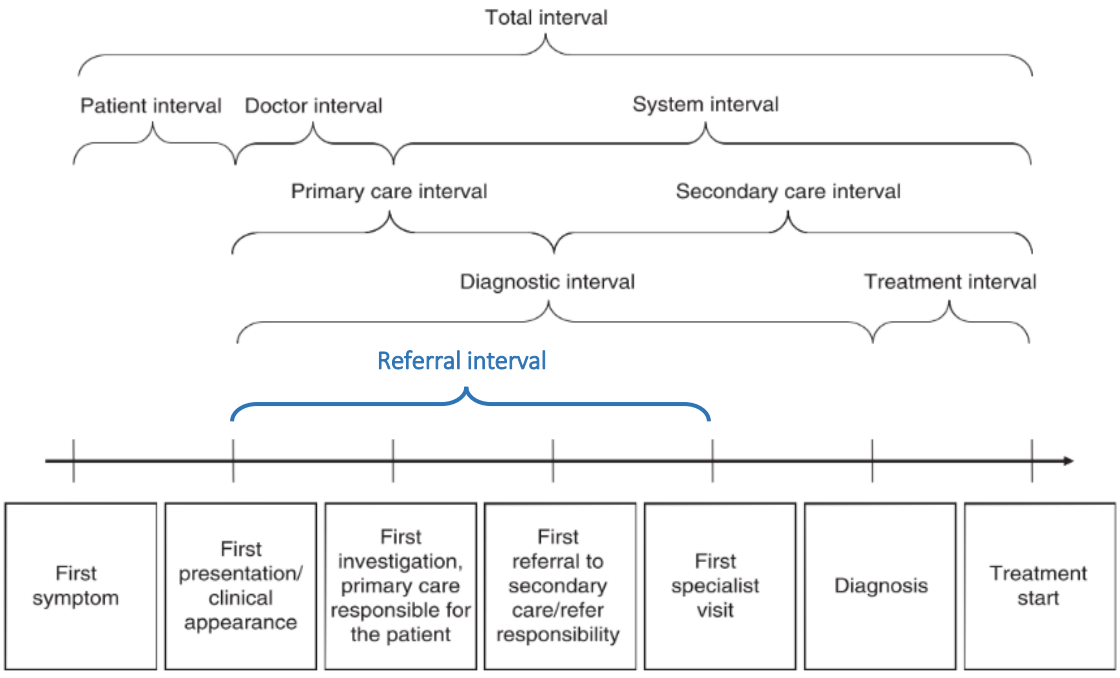


Figure 8.2 The Aarhus statement pathway model (Weller et al, 2012) with the referral interval as defined in this chapter in blue
NB the referral interval (time from first presentation in primary care to first specialist consultation in secondary care) **includes** the primary care interval.

8.4 Results

As reported in Chapter 7, there were 3,661/15,956 (23%) patients who presented with one or more of the eight abdominal symptoms (see Table 7.2, page 104). The cancer site case-mix varied by symptom (see Table 8.2 and Figure 8.3 below).

8.4.1 The cancer signature of abdominal alarm symptoms

The three examined abdominal alarm symptoms were associated with a specific cancer site, however a substantial percentage of patients had cancers of other organs.

The majority (92%) of cancer patients who had presented with rectal bleeding were later diagnosed with colorectal cancer, though 1 in 12 patients (n=63) were diagnosed with one of 12 different cancer types (mostly abdominal or adjacent organ cancers).

Similarly, the majority (78%) of cancer patients who had presented with change in bowel habit were later diagnosed with colorectal cancer, though more than a fifth (22%) were diagnosed with one of 19 cancers including ovarian cancer (5%), prostate cancer (3%), and pancreatic cancer (3%).

Lastly, 18% of patients presenting with dysphagia were diagnosed with a cancer other than oesophageal or stomach cancer, including oropharyngeal cancer (5%) and lung cancer (4%).

Table 8.2 Proportion of abdominal or adjacent organ, other solid tumours, and haematological cancers diagnosed following presentation with an abdominal symptom

	Abdominal or adjacent organ cancer†	Other solid tumour cancer†	Haematological cancer†
Abdominal alarm symptom			
Change in bowel habit	94%	4%	2%
Rectal bleeding	95%	3%	1%
Dysphagia	85%	13%	2%
All alarm symptoms	92%	6%	2%
Abdominal non-alarm symptom			
Abdominal pain	84%	9%	6%
Nausea or vomiting	76%	18%	6%
Dyspepsia	90%	5%	5%
Bloating or distension	89%	8%	4%
Reflux	96%	3%	1%
All non-alarm symptoms	85%	10%	6%
Any abdominal symptom	89%	8%	4%

†**Abdominal or adjacent organ cancers:** colorectal, oesophageal, ovarian, pancreatic, stomach, prostate, renal, bladder, liver, small intestine, gallbladder, and endometrial cancers. **Other solid tumour cancers (non-abdominal cancers and other cancers excluding oesophageal and prostate cancer):** lung, oropharyngeal, breast, laryngeal, brain, cervical, sarcoma, testicular, melanoma, mesothelioma, thyroid, vulval, and 'other' cancers. **Haematological cancers:** lymphoma, leukaemia, and myelomas.
NB It is likely that a proportion of sarcomas and lymphomas were intra-abdominal but information regarding their exact location was not available.

8.4.2 The cancer signature of abdominal non-alarm symptoms

Abdominal non-alarm symptoms tended to have a more heterogeneous cancer site composition among incident cases, including greater proportions of other solid tumour cancers and haematological cancers.

Among cancer patients presenting with abdominal pain, a third (33%) were subsequently diagnosed with colorectal cancer, while 12% were diagnosed with ovarian cancer, and 10% with pancreatic cancer.

Similarly, patients who presented with nausea or vomiting were most commonly diagnosed with colorectal cancer (24%), followed by pancreatic cancer (13%), oesophageal cancer (11%), and stomach cancer (11%). Around a quarter (24%) of patients who presented with nausea or vomiting were diagnosed with an 'other solid tumour' cancer or haematological cancer.

Patients who had presented with dyspepsia were subsequently diagnosed with oesophageal cancer (37%), stomach cancer (24%), colorectal cancer (12%), or pancreatic cancer (10%), with a further 11% of patients being diagnosed with cancer of another abdominal or adjacent organ.

Patients presenting with bloating or distension were most likely to be diagnosed with ovarian cancer (45%) than colorectal cancer (20%), and a further 27% of patients were diagnosed with a cancer of an abdominal or adjacent organ. Stratification by sex indicated that among women presenting with bloating or distension, 61% were diagnosed with ovarian cancer while 14% were diagnosed with colorectal cancer (by comparison, 37% of men who presented with bloating or distension were diagnosed with colorectal cancer, see Appendix 8.2).

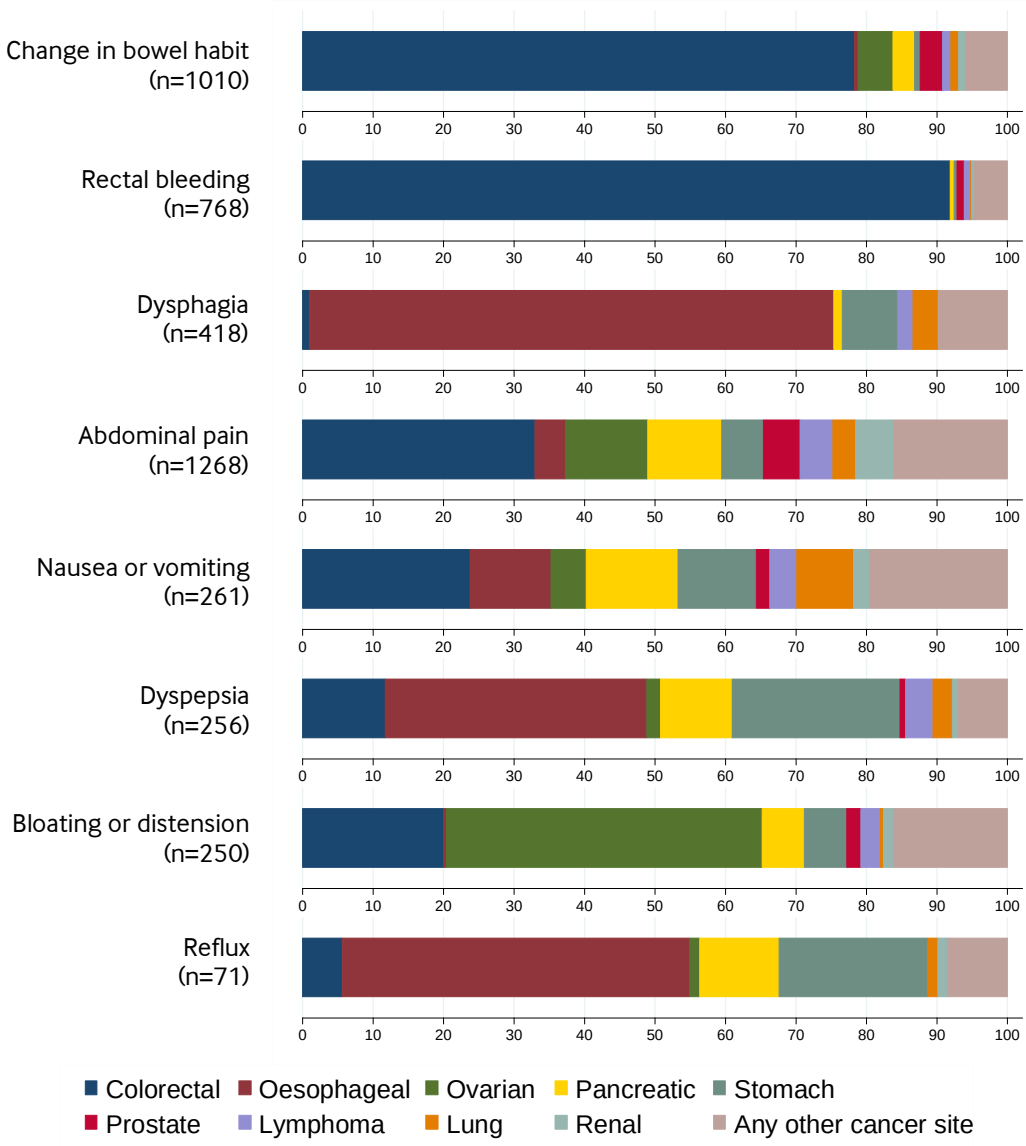


Figure 8.3 The cancer site case-mix of abdominal alarm and abdominal non-alarm symptoms
See Appendix 8.2 for exact proportions.

8.4.3 Primary care interval by presenting abdominal symptom

Among cancer patients with a single presenting abdominal symptom (n=2,253), the length of the primary care interval varied by individual symptom (see Figure 8.4 and Table 8.3). Dysphagia had the shortest median primary care interval (median (IQR) interval: 1 (0–14) days) followed by rectal bleeding (median (IQR) interval: 1 (0–12) days), while patients with dyspepsia and reflux had the longest such intervals before referral to specialist care (median (IQR) interval: 22.5 (5–75) days and 23.5 (7–69) days respectively).

The proportion of patients with each abdominal symptom who experienced 3 or more pre-referral consultations was largely concordant with the ranking of the length of the primary care interval by symptom (Table 8.3). Regression analyses indicated that the variation in interval

length by abdominal symptom remained after adjusting for age group, sex, and ethnicity (joint Wald test p<0.001) (Table 8.3).

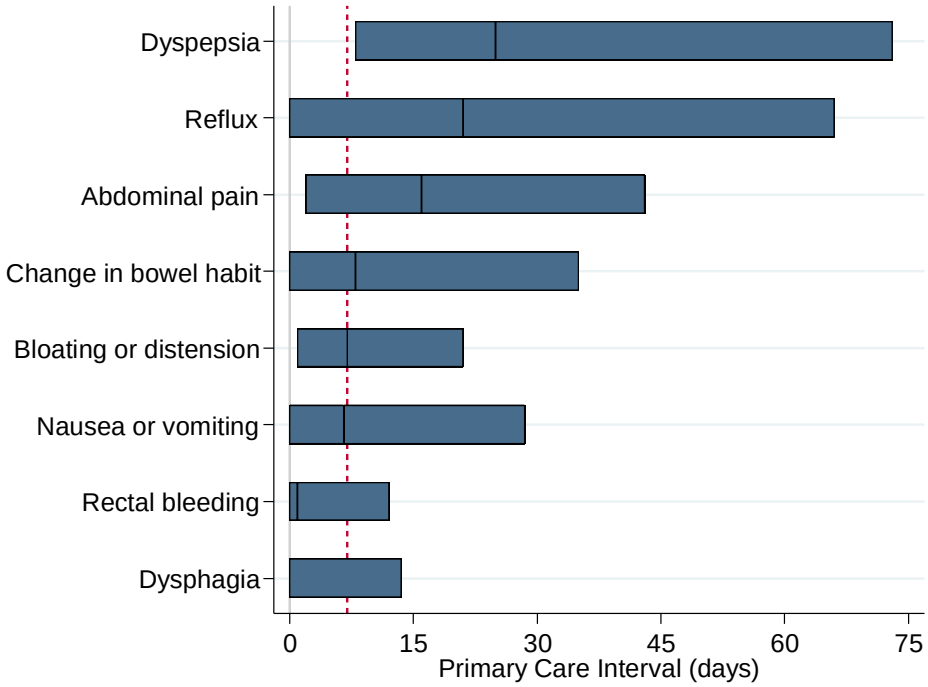


Figure 8.4 The length of the primary care interval by presenting abdominal symptom
Among patients with a single abdominal symptom (n=2,017 as 10% had missing primary care interval values). Bar length = IQR, vertical line = median value; the red dashed vertical line represents the median primary care interval value (7 days) across all abdominal symptoms. Symptoms are ordered by median interval length.

Table 8.3 Summary statistics of the primary care interval (measured in days), proportion of 3+ pre-referral consultations, and adjusted GLM coefficient by abdominal symptom among patients with a single abdominal symptom

Symptom	N	Mean	25th	50th	75th	90th	% 3+ consultations	Adjusted GLM coefficient†[1]
Dyspepsia	118	53	8	25	73	124	36%	1.6 (1.0–2.8)
Reflux	29	55	0	21	66	148	35%	1. (0.5–2.0)
Abdominal pain	670	36	2	16	43	85	36%	Ref.
Change in bowel habit	525	34	0	8	35	100	23%	1.0 (0.7–1.4)
Bloating or distension	96	24	1	7	21	58	23%	0.6 (0.3–1.1)
Nausea or vomiting	53	25	0	7	29	52	40%	0.8 (0.4–1.6)
Rectal bleeding	495	23	0	1	12	63	8%	**0.6 (0.4–0.8)**
Dysphagia	267	14	0	0	14	45	9%	**0.4 (0.3–0.6)**
All abdominal symptoms	2253	30	0	7	32	79	23%	-

† 21% of observations had missing information on the primary care interval or ethnicity resulting in a final model size of n=1770.
1 Joint Wald test of variation in primary care interval by symptom: p<0.001. The exponentiated coefficient represents the factor difference applicable to the geometric mean patient interval of the reference group. For example, an exponentiated coefficient of 2 for a symptom would mean that on average, the patient interval was 2 times longer among patients with that symptom compared to patients with abdominal pain (reference group), adjusted for sex, age group, and ethnicity.

There was no difference in the median primary care interval between cancer patients who presented with an abdominal alarm symptom and were diagnosed with the 'typical' cancer, and cancer patients who were diagnosed with a different cancer (Table 8.4).

Table 8.4 Median and IQR primary care interval values of cancer patients who presented with an abdominal alarm symptom and were diagnosed with the 'typical' cancer versus 'non-typical' cancer

Symptom	N	Median (IQR) primary care interval (days)	Wilcoxon rank-sum test for median values
Cancer patients with CIBH	489	8 (0–35)	
CRC patients with CIBH	412	9 (0–38)	0.971
Other cancer patients with CIBH	77	7 (1–28)	
Cancer patients with rectal bleeding	454	1 (0–12)	
CRC patients with rectal bleeding	422	1 (0–13)	0.416
Other cancer patients with rectal bleeding	32	2 (0–10)	
Cancer patients with dysphagia	256	0(0–14)	
OG cancer patients with dysphagia	227	0 (0–14)	0.913
Other cancer patients with dysphagia	29	0 (0–7)	

CRC: colorectal cancer; CIBH: change in bowel habit; OG: oesophago-gastric cancer

8.4.4 Supplementary analysis: patients with multiple abdominal symptoms

Further analyses examining patients with four most common abdominal symptom combinations (as described in Section 7.4.4 with regards to the patient interval) indicated largely comparable cancer signatures and associations with the primary care interval (see Appendix 8.3).

8.4.5 Supplementary analysis: variation in the referral interval by abdominal alarm symptom

There was no evidence of variation in the length of the referral interval (time from first presentation in primary care to first secondary care consultation) between patients who presented with one of the three abdominal alarm symptoms and were diagnosed with 'typical' (for those symptoms) cancer, and those who presented with the same symptoms but were diagnosed with a 'non-typical' cancer (see Appendix 8.4).

8.5 Discussion

8.5.1 Main findings

The cancer signatures of the three examined abdominal alarm symptoms (change in bowel habit, rectal bleeding, and dysphagia) were dominated by colorectal and oesophago-gastric cancers – each comprising more than 75% of all cancer cases who presented with the respective symptom. However, a substantial minority of patients were subsequently diagnosed with other cancers which were 'non-typical' for the respective presenting symptom. By contrast, abdominal non-alarm symptoms tended to have a more heterogeneous cancer signature, including a relatively high frequency of patients with non-abdominal or adjacent organ cancers.

The length of the primary care interval varied by abdominal symptom. There was no variation in the length of the primary care interval between patients who presented with an abdominal alarm symptom and were diagnosed with the 'typical' cancer compared with those who presented with the same symptom but were diagnosed with a 'non-typical' cancer. Supplementary analyses did not find any difference in length of the referral interval.

8.5.2 Comparison with prior evidence

The majority of cancer patients who presented with change in bowel habit, rectal bleeding, or dysphagia were subsequently diagnosed with colorectal cancer or oesophago-gastric cancer respectively. However, a proportion of patients were diagnosed with other 'non-typical' cancers including other solid tumour cancers and haematological cancers. The findings are concordant with evidence indicating that change in bowel habit, rectal bleeding, and dysphagia have relatively high predictive values for multiple cancer sites (Hippisley-Cox & Coupland, 2013a, 2013b), but this study is the first in my knowledge to document the frequencies of these 'non-typical' cancers.

Variation in time to referral by abdominal symptom was consistent to evidence on time to diagnosis and the presenting symptoms of colorectal cancer patients as described in three relevant studies commented in Chapter 3 (Pruitt *et al*, 2013; Walter *et al*, 2016a; Leiva *et al*, 2017). Rectal bleeding and change in bowel habit tended to be associated with shorter intervals while dyspepsia-like symptoms, and vomiting or nausea were associated with a longer interval (Pruitt *et al*, 2013; Walter *et al*, 2016a). Similar patterns of variation were reported among Spanish colorectal cancer patients based on diagnostic intervals based on patient records and primary care records, although those based on hospital records indicated that vomiting and abdominal pain were associated with shorter time to diagnosis than rectal bleeding (Leiva *et al*, 2017).

There was no evidence to support a difference in the length of the primary care interval (and the related referral interval) between cancer patients who presented with an abdominal alarm symptom and were diagnosed with a 'typical' cancer, compared to those who presented with the same symptom but were diagnosed with a 'non-typical' cancer. The findings suggest that diagnostic timeliness is not affected by the subsequently diagnosed cancer for patients who present with abdominal alarm symptoms even if referral pathways tend to be site-specific. This is consistent with recent findings from a prospective study of colorectal cancer, which identified minimal differences in time from first presentation to diagnosis between patients who presented with rectal bleeding and were diagnosed with colorectal cancer, versus those who presented with rectal bleeding but were not subsequently diagnosed with cancer (Walter *et al*, 2016a). However, the primary care interval and referral interval may not have fully captured the potential complexity of the diagnostic pathway for cancer patients presenting with abdominal

symptoms. Recently published evidence from Denmark indicates that the most common urgent referral pathway made within 6 months of a negative first referral for cancer is for the gastro-intestinal system, suggesting that cancer site-specific referral systems could lead to more complex (and longer) diagnostic pathways among those presenting with abdominal symptoms (Nielsen *et al*, 2018).

8.5.3 Strengths and limitations

There is no universal consensus on how abdominal symptoms should be defined; I focused on eight abdominal symptoms (as in Chapter 7) based on their inclusion in clinical guidelines on referral for suspected cancer (NICE, 2015). Similarly, there is no uniform consensus on the definition of an 'alarm' symptom of cancer, and so my operational definitions were based on NICE clinical guidelines. I further explore the ways in which an alarm symptom may be defined in Chapter 9 (Cancer alarm symptoms and stage at diagnosis) (see Section 9.3.4, page 134).

The findings are based on patients who were subsequently diagnosed with cancer, which provides only a partial picture of diagnostic timeliness among individuals who seek help for abdominal symptoms; as also discussed in Chapter 7, abdominal symptoms are highly prevalent in primary care and can be a sign of other serious but benign diseases (Holtedahl *et al*, 2017; Stapley *et al*, 2017).

8.5.4 Implications

The relative contribution of the patient and primary care intervals to the combined pre-referral interval have previously been described by cancer site (Lyratzopoulos *et al*, 2015). The findings of this chapter in combination with those of Chapter 7 highlight a contrasting pattern of variation in the length of the patient and the primary care intervals by presenting abdominal symptom (see Figure 8.5).

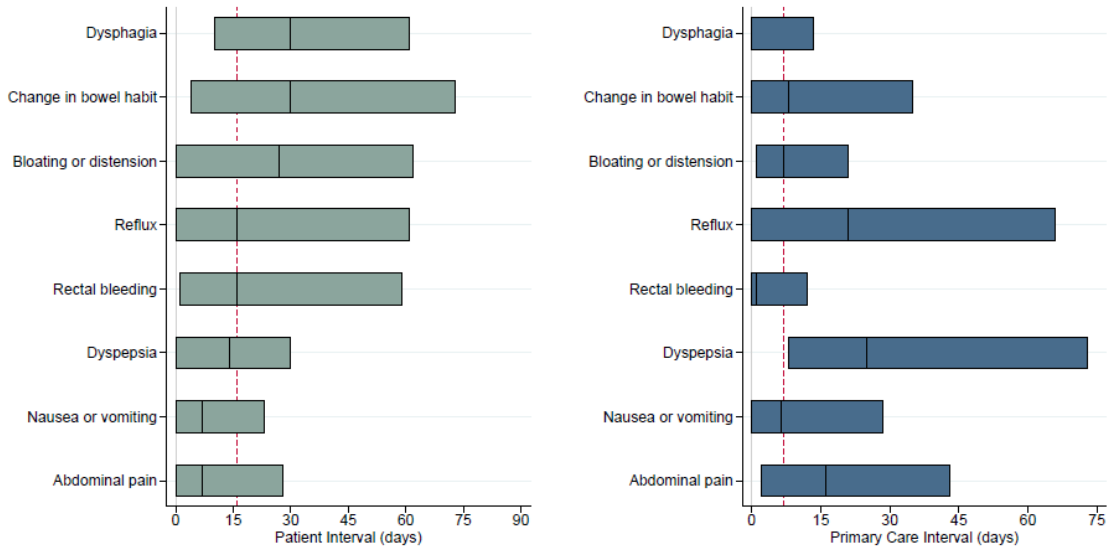


Figure 8.5 The length of the patient and primary care interval by abdominal symptom at presentation
As presented in Figure 7.3 and Figure 8.4 respectively; symptoms are ordered by median patient interval.

For example, dysphagia had the shortest primary care interval (median: 0 days) among the studied abdominal symptoms and one of the longest patient intervals (median: 30 days). Similar observations can be made for rectal bleeding (a median of 16 days for the patient interval, and 1 day for the primary care interval). For these abdominal alarm symptoms, early diagnosis efforts targeting the patient interval may be more meaningful over efforts directed at shortening post-presentation intervals, given that for most patients, the patient interval accounted for a greater proportion of the combined interval and referral was usually prompt.

By contrast, cancer patients with presenting symptoms such as dyspepsia and abdominal pain tended to present relatively promptly but on average, experienced prolonged intervals to referral. Prompt presentation is likely to be explained by the relatively serious or persistent nature of the symptom in the context of the patient's subsequently diagnosed cancer; for example, nausea and vomiting caused by malignancy is likely to be persistent and non-resolving which would therefore lead to relatively prompt help-seeking compared to other abdominal symptoms. Such symptoms are known to have low predictive values for cancer (Astin *et al*, 2011, 2015) and are not recommended for urgent referral in current clinical guidelines in the absence of other symptoms (NICE, 2015).

Improving access to imaging and other investigative services in primary care may help to reduce clinical uncertainty and accelerate diagnostic resolution (Independent Cancer Taskforce, 2015; Vedsted & Olesen, 2015; London Cancer, 2017; Rubin *et al*, 2018). Indeed, the findings of this chapter could support the design of such multidisciplinary diagnostic centres (MDCs) for serious non-specific symptoms. For example, patients with abdominal pain could be investigated for colorectal cancer (e.g. with a colonoscopy) as a first priority, given that the results indicate that – in the context of symptomatic cancer diagnosis – one in three patients would be subsequently diagnosed with colorectal cancer.

Relatedly, together with other evidence, the findings can also contribute to the choice of diagnostic strategies for patients referred to secondary care in order to maximise the effectiveness and cost-effectiveness of these services, for example, the selection and sequence of diagnostic tests used by multidisciplinary diagnostic centres (MDCs) for serious non-specific symptoms (Independent Cancer Taskforce, 2015; Ingeman *et al*, 2015; London Cancer, 2017).

8.6 Chapter summary

Time to referral (the primary care interval) varied by abdominal symptom, but the pattern of variation contrasted the variation in time to presentation (patient interval) observed for patients with the same abdominal symptoms as described in Chapter 7. Understanding such associations can guide the targeting of early diagnosis interventions in healthcare, such as the design of novel diagnostic pathways and investigation cascades.

Chapter 9: Cancer alarm symptoms and stage at diagnosis

Following the previous chapters that have examined the epidemiology of symptomatic presentation and associated lengths of the patient and primary care intervals, this chapter explores the relationship between alarm symptoms at presentation and recorded stage at diagnosis of subsequently diagnosed cancer.

9 Cancer alarm symptoms and stage at diagnosis

9.1 Rationale for this chapter

The previous chapters of this thesis examined presenting symptoms of cancer patients and associated diagnostic timeliness in order to contribute to the evidence base underlying early diagnosis interventions, which are chiefly centred on the presenting symptoms of cancer.

This chapter continues exploring the epidemiology of presenting symptoms among cancer patients by examining their associations with stage at diagnosis. While this enquiry is relevant for any of the multitude of presenting symptoms in cancer patients, I chose to focus on a range of 'alarm' symptoms of cancer given that they tend to be the focus of most early diagnosis interventions.

9.2 Introduction

As alluded to in Chapter 1, interventions that promote the earlier diagnosis of cancer typically target individuals who present with symptoms indicative of cancer. These are most often 'alarm' symptoms with relatively high predictive value for cancer.

For example, public health education campaigns focus on raising the awareness of possible cancer alarm symptoms such as rectal bleeding, haematuria, and post-menopausal bleeding (Danish National Board of Health, 2005; Scottish Government, 2012; Cancer Australia, 2013; Public Health England, 2016a; CDC, 2017). Similarly, clinical referral guidelines and associated diagnostic pathways mandate referral and/or investigation for symptomatic patients conditional on the presence or absence of an alarm symptom (IKNL [Netherlands Comprehensive Cancer Organisation], 2012; Probst *et al*, 2012; Healthcare Improvement Scotland, 2014; NICE, 2015; Cancer Council Austalia, 2017).

Diagnosing cancer earlier is likely to be beneficial for patient experience and cost-effectiveness of treatment (Smith & Hillner, 2011; Brocken *et al*, 2012; Robinson *et al*, 2012; Blumen *et al*, 2016; Mendonca *et al*, 2016; Dahl *et al*, 2017), but interventions principally aim to detect cancer at an earlier stage, thereby improving cancer survival (Coleman *et al*, 2011; Hiom, 2015). For example, the Detect Cancer Early (DCE) Programme in Scotland aimed to help increase the proportion of breast, colorectal, and lung cancers diagnosed in TNM stage I by 25% (Information Services Division, 2017).

However, alarm symptoms may not represent early stage disease (Jensen *et al*, 2015; Neal *et al*, 2015). For example, while dysphagia is considered an alarm symptom for oesophago-gastric cancer (Stapley *et al*, 2013), it is also perceived as a sign of advanced disease stage (Maconi *et al*, 2008; Bird-Lieberman & Fitzgerald, 2009).

Given the above, I aimed to examine the association between alarm symptom status at presentation and stage at diagnosis, and also compare the relative variation in odds of late stage disease between different alarm symptoms.

9.3 Methods

9.3.1 Study population

The analysis sample comprised cancer patients included in the NACDPC with complete and valid information on age (among patients aged fifteen years or older), sex, presenting symptoms, and stage at diagnosis (see Figure 9.1 for sample derivation). Patients with a patient interval of 30 days or longer were excluded (see Section 9.3.3 below for justification). As described in Chapter 4, individuals diagnosed incidentally or asymptomatically and those with cancer sites categorised as "No information" and "Unknown Primary" were excluded from the analysis (see Figure 9.1 for sample derivation).

Additionally, leukaemia, lymphoma, and multiple myeloma patients were excluded from the study given that information on stage at diagnosis was available based on an adapted SEER LRD classification which is not optimised for haematological cancers (Young *et al*, 2001). The final analysis sample therefore comprised 6,988 patients (see Figure 9.1).

9.3.2 Outcome of interest

Information on the stage at diagnosis of patients included in the NACDPC was provided by the auditing clinicians, based on information included in clinical records, using an adaptation of the SEER (local/regional/distant disease) classification (Rubin *et al*, 2011).

For this analysis, stage was parameterised as early (local/regional disease) or late (distant). Information on stage was available for 91% of the patients who otherwise met criteria for inclusion in the analysis sample, enabling complete case analysis (Figure 9.1). Nevertheless, I conducted supplementary analyses assuming that all patients with missing stage information had had late stage at diagnosis (see Section 9.3.6.1 below).

9.3.3 Additional considerations of the analysis sample

Early diagnosis interventions aim to achieve prompt presentation and referral of patients soon after symptom onset. For example, the Be Clear on Cancer campaign promotes help-seeking either immediately (for acute symptoms such as rectal bleeding or haematuria) or 3 weeks after onset (for persistent symptoms such as change in bowel habit) (Public Health England, 2016a). Therefore, the ideal study for examining associations between presenting symptoms of cancer and stage at diagnosis would measure stage as close to symptom onset as possible, thereby focusing on the patient population of greatest prior interest for early diagnosis interventions.

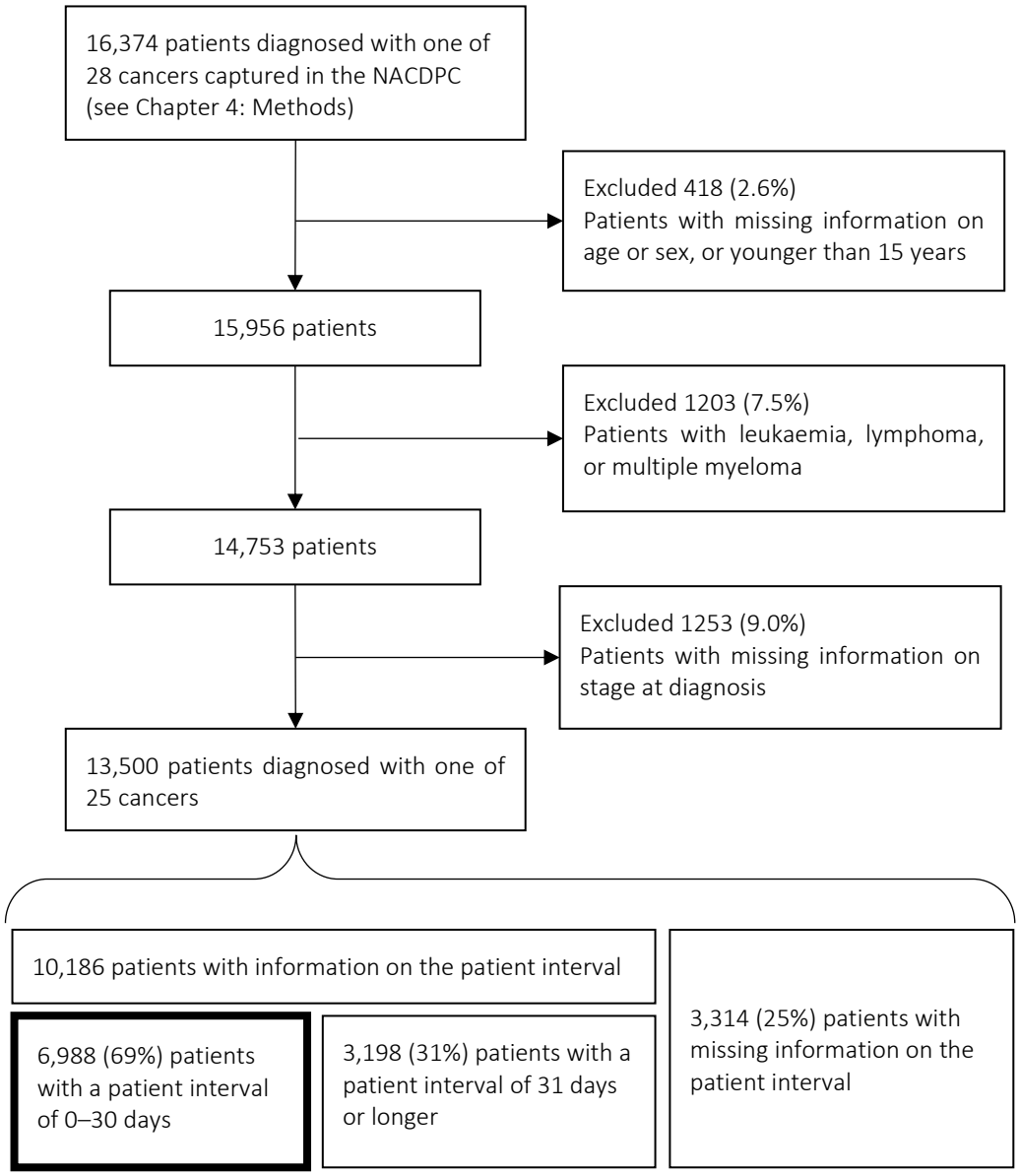


Figure 9.1 Flow diagram of sample derivation for the study population of this chapter

In the absence of information on the symptom onset-to-diagnosis interval (given the lack of information on diagnosis date in the audit dataset), I chose to restrict this analysis to patients with a patient interval of up to 30 days. Doing so additionally minimises the risk of potential confounding of the association of interest (between symptoms and stage) by time (Crawford, 2002; Tørring *et al*, 2011, 2017; Forrest *et al*, 2014; Neal *et al*, 2015)[1].

[1] In post-hoc analysis, I found that a lower proportion of patients with a patient interval of 0–30 had late stage compared to patients with a patient interval of 31+ days in the majority of cancer sites, supporting such confounding.

As described previously (see section 7.3.3 and Appendix 7.2), information on the patient interval was available for 75% of patients: those with missing patient interval values were more likely to be older males, and to have presented with a non-alarm symptom in a healthcare facility other than their general practice.

9.3.4 Variables of interest

The exposure of prior interest was alarm symptom at presentation. While the concept of 'alarm' symptom for suspected cancer is ubiquitous in early diagnosis literature, there is no universally accepted consensus regarding its definition. Therefore, alarm symptoms were selected for this analysis on the basis of satisfying at least four of the following five conditions, which aimed to provide an operational definition:

a) Mentioned in at least one English "Be Clear on Cancer" public health awareness campaign as a primary or secondary symptom (Public Health England, 2016b);
b) Association with a relatively high predictive value for cancer, as reported in the literature (based on papers identified in Chapter 3: Literature review);
c) Inclusion in NICE guidelines (2005) with recommendations mandating two-week-wait referral or urgent investigation for suspected cancer;
d) Inclusion in updated NICE guidelines (2015) with recommendations mandating two-week-wait referral or urgent investigation for suspected cancer; and
e) Sufficiently high frequency among the NACDPC study population to support analysis.

These criteria represent a pragmatic decision, designed to capture the greatest number of symptoms which would be considered as 'alarm' symptoms by clinicians, researchers, and policy-makers (a-d), while ensuring that they could be examined with sufficient precision using the NACDPC data (e). Out of an initial 22 symptoms that were considered, ten were selected: abnormal mole, breast lump, rectal bleeding, haematuria, haemoptysis, post-menopausal bleeding, change in bowel habit, dysphagia, jaundice, and weight loss (see Table 9.1, and Appendix 9.1 for excluded symptoms). All other symptoms were considered as 'non-alarm' and formed the comparison group.

There is some evidence to support an association between ethnicity and stage at diagnosis, as differences in symptom knowledge, appraisal, and recognition have also been described (Ward *et al*, 2004; Halpern *et al*, 2008; Niksic *et al*, 2016; Vrinten *et al*, 2016). However, there was no evidence to support a difference in stage by ethnicity in crude analyses in the study population (unadjusted OR: 0.8 (0.6–1.1) p=0.117), and so ethnicity was not included in the main analysis.

Table 9.1 Description of how the ten symptoms defined as 'alarm symptoms' in this analysis fit the selection criteria; see Appendix 9.1 for excluded symptoms.

Alarm symptom	Criterion a) Included in BCOC campaign[1]	Criterion b) Positive predictive value (PPV) reported in the literature[2]	Criterion c) Mentioned in NICE 2005 and associated with mandated action[3]	Criterion d) Mentioned in NICE 2015 and associated with mandated action[3]	Criterion e) NACDPC patients reporting the symptom (n)[4]
Abnormal mole	Primary symptom (skin cancer campaign)	None reported in available literature	Refer urgently for suspected melanoma	Refer urgently for suspected melanoma	525
Breast lump	Secondary symptom (breast cancer campaign)[5]	4.9% for breast cancer (Walker *et al*, 2014) 8.1% for breast cancer (Eberl *et al*, 2008)	Refer urgently for suspected breast cancer	Refer urgently for suspected breast cancer	2195
Change in bowel habit	Secondary symptom (ovarian cancer campaign)	2.4% for CRC (Lawrenson *et al*, 2006) 2.4% for CRC (Hamilton *et al*, 2005b)	Refer urgently for suspected CRC cancer	Refer urgently for suspected CRC	1010
Dysphagia	Secondary symptom (OG cancer campaign)	4.8% for OG cancer (Stapley *et al*, 2013) 6.0% for oesophageal cancer (Jones *et al*, 2007)	Refer urgently for endoscopy for suspected OG cancer	Refer urgently for endoscopy for suspected OG cancer	418
Haematuria	Primary symptom (bladder & kidney cancer campaign)	2.6% for bladder cancer (Shephard *et al*, 2012) 5.9% for a urinary tract cancer[6] (Jones *et al*, 2007) 6.5% for a cancer of the urological tract[7] (Bruyninckx *et al*, 2003)	Refer urgently for suspected bladder or renal cancer	Refer urgently for suspected bladder or renal cancer	914
Haemoptysis	Secondary symptom (lung cancer campaign)	2.4% for lung cancer (Hamilton *et al*, 2005a) 8.4% for a respiratory tract cancer (Jones *et al*, 2007) 6.4% for lung cancer (Hippisley-Cox & Coupland, 2011b)	Refer urgently for suspected lung cancer	Refer urgently for suspected lung cancer	222

Alarm symptom	Criterion a) Included in BCOC campaign[1]	Criterion b) Positive predictive value (PPV) reported in the literature[2]	Criterion c) Mentioned in NICE 2005 and associated with mandated action[3]	Criterion d) Mentioned in NICE 2015 and associated with mandated action[3]	Criterion e) NACDPC patients reporting the symptom (n)[4]
Jaundice	Not included	21.6% for pancreatic cancer (Stapley *et al*, 2012)	Refer urgently for suspected OG cancer	Refer urgently for suspected pancreatic cancer	162
Rectal bleeding	Primary symptom (CRC campaign)	2.4% for CRC (Shapley et al, 2010)[8]	Refer urgently for suspected CRC	Refer urgently for suspected CRC	768
Post-menopausal bleeding	Not included	5.4% for endometrial cancer (Parker et al, 2007)	Refer urgently for suspected gynaecological cancer	Refer urgently for suspected endometrial cancer	322
Weight loss	Secondary symptom (bladder & kidney cancers, CRC, lung, OG, and ovarian cancer campaigns)	1.2% for CRC cancer (Hamilton *et al*, 2005b) 1.1% for lung cancer (Hamilton *et al*, 2005a) 0.8% for OG cancer (Stapley *et al*, 2013) 0.8% for pancreatic cancer (Stapley *et al*, 2012) 0.2% for multiple myeloma (Shephard *et al*, 2015a)	Refer urgently for chest x-ray for suspected lung cancer Refer urgently for suspected OG cancer (with upper abdominal pain)	Refer urgently for suspected cancer (several types) With abdominal pain: refer urgently for suspected CRC Alone or with chest pain, cough, dyspnoea, appetite loss, fatigue: offer urgent chest x-ray for suspected lung cancer or mesothelioma With upper abdominal pain, reflux, or dyspepsia: offer urgent direct access endoscopy for suspected OG cancer	751

BCOC: Be Clear on Cancer; CRC: colorectal cancer; OG: oesophago-gastric cancer

1 The 'primary' or 'secondary' status of symptoms was inferred based on the design and phrasing of campaign materials: symptoms that were used to headline individual campaigns or that were described as 'key' were considered to be primary symptoms, while other symptoms mentioned in supporting material were considered to be secondary symptoms

2 where individual PPVs were presented by age group or sex, the lowest value has been reported

3 where phraseology indicated mandated action, usually two-week-wait referral ("refer") or urgent investigation ("offer"). Excludes symptoms for which guidance begins "consider"

4 among 15,956 cancer patients in the NACDPC

5 the primary focus of this campaign was to raise awareness of breast cancer among 70+ year old women rather than on raising awareness of particular presenting symptoms.

6 including urethra, bladder, ureter, and kidney cancers

7 including all cancers of the urological tract

8 smallest PPV of a range identified by Shapley et al (2010) as part of a systematic review, based on 13 studies on rectal bleeding and colorectal cancer

9.3.5 Statistical analyses

Firstly, the frequency of alarm symptoms among the study population was estimated, together with accompanying exact confidence intervals.

Subsequently, the association between individual alarm symptoms and stage at diagnosis was examined using logistic regression (i.e. comparing odds of late versus early stage disease). Patients were first categorised according to whether they had each of the ten alarm symptoms at presentation, or none; those with multiple alarm symptoms were excluded from the regression analyses (n=131, 4% of all patients recorded as presenting with an alarm symptom).

Crude logistic regression models were run before running multivariate analyses, using patients with non-alarm symptoms as the reference group. The first multivariate model included two covariates, age group (parameterised as 15–49 years, 50–69 years, and 70+ years), and sex (parameterised as men, women). Variation in odds for late stage across alarm symptoms and age group was examined using joint Wald tests, with statistical significance at the 5% level. This model examined the association between alarm symptom category and stage (adjusted for age and sex) among incident cancer patients with any potential tumour (other than haematological, see section 9.3.1).

A second multivariate model including cancer site as an additional covariate was conducted; this model estimated the association between each of the alarm symptoms and stage after taking into account potential confounding by cancer site. For ease of interpretation, cancer site was parameterised as a categorical variable comprising the 15 most common cancer sites in the study population as distinct categories, together with a single 'other cancer site' category for all other cancer sites excluding haematological cancers.

9.3.6 Supplementary analyses

9.3.6.1 Restricting analyses to patients who presented and were referred within 30 days

In order to further control for the potential confounding effect of diagnostic timeliness on the association between presenting symptoms and stage, I repeated the analyses on patients who had a combined patient and primary care interval of 0–30 days.

9.3.6.2 Assuming patients with missing stage had late stage

In order to examine the potential of stage information being missing not at random, I repeated the above analyses assuming all patients with missing stage information had distant stage at diagnosis. This 'worst case scenario' sensitivity analysis has been shown to be useful for cases where there only a few missing values of a binary outcome (Sterne *et al*, 2009).

9.4 Results

9.4.1 The prevalence of alarm symptoms among patients who presented within 30 days of symptom onset

A total of 3,496/6,988 (50%) cancer patients who presented within 30 days of symptom onset had one or more of the ten studied alarm symptoms (see Table 9.2). Breast lump (18%) was the most common presenting alarm symptom in the analysis cohort, followed by haematuria (9%) and change in bowel habit (6%), while jaundice was reported as a presenting symptom in 110 patients (2%).

Table 9.2 Frequency of alarm symptoms among the studied cancer patient population (n=6,988)

Alarm symptom	N	Proportion (95% CI)
Breast lump	1282	18% (17–19%)
Haematuria	620	9% (8–10%)
Change in bowel habit	421	6% (5–7%)
Rectal bleeding	333	5% (4–5%)
Weight loss	240	3% (3–4%)
Post-menopausal bleeding	171	2% (2–3%)
Abnormal mole	164	2% (2–3%)
Dysphagia	161	2% (2–3%)
Haemoptysis	128	2% (2–2%)
Jaundice	108	2% (1–2%)
Any alarm symptom	3496	50% (49–51%)
Non-alarm symptoms[1]	3492	50% (49–51%)
Total number of patients	6988	100%

1 Any symptom that was not considered an alarm symptom

9.4.2 Associations between alarm symptoms and stage at diagnosis

Among patients who had presented within 30 days of symptom onset with non-alarm symptoms, over a quarter (27%) were recorded as having distant stage at diagnosis (see Table 9.3). In comparison, the proportion of distant stage at diagnosis among patients who had presented with an alarm symptom within 30 days of onset ranged from 2% (abnormal mole) to 45% (weight loss).

Crude (unadjusted) logistic regression analyses indicated evidence that five symptoms (abnormal mole, post-menopausal bleeding, breast lump, haematuria, and rectal bleeding) were associated with substantially lower odds of late stage at diagnosis compared to patients with non-alarm symptoms (crude OR values <0.6 for all five, with upper bound of 95% CIs being <1.0). In comparison, there was evidence that weight loss was associated with greater odds of late stage (crude OR (95% CI): 2.1 (1.6–2.9), p<0.001). For patients who presented with one of four alarm symptoms (dysphagia, haemoptysis, change in bowel habit, and jaundice), there was no significant difference in odds of advanced stage disease compared to patients with non-alarm symptoms.

The pattern of variation in odds of late stage disease by alarm symptom did not change substantially after adjustment for age group and sex (joint Wald test p<0.001). The five alarm symptoms described above remained associated with lower odds of late stage at diagnosis: abnormal mole was associated with lowest adjusted OR (95% CI) of 0.07 (0.02–0.2), while weight loss remained associated with greater odds of distant stage (adjusted OR (95% CI): 2.2 (1.6–3.0), p<0.001) (see Figure 9.2 and Table 9.3).

The multivariate analysis also indicated strong evidence for an association between age group and stage at diagnosis: patients aged 15–49 years were less likely to be diagnosed with late stage disease than those aged 50–69 years (adjusted OR (95% CI): 0.6 (0.5–0.8), p<0.001). Men had been more likely to be diagnosed with late stage disease compared to women in univariate analyses, but adjusting for symptoms and age, women were more likely to be diagnosed with late stage disease (see Table 9.3).

Table 9.3 Proportion of late (distant) stage, and crude/adjusted odds ratios of late stage at diagnosis among patients who presented within 30 days of symptom onset (n=6,857[1])

Variable	N	% distant stage	Crude OR[1] (95% CI)	P-value[2]	Adjusted OR[1] (95% CI)	P-value[2]
Alarm symptom						
No alarm symptoms	3492	27% (25–28%)	Ref.		Ref.	
Abnormal mole	164	2% (1–6%)	**0.1 (0.03–0.2)**		**0.1 (0.03–0.2)**	
PMB	169	5% (3–10%)	**0.2 (0.1–0.3)**		**0.1 (0.1–0.3)**	
Breast lump	1281	6% (4–7%)	**0.2 (0.1–0.2)**		**0.2 (0.1–0.2)**	
Haematuria	618	8% (6–11%)	**0.3 (0.2–0.3)**		**0.3 (0.2–0.3)**	
Rectal bleeding	267	15% (11–20%)	**0.5 (0.3–0.7)**	**<0.001**	**0.5 (0.3–0.6)**	**<0.001**
Dysphagia	143	22% (16–30%)	0.8 (0.5–1.2)		0.8 (0.5–1.2)	
Haemoptysis	124	28% (21–37%)	1.1 (0.7–1.6)		1.1 (0.7–1.6)	
CIBH	330	28% (24–33%)	1.1 (0.8–1.4)		1.0 (0.8–1.3)	
Jaundice	96	35% (27–45%)	1.5 (1.0–2.3)		1.4 (0.9–2.2)	
Weight loss	173	45% (38–53%)	**2.2 (1.6–3.1)**		**2.2 (1.6–3.0)**	
Age group						
15–49 years	982	11% (9–13%)	**0.5 (0.4–0.6)**		**0.6 (0.5–0.8)**	
50–69 years	2660	21% (19–22%)	Ref.	**<0.001**	Ref.	**<0.001**
70+ years	3346	23% (21–24%)	1.1 (1.0–1.3)		1.1 (0.9–1.2)	
Sex						
Male	3284	22% (21–24%)	Ref.		Ref.	
Female	3704	19% (17–20%)	**0.8 (0.7–0.9)**	**<0.001**	**1.3 (1.1–1.5)**	**<0.001**

CIBH: change in bowel habit; PMB: post-menopausal bleeding
1excludes 131 patients with multiple alarm symptoms
2Joint Wald test p-value

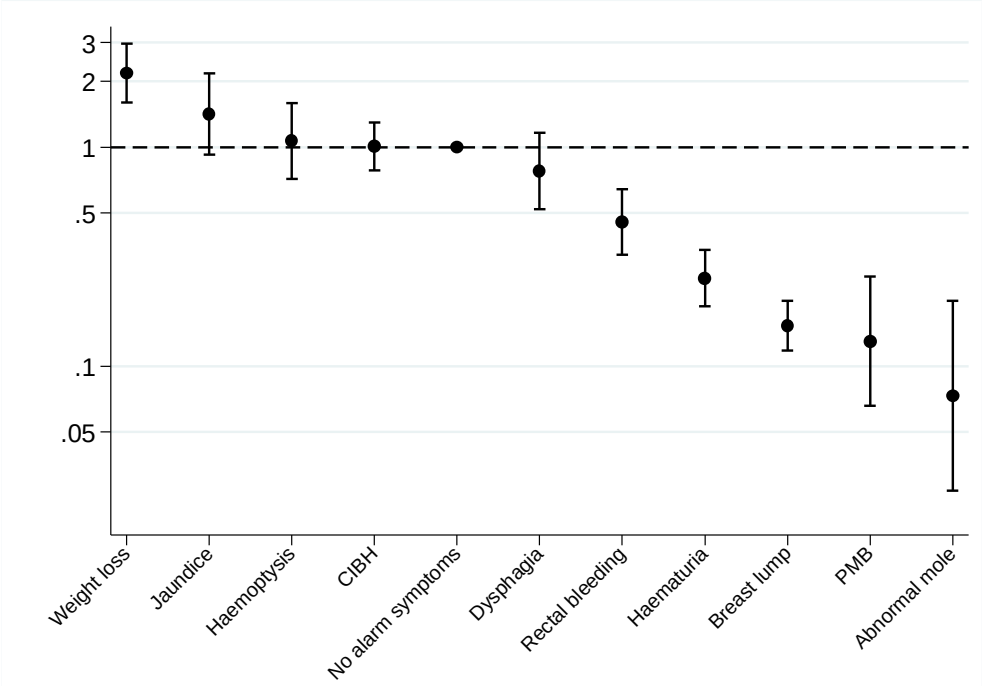


Figure 9.2 Adjusted odds ratios for late stage disease versus early stage disease by symptom
Adjusted for age group and sex; Y-axis is on a log scale to aid visualisation.
CIBH: change in bowel habit; PMB: post-menopausal bleeding

9.4.3 Adjusting associations between alarm symptoms and stage at diagnosis by cancer site

In the model that further adjusted for cancer site, there were a few differences to the output of the main analysis (see Table 9.4). Haematuria was only weakly associated with earlier stage at diagnosis after adjusting for cancer site (adjusted OR (95% CI): 0.7 (0.5–0.1), p=0.055). Further, after adjusting for cancer site, haemoptysis was associated with lower odds of late stage disease compared to the main model which indicated no evidence for an association (adjusted OR (95% CI): 0.5 (0.3–0.8), p=0.001).

There was no evidence for an association between age group or sex with stage at diagnosis adjusting for cancer site, indicating the age and sex variation observed in the first model (which only adjusted for these two variables and symptom category) could be accounted for by cancer site.

Fewer than 10% of cancer patients diagnosed with oropharyngeal, testicular, bladder, endometrial, melanoma and breast cancer had advanced stage disease, whereas over two-fifths of patients diagnosed with ovarian, lung, or pancreatic cancer were diagnosed with late stage disease. This substantial variation in proportion of late stage disease by cancer site remained after adjusting for alarm symptom category, age group, and sex: adjusted ORs (95% CI) ranged from 0.1 (0.02–0.3) among patients diagnosed with oropharyngeal cancer and 2.2 (1.6–3.1) for patients with ovarian cancer compared to colorectal cancer (joint Wald test p<0.001).

Table 9.4 Proportion of late (distant) stage, and crude/adjusted odds ratios of late stage at diagnosis among patients who presented within 30 days of symptom onset, including cancer site as an additional covariate (n=6,857[1])

Variable	N	% distant stage	Crude OR[1] (95% CI)	P-value[2]	Adjusted OR[1] (95% CI)	P-value[2]
Alarm symptom						
No alarm symptoms	3492	27% (25–28%)	Ref.		Ref.	
Abnormal mole	164	2% (1–6%)	**0.1 (0.03–0.2)**		**0.1 (0.04–0.4)**	
PMB	169	5% (3–10%)	**0.2 (0.1–0.3)**		**0.3 (0.1–0.6)**	
Breast lump	1281	6% (4–7%)	**0.2 (0.1–0.2)**		**0.2 (0.1–0.2)**	
Haematuria	618	8% (6–11%)	**0.3 (0.2–0.3)**		0.7 (0.5–1.0)	
Rectal bleeding	267	15% (11–20%)	**0.5 (0.3–0.7)**	**<0.001**	**0.4 (0.3–0.6)**	**<0.001**
Dysphagia	143	22% (16–30%)	0.8 (0.5–1.2)		0.6 (0.4–1.0)	
Haemoptysis	124	28% (21–37%)	1.1 (0.7–1.6)		**0.5 (0.3–0.8)**	
CIBH	330	28% (24–33%)	1.1 (0.8–1.4)		0.9 (0.7–1.2)	
Jaundice	96	35% (27–45%)	1.5 (1.0–2.3)		0.8 (0.5–1.4)	
Weight loss	173	45% (38–53%)	**2.2 (1.6–3.1)**		**1.7 (1.2–2.3)**	
Age group						
15–49 years	982	11% (9–13%)	**0.5 (0.4–0.6)**		0.8 (0.6–1.0)	
50–69 years	2660	21% (19–22%)	Ref.	**<0.001**	Ref.	0.054
70+ years	3346	23% (21–24%)	1.1 (1.0–1.3)		1.0 (0.9–1.2)	
Sex						
Male	3284	22% (21–24%)	Ref.		Ref.	
Female	3704	19% (17–20%)	**0.8 (0.7–0.9)**	**<0.001**	1.0 (0.9–1.2)	0.926
Cancer diagnosis[3]						
Oropharyngeal	81	2% (1–9%)	**0.1 (0.02–0.3)**		**0.1 (0.02–0.3)**	
Testicular	77	3% (1–9%)	**0.1 (0.02–0.3)**		**0.1 (0.02–0.3)**	
Bladder	504	4% (3–7%)	**0.1 (0.1–0.2)**		**0.1 (0.1–0.3)**	
Brain	125	5% (2–10%)	**0.1 (0.1–0.3)**		**0.1 (0.1–0.3)**	
Endometrial	198	6% (3–10%)	**0.2 (0.1–0.3)**		**0.3 (0.1–0.5)**	
Melanoma	245	7% (4–11%)	**0.2 (0.1–0.4)**		**0.5 (0.3–0.9)**	
Breast	1506	9% (8–10%)	**0.3 (0.2–0.4)**		0.9 (0.6–1.3)	
Prostate	858	18% (15–20%)	**0.6 (0.5–0.8)**		**0.5 (0.4–0.7)**	
Other cancer[1]	545	19% (16–22%)	**0.7 (0.5–0.9)**		**0.6 (0.5–0.8)**	
Colorectal	1033	25% (22–28%)	Ref.	**<0.001**	Ref.	**<0.001**
Stomach	133	29% (22–37%)	1.2 (0.8–1.8)		1.0 (0.6–1.5)	
Oesophageal	244	29% (24–35%)	1.2 (0.9–1.7)		1.2 (0.8–1.9)	
Renal	182	33% (27–40%)	**1.4 (1.0–2.0)**		1.3 (0.9–2.0)	
Lung	874	43% (40–46%)	**2.2 (1.8–2.7)**		**1.9 (1.5–2.4)**	
Pancreatic	199	44% (37–51%)	**2.4 (1.7–3.4)**		**2.1 (1.4–3.0)**	
Ovarian	184	45% (38–52%)	**2.4 (1.8–3.4)**		**2.2 (1.6–3.1)**	

CIBH: change in bowel habit; PMB post-menopausal bleeding
1 Excludes 131 patients with multiple alarm symptoms
2 Joint Wald test p-value
3 Includes 12 specified cancer sites (brain, gallbladder, laryngeal, liver, lymphoma, mesothelioma, sarcoma, small intestine, thyroid, and vulval cancer) and unspecified 'other' cancers

9.4.4 Supplementary results: restricting analyses to patients who presented and were referred within 30 days

Crude and adjusted logistic regression analyses were conducted on patients who had a total pre-referral interval (sum of the patient interval and primary care interval) of 0–30 days (see Appendix 9.2). Given the more restrictive definition, this analysis was run on a smaller sample of patients compared to the main analysis (n=4,909, 70% of all patients who presented within 30 days). The resulting odds ratios were largely comparable: abnormal mole, post-menopausal bleeding, breast lump, haematuria, and rectal bleeding remained associated with lower odds of late stage disease compared to patients with non-alarm patients, while those with weight loss were more likely to be diagnosed at late stage (adjusted OR (95% CI): 2.7 (1.8–4.1)). There was weak evidence to support that jaundice was associated with late stage disease (adjusted OR (95% CI): 1.7 (1.0–2.7)) (see Appendix 9.2).

9.4.5 Supplementary results: assuming patients with missing stage had late stage

I repeated the main analyses whereby patients with missing information on stage at diagnosis were assigned to late stage (n=7,467). The resulting multivariate logistic regression model indicated negligible differences to the previously observed odds ratios by alarm symptom category (see Appendix 9.3).

As observed in the main analysis, there was evidence that cancer patients who presented with one of five alarm symptoms (abnormal mole, post-menopausal bleeding, breast lump, haematuria, and rectal bleeding) had lower odds of being diagnosed at late stage compared with patients who presented without alarm symptoms (OR values <0.7 for all, with upper bound 95% CIs <1.0). Again, as observed in the main analysis, there was evidence that cancer patients who presented with weight loss within 30 days of symptom onset were more than twice as likely to be diagnosed with late stage disease compared to cancer patients who presented with non-alarm symptoms (adjusted OR (95% CI): 2.2 (1.6–2.9)) (see Appendix 9.3). There was no evidence to suggest that patients who presented with one of four symptoms (dysphagia, change in bowel habit, haemoptysis, and jaundice) were more likely to be diagnosed with late stage disease compared to those who presented with non-alarm symptoms.

9.5 Discussion

9.5.1 Main findings

Almost half (50%) of symptomatic cancer patients who sought help within a month from symptom onset had one or more of the studied alarm symptoms at presentation before diagnosis. The proportion of patients diagnosed with late stage disease among this group varied greatly by the studied alarm symptoms, following three distinct patterns.

Firstly, cancer patients with one of five alarm symptoms (abnormal mole, PMB, breast lump, haematuria, and rectal bleeding) had significantly lower odds of distant disease after adjusting

for the effect of age group and sex compared to cancer patients who presented with other non-alarm symptoms. Secondly, cancer patients who presented with dysphagia, change in bowel habit, haemoptysis, or jaundice had odds of advanced stage disease that were similar to patients with non-alarm symptoms. Lastly, patients who presented with weight loss were significantly more likely to have late stage disease.

9.5.2 Comparison with prior evidence

This study is the first to my knowledge to describe associations between presenting symptoms of cancer and stage at diagnosis across a large incident cohort of cancer patients diagnosed with one of 25 different cancer sites.

Evidence about associations between symptoms and stage at diagnosis is limited. A small number of studies point to differences in symptomatic profiles of cancer patients diagnosed with early versus late stage disease: for example, women with ovarian cancer who presented with abnormal vaginal bleeding have been shown to be more likely to have localised disease than regional/distant disease in women with ovarian cancer without that symptom (Ryerson *et al*, 2007). Similarly, among colorectal cancer patients, those presenting with rectal bleeding were more likely to be diagnosed with early stage disease compared with patients with other presenting symptoms (Alexiusdottir *et al*, 2012). Generally, evidence tends to be restricted to only a few cancer sites, and examines presence versus absence of symptoms, without consideration of the relative differences between patients with different presenting symptoms (Goff *et al*, 2000; Ryerson *et al*, 2007; Lurie *et al*, 2010; Alexiusdottir *et al*, 2012).

Cancer-site specific evidence such as the above can only provide a partial picture of the association between presenting symptoms and stage at diagnosis in the entire incident cohort of cancer patients, given that not all cancer patients presenting with an alarm symptom are subsequently diagnosed with the associated cancer site (as discussed in Chapter 8).

9.5.3 Strengths and limitations

In this study, I focused on ten common 'alarm' symptoms (Table 9.2). This was based on the consideration of five factors, which incorporated epidemiological/clinical factors (known predictive value of a symptom for cancer), relevance to current policy and practice in England (inclusion in Be Clear on Cancer campaigns and mention in NICE clinical guidelines), and pragmatic factors such as statistical precision (available sample size among the NACDPC study population). However, the application of this research question to other common presenting symptoms of cancer also warrants further examination, particularly as the majority of cancer patients present with non-specific or vague symptoms as discussed in Chapter 3 (Literature review). Furthermore, power limitations may have precluded differences between early and late stage disease among patients presenting with rarer alarm symptoms (such as haemoptysis and jaundice) from being detected.

Relatedly, the reference group comprised cancer patients who presented with symptoms other than the 'alarm' symptoms as defined in this analysis (see Section 9.3.4 and Appendix 9.1). Misclassification of a symptom as non-alarm (when it should have been considered as an alarm symptom) could have attenuated the association with stage at diagnosis with alarm symptom status ('alarm' versus 'non-alarm'). In spite of this possibility, the findings indicate substantial relative variation in association with late stage disease between the ten alarm symptoms, independent of the reference group.

The findings are reliant upon the validity and completeness of staging information that was available. At the time of data collection (2009–10), population-based data on the stage at diagnosis of incident cancer cases in England was highly incomplete, particularly for rarer cancer sites. Therefore GPs who took part in the NACDPC provided information on stage at diagnosis using the SEER LRD system based on practice records, including hospital correspondence. The principal categorisation used by the SEER LRD system is a widely used method of staging; indeed stage information from the NACDPC dataset was used as part of a recent international comparison of diagnostic timeliness and stage among colorectal cancer patients (Tørring *et al*, 2017).

However, the interpretation of local, regional, and distant disease can vary substantially across cancer site (Young *et al*, 2001). Moreover, GPs had to assign stage based on information in practice records and hospital correspondence, whereas information on stage included in population-based cancer registries typically includes information from a wide range of clinical, imaging, laboratory, and pathology reports. Extending the enquiry of associations between presenting symptoms and stage at diagnosis in more recent incident cohorts with more complete stage at diagnosis information derived by cancer registration protocols should therefore be prioritised. Nevertheless, analyses assuming that all patients with missing information on stage at diagnosis had late stage found the associations between alarm symptoms and stage largely unchanged (see Appendix 9.3), while retrospective comparison of the proportion of cancer patients with late stage and missing stage between the NACDPC patients and a more recent incident cohort (2015) for several cancer sites lends support to the SEER LRD system (data not shown) (NCRAS and ONS, 2017).

In order to ensure findings were relevant to early diagnosis interventions promoting timely help-seeking (typically within several weeks of symptom onset), the analysis was restricted to patients who presented within 30 days of symptom onset. This restriction excluded patients for whom disease progression (as measured by stage at diagnosis) may have been influenced by prolonged time to presentation, minimising the confounding effect of the "waiting time paradox" on findings. Further, the risk of length time bias (a form of selection bias which could

have led to the over-representation of slow growing tumours in the study population) was minimised through this restriction.

The above restriction was justified given the aims of this study, but future work could also aim characterise the association between presenting symptoms and stage at diagnosis among patients with varying diagnostic intervals to further address the potential effect of the "sicker quicker" paradox for a range of cancer sites (Crawford, 2002; Tørring *et al*, 2011, 2017; Forrest *et al*, 2014; Neal *et al*, 2015).

9.5.4 Implications

Symptoms are, by definition, considered to be the first physical manifestations of disease. Being able to examine how a symptom (at onset, or soon after onset) relates to the extent of disease is therefore critical for guiding strategies to achieve earlier diagnosis of cancer. This study aimed to examine whether the studied alarm symptoms could be distinguished from non-alarm symptoms on the basis of their association with stage at diagnosis, and also to examine variation in associations between different alarm symptoms and stage.

The findings indicate that the studied alarm symptoms have highly heterogeneous associations with extent of disease. Among patients who had presented within 30 days of symptom onset, compared to patients who presented with non-alarm symptoms, those who had presented with one of five alarm symptoms (abnormal mole, PMB, breast lump, rectal bleeding, and haematuria) were less likely, while those who had presented with weight loss were more likely, to be diagnosed with distant disease. Therefore, this suggests that the studied alarm symptoms should not be treated as a single construct.

Some of the studied alarm symptoms, such as dysphagia, haemoptysis, jaundice, and weight loss are often considered a sign of late-stage disease (Stapley *et al*, 2012; Shim *et al*, 2014; Walter *et al*, 2015; Fang *et al*, 2016; Ewing *et al*, 2018). While I found they were more strongly associated with advanced stage compared with other alarm symptoms, a large group of patients who presented promptly (within 30 days) with one of these symptoms were diagnosed with local or regional disease. These findings do not support excluding these symptoms from early diagnosis interventions.

In comparison, the majority of patients who had presented with one of the five alarm symptoms mentioned above (abnormal mole, PMB, breast lump, rectal bleeding, and haematuria) were diagnosed with early stage disease, with only a small proportion (less than 15%) of patients were diagnosed with late stage disease. These findings support the continued inclusion of these symptoms in public health awareness campaigns, and healthcare interventions to promote earlier diagnosis in such patients.

9.6 Chapter summary

There is variation in stage at diagnosis associated with different alarm symptoms among promptly presenting cancer patients. The findings indicate a variable association between alarm symptoms and disease stage, and that some alarm symptoms often considered to be associated with late stage disease may reflect local or regional disease in certain patients.

This study highlights the value of using epidemiological approaches to examine the theoretical (often implicit) assumptions that underpin early diagnosis interventions, to help strengthen the rationale and evidence base.

Chapter 10: Discussion

This chapter summarises the main findings of this thesis and reflects on its limitations. Finally, the implications of the conceptual contributions and empirical findings of this thesis for public health, health policy, and future research are discussed.

10 Discussion

This thesis aimed to describe the epidemiology of symptoms experienced by cancer patients before diagnosis, and how they relate to diagnostic timeliness. The majority of cancer patients present with symptoms, and these symptoms often serve as the starting point of their diagnostic pathway. Understanding the frequency and nature of symptoms, and how they may be associated with time to help-seeking or referral from primary care could be insightful for public health practice and research.

In this final chapter, I summarise the findings of preceding chapters and reflect upon their contribution to the main aims of this thesis. Further, I address important assumptions and limitations of the empirical findings, before discussing implications for public health practice and research.

10.1 The epidemiology of cancer symptoms and associated diagnostic timeliness

The literature review (presented in Chapter 3) helped crystallise my understanding of the challenges in examining the presenting symptoms of patients subsequently diagnosed with cancer in patient populations in the context of epidemiological research. Information on symptoms experienced by cancer patients before diagnosis can be obtained either from patients directly through self-report (Burgess *et al*, 2006; Evans *et al*, 2014; Forbes *et al*, 2014; Lim *et al*, 2014; Walter *et al*, 2014; Howell *et al*, 2015; McLachlan *et al*, 2015; Queenan *et al*, 2017) or their health records (Hippisley-Cox *et al*, 2004; Hamilton, 2009b; Blak *et al*, 2011; Herrett *et al*, 2015). Both approaches have distinct strengths and weaknesses with implications for analysis and interpretation.

The review highlighted evidential uncertainties and gaps, in particular several cancer sites were bereft of epidemiological evidence on presenting symptoms, and evidence characterising associations between presenting symptoms and intervals to diagnosis was scarce. Nevertheless, the limited cancer site-specific evidence pointed to marked variation in diagnostic timeliness between patients with different symptoms (Pruitt *et al*, 2013; Walter *et al*, 2015, 2016a, 2016b; Leiva *et al*, 2017).

These findings supported the initial hypothesis that motivated this PhD and demonstrated the need for further detailed examination of the presenting symptoms of cancer patients and associated diagnostic intervals in representative incident cohorts of patients with a range of cancer sites.

10.2 Cataloguing the presenting symptoms (and symptom status) of cancer patients

The empirical outputs of my PhD are based on the secondary analysis of cross-sectional data collected as part of a national audit of cancer diagnosis in primary care in England (NACDPC). I undertook the coding of free-text information on the presenting symptoms of cancer patients in the NACDPC dataset. As summarised in Chapter 4, this was an iterative process that involved the identification of different phenotypic expressions of the same symptom entity (construct), and the development of a rules-based algorithm for further systematic coding of symptom information.

Although my original aim was to examine the epidemiology of presenting symptoms, in doing so, I identified a considerable number of patients who had either been asymptomatic, or had not presented with symptoms associated with the subsequently identified cancer before diagnosis. I deemed it important to investigate the characteristics of this population given the dearth of epidemiological evidence regarding such patients and the circumstances surrounding their diagnosis (Chapter 5).

Examination of this group identified substantial variation in the probability of an atypical diagnosis by cancer site: over a third of leukaemia patients were diagnosed incidentally or asymptomatically, and liver, renal, and myeloma cancer patients were also relatively more likely to have had an atypical diagnosis. Exploration of the circumstances preceding cancer diagnosis of these patients (where further information was available) identified several common clinical scenarios leading to an atypical diagnosis of cancer.

10.3 The symptom signature of cancer, the cancer signature of symptoms, and diagnostic timeliness

I aimed to explore the symptom signatures, cancer signatures, and associated diagnostic timeliness among the NACDPC patient population. These concepts can be applied to a range of symptoms and cancers; within the limited time and resources available to completing my doctoral studies, I restricted the application of these inquiries to specific cancers and symptoms to serve as exemplar studies.

10.3.1 The symptom signature of breast cancer

In Chapter 6 I examined the spectrum of presenting symptoms among women subsequently diagnosed with breast cancer, and the patient and primary care intervals associated with different symptoms. Breast cancer was chosen as the focus of this study as although it is largely characterised by a 'narrow symptom signature' (as shown in Chapter 3), previous evidence regarding intervals to diagnosis and how they may be influenced by presenting symptoms

comprised studies from mostly small secondary care populations, or qualitative examinations (Ramirez *et al*, 1999; Friedman *et al*, 2006; Webber *et al*, 2017). Further, while most patients diagnosed with breast cancer have good prognosis relative to other cancer types, it is associated with high morbidity and mortality due to its high incidence and remains a meaningful target for further interventions (Walters *et al*, 2013; Kaushal *et al*, 2016; Cancer Research UK, 2017).

I identified a large number of presenting symptoms other than breast lump, encompassing a total of 56 symptoms in 95 unique combinations. This is a much richer symptomatic picture of breast cancer including rarer symptoms and symptom combinations in contrast to previous studies as identified in Chapter 3, where selected symptoms (determined *a priori*) or larger symptom categories had been described. Examination of diagnostic timeliness indicated variation in time to presentation by symptom type: women with non-lump breast symptoms had longer patient intervals than women with breast lump alone. Further, by examining symptom combinations (rather than presence/absence of a symptom in isolation), I was able to elucidate that women presenting with both breast lump and non-lump breast symptoms presented later than those with breast lump alone.

10.3.2 The cancer signature of abdominal symptoms and diagnostic timeliness

Presenting symptoms can be indicative of different malignancies; examining diagnostic timeliness based on symptoms rather than the subsequently diagnosed cancer can elucidate important insights (Dobson *et al*, 2014). This is perhaps most effectively exemplified by abdominal symptoms, which are of particular interest to policymakers in early diagnosis. In England, this is demonstrated by the recent regional pilot Be Clear on Cancer campaign for 'abdominal symptoms' (Public Health England, 2017b), and multi-disciplinary diagnostic centre initiatives aiming to expedite diagnostic resolution in patients with abdominal symptoms (Fuller *et al*, 2016; London Cancer, 2017). Therefore, focusing on cancer patients with abdominal symptoms, I examined the spectrum of cancer sites diagnosed following a particular symptomatic presentation (the 'cancer signature'), and variation in diagnostic timeliness by symptom in Chapters 7 and 8.

Incident cancer patients who presented with one or more abdominal symptoms were subsequently diagnosed with a spectrum of abdominal, non-abdominal, and haematological cancers. Examination of symptom-specific cancer signatures indicated that a variable proportion of patients who presented with an abdominal alarm symptom were not diagnosed with the most commonly associated cancer site. Non-alarm abdominal symptoms had much more variable subsequent cancer diagnoses.

Time to presentation and referral (and their relative contributions to the combined pre-referral interval) varied significantly among patients with different abdominal symptoms. Abdominal pain was associated with relatively short patient intervals but longer primary care intervals. On the other hand, cancer patients who presented with dysphagia or change in bowel habit tended to wait almost a month before seeking help but were then promptly referred after presentation.

10.4 Presenting symptoms of cancer and stage at diagnosis

Alarm symptoms play a significant role in early diagnosis activities, but the degree by which these symptoms represent early stage cancer is poorly understood. I therefore sought to contribute to this evidence gap in Chapter 9, focusing my enquiry on prompt presenters (patients who had presented within 30 days of symptom onset).

The findings indicated that, beyond signifying a strong association with cancer, the 'alarm' symptom label did not always indicate early stage disease. Cancer patients who presented with abnormal mole, post-menopausal bleeding, breast lump, rectal bleeding, or haematuria were less likely to be diagnosed with late stage disease, compared to patients with non-alarm symptoms, while patients who presented with weight loss were more likely to be diagnosed with late stage disease. Patients who presented with change in bowel habit, dysphagia, haemoptysis, or jaundice had similar odds of late stage disease as those with non-alarm symptoms. This variation remained unchanged after adjusting for cancer site.

While the relative odds of late stage disease among alarm symptoms (compared to non-alarm symptoms) was variable, substantial proportions of patients with alarm symptoms typically thought to represent advanced stage disease (e.g. dysphagia, jaundice, and weight loss) were recorded as having local or regional disease.

10.5 Limitations of this thesis

The limitations of this thesis largely fall into one of two categories: characteristics of the data source, and methodological or statistical considerations.

10.5.1 Characteristics of the NACDPC dataset

The external validity (generalisability) of the audited study population and general practices that participated in the NACDPC has been examined previously (Rubin *et al*, 2011; Lyratzopoulos *et al*, 2013a). Nevertheless, there are several additional limitations to consider in the context of the original research presented in this thesis.

10.5.1.1 How presenting symptoms were measured

As discussed in Chapter 3, the validity and completeness of symptom information collected in primary care records is dependent on patients disclosing relevant symptoms during consultations before diagnosis, and on their accurate elicitation, interpretation, and recording

by doctors during (or soon after) the consultation (Hripcsak & Albers, 2013). Therefore, compared to other sources of symptom information such as patient surveys, the information on symptoms captured in the NACDPC may have underestimated the true symptom burden of patients diagnosed with cancer (Lim *et al*, 2016; Leiva *et al*, 2017).

Additionally, data extraction (by the GPs participating in the audit) was performed retrospectively, and in the knowledge of the patient's subsequent cancer diagnosis. Consequently, rarer symptoms or non-specific symptoms may have been under-recorded, while the opposite may be true for more typical symptoms and alarm symptoms.

On the other hand, the clinicians who participated in the audit were able to capture the most pertinent details of symptomatic presentation, and had the practical capability or doing so, given their knowledge of each cancer patient and direct access to the patient's primary care record. Auditors had full access to both coded and free-text data, which could have minimised the risk of bias associated with calculating symptom frequencies on coded data alone (Price *et al*, 2016). The anonymous nature of the analysed data precluded the potential for independent verification, but this data collection method is established as part of clinical audit protocols (Baughan *et al*, 2009; Hansen *et al*, 2011; Rubin *et al*, 2011; Swann *et al*, 2018).

10.5.1.2 Missing values of diagnostic timeliness

As is often the case for epidemiological studies, some patients had missing values for the key outcomes of interest (particularly regarding the patient and the primary care intervals). The degree of missing data for these two intervals was comparable (and often lower) than those reported by other studies of symptomatic cancer patients (Hansen *et al*, 2011; Walter *et al*, 2016a, 2016b; Leiva *et al*, 2017). It should be noted that for some patients, interval data may have been 'appropriately' missing. For example, patients diagnosed without (relevant) symptomatic presentations would not have a date of symptom onset, nor a relevant patient interval (as explored in Chapter 5).

Multiple imputation of interval values in Chapters 6–8 was not considered optimal, because the regression analyses examined intervals as an outcome variable. In such circumstances, multiple imputation of the outcome variable is of little value if variables included in the imputation model are also included as covariates in the final analysis model, as would have been the case (Sterne *et al*, 2009). Therefore, I undertook complete case analysis for the analyses presented in this thesis, although I have also described the degree and patterns of missing interval values (see Appendices 6.2, 7.2, and 8.1).

10.5.1.3 Unmeasured confounders

The NACDPC did not collect any information regarding the socio-economic status or the level of health literacy of the included patients, which tend to be associated with symptom

knowledge, appraisal, and barriers to presentation and could have therefore influenced the length of the patient interval (Macleod *et al*, 2009; Niksic *et al*, 2015; Marcu *et al*, 2016). Similarly, there was no information on comorbidities which may have influenced the appraisal of symptoms by both the patient and the consulting general practitioner, thereby affecting the length of the patient and the primary care intervals (Friese *et al*, 2009; Walter *et al*, 2016a; Mounce *et al*, 2017; Salika *et al*, 2017).

Some of the variation in diagnostic timeliness by symptoms described throughout this thesis may be confounded by the above (and other unmeasured) variables. However, given the large size of the observed symptom-related variation in the examined intervals after adjustment for age, sex, and ethnicity, it is very unlikely that the observed variation in intervals by symptom could by fully explained by other potential (unmeasured) confounders.

10.5.1.4 External validity of findings

A further consideration is the generalisability of the findings given the time and context of data collection in the NACDPC. Regarding the nature of presenting symptoms (for example, the symptom signature of breast cancer), the findings are unlikely to have changed substantially over time, given that they reflect underlying tumour factors and biological mechanisms.

However, whether the reported associations between presenting symptoms and intervals remained stable over time deserves further discussion. The NACDPC data were collected in 2009–10 before the Be Clear on Cancer campaigns were launched (in 2012), and the NICE clinical guidelines were updated (in 2015) (NICE, 2015; Public Health England, 2016b). It is therefore plausible (and indeed desirable) to hypothesise that increased knowledge of likely cancer symptoms among the general population (including those who are diagnosed with cancer), and increased awareness of cancer diagnosis among healthcare professionals may have resulted in shorter patient and primary care intervals among more contemporary cohorts of cancer patients than those described among the cancer patients captured in the NACDPC.

Nevertheless, recently published findings from the 2014 National Cancer Diagnosis Audit (NCDA) indicate comparable patterns of time to referral and number of pre-referral consultations by cancer site (Swann *et al*, 2018). Unfortunately, the NCDA did not collect information regarding the patient interval, and so the potential effect of symptom awareness campaigns on time to help-seeking cannot be compared using audit data. However, recent evidence indicates that levels of cancer symptom awareness among the general population have remained largely unchanged (Cancer Research UK, 2016b).

10.5.2 Methodological considerations

10.5.2.1 Information on presenting symptoms

The availability of free-text information on symptoms is unique to the NACDPC; this was coded at the lowest level possible to maximise granularity of the coded data. Symptom information was invalid in 577 patients (see Figure 4.1 in Chapter 4); a proportion of these patients may have been diagnosed incidentally or asymptomatically, in the absence of tumour-related symptoms (as investigated in Chapter 5).

10.5.2.2 Information on intervals

The distribution in length of the patient and primary care intervals among the NACDPC study population was significantly skewed (see Figure 10.1). Analysis with traditional linear regression models would have been unsuitable, due to violations of the assumptions of normality of the corresponding residuals.

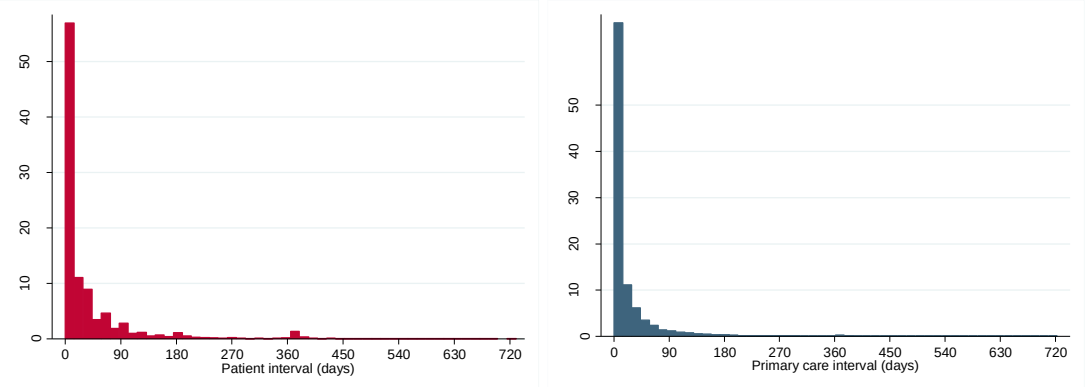


Figure 10.1 Distribution of patient interval (left) and primary care interval (right) values among the NACDPC patient population
Tests for skew and kurtosis found p<0.001 for both intervals.

There are several established approaches for examining diagnostic interval data which are typically right-skewed in early diagnosis research. Variation in diagnostic timeliness is often described with summary metrics (e.g. mean, median and centile values) without use of regression frameworks (Baughan *et al*, 2009; Hansen *et al*, 2011; Esteva *et al*, 2013; Dobson *et al*, 2014; Lyratzopoulos *et al*, 2015; Helsper *et al*, 2017). Alternatively, when regression analysis is used, interval data are often dichotomised as 'timely' or 'untimely' using cut-off points of prior assumed interest (Dobson *et al*, 2014; Keeble *et al*, 2014; Lim *et al*, 2014). Both approaches are useful, but I deemed it important to be able to examine differences in diagnostic intervals in greater detail after adjustment for possible confounders.

For the first empirical study examining diagnostic timeliness in this thesis (Chapter 6: The symptom signature of breast cancer and associated diagnostic intervals) I chose to use quantile regression. This approach allows between-group comparisons at different specified points of the distribution and has previously been used to examine diagnostic timeliness (Jensen *et al*, 2014; Herbert *et al*, 2018), although as it is a non-parametric method, the resulting model had less statistical power than parametric methods.

For the second and third studies that examined diagnostic timeliness (Chapter 7 and 8), I used log-linked generalised linear models (GLM). As GLM is a parametric approach, the resulting models provided greater statistical power compared to quantile regression. Retrospective re-analysis of Chapter 6 (The symptom signature of breast cancer and associated diagnostic intervals) using GLM regression found comparable results (data not shown).

Another possible methodological approach could have been the use of time-to-event (survival) regression frameworks encompassing semi-parametric methods such as Cox regression, or flexible parametric survival modelling, where the date of first presentation (with regard to the patient interval) or date of referral (with regard to the primary care interval) is treated as the event of interest (Walter *et al,* 2015, 2016a, 2016b).

10.6 Implications for practice and research

As outlined in Chapter 1, early diagnosis interventions can be considered complex in nature, by virtue of being conducted at population level, and involving a multi-modal strategy with different likely mechanisms (Campbell *et al,* 2007; Craig *et al,* 2008). Emerging evidence relating to early diagnosis interventions point to the challenges of translating the theoretical understanding of symptomatic cancer diagnosis into a real-world setting (Ironmonger *et al,* 2014; Moffat *et al,* 2015; Jensen *et al,* 2016; Emery *et al,* 2017). Understanding associations between symptomatic presentations and diagnostic timeliness among individuals who are diagnosed with cancer could strengthen the assumptions underpinning these early diagnosis interventions.

10.6.1 How does this affect early diagnosis intervention design and evaluation?

Symptom awareness campaigns continue to be a common and popular feature of the "early diagnosis toolbox" at national and international levels (Scottish Government, 2012; Cancer Australia, 2013; Public Health England, 2016b; World Wide Breast Cancer, 2016; CDC, 2017). Findings from Chapter 6 and 7 suggest that among individuals diagnosed with cancer, certain symptoms (e.g. non-lump breast symptoms and dysphagia) are associated with longer time to help-seeking than others; this could inform a more evidence-based approach to campaign design. For example, with regard to awareness campaigns promoting the earlier diagnosis of breast cancer, the findings of Chapter 6 iterate the importance of encompassing non-lump symptoms such as nipple abnormalities, breast skin abnormalities, or breast pain in public health education campaigns.

Furthermore, the findings of Chapter 9 suggest that alarm symptoms are variably associated with early stage disease among prompt presenters. In other words, interventions promoting the timely presentation and investigation of alarm symptoms alone are unlikely to result in earlier

diagnosis of the cancer patient population, all else being equal, particularly given that a large proportion of patients diagnosed with cancer do not initially present with alarm symptoms. Evidence regarding variation in time to referral by cancer symptom (as investigated in Chapter 8) could guide the choice of non-specific symptom referral criteria and diagnostic test cascades for clinical care pathways and multidisciplinary diagnostic centres (MDCs) targeting such presentations (Fuller *et al*, 2016; Moseholm & Lindhardt, 2017; Næser *et al*, 2017).

Nonetheless, the epidemiology of presenting symptoms and associated diagnostic timeliness constitutes only one of many factors that need to be taken into account when designing an intervention promoting the early diagnosis of cancer. Other factors must also be taken into consideration, akin to the criteria proposed by Wilson and Jungner when considering population-based screening programmes (Wilson & Jungner, 1968) and more recently, in acknowledgement of more holistic systems science approaches (Mooney, 2017).

Firstly, most individuals presenting with a particular symptom will not be subsequently diagnosed with cancer. Thus, the prevalence of a symptom among the general population (and not just cancer patients, as examinable in the NACDPC data) and the resulting predictive value for malignancy must be considered. Secondly, promoting the earlier presentation of individuals with possible cancer symptoms to improve clinical outcomes is conditional on the fact that the given symptoms are associated with early stage disease at diagnosis. Findings from Chapter 9 indicate that this may not be as clear as previously assumed. Further, for healthcare interventions, downstream capacity for the anticipated increase in investigations or referrals will be an important factor influencing effectiveness and success of the intervention. Relatedly, broader contextual factors such as available resource and opportunity costs at population level are likely to be important in the design, delivery, and evaluation of interventions promoting earlier diagnosis.

An additional pre-requisite is the effectiveness of the interventions in achieving the desired behaviour change among the target population. Raising awareness of the likely symptoms of cancer may be insufficient in encouraging prompt help-seeking among individuals; other psychosocial factors such as 'expectation' (which encompasses cancer fear and fatalism) could impede prompt presentation regardless of symptom awareness (Whitaker *et al*, 2015b; Vrinten *et al*, 2017). Post-presentation, cognitive biases and the physician's 'gut feeling' may have greater influence over the decision to utilise fast-track referral pathways than the presence or absence of particular symptoms, which could also affect the timeliness of investigation and diagnosis (Stolper *et al*, 2011; Jensen *et al*, 2014; Donker *et al*, 2016; Holtedahl *et al*, 2017).

Many of the implications for design as discussed above also apply to evaluative strategies. Further, evidence regarding the cancer signature associated with individual symptoms (as demonstrated in Chapter 7 and 8) may be helpful for evaluation of early diagnosis interventions.

This is particularly the case if symptom-based campaigns (rather than approaches centred on a given cancer-site) are to be adopted, as signalled by the recent 'abdominal symptoms' campaign by Be Clear on Cancer (Public Health England & Department of Health, 2015; Public Health England, 2017b).

10.6.2 Future directions

The findings of this thesis have several implications for future research in early diagnosis of symptomatic cancer. Firstly, examining associations between presenting symptoms and diagnostic intervals using different data would be a useful comparison to the findings presented in this thesis. Drawing upon self-reported symptom and interval data may be particularly useful given the opposing strengths and limitations as compared to audit data as discussed in Chapter 3 and 4. Nevertheless, capturing such information among a large number of cancer patients is likely to be costly and resource intensive, and so continuing enquiries in secondary data is likely to remain prudent. Electronic health systems were not created to collect data for research, and therefore analyses are at risk of many biases (Verheij *et al*, 2018). The development of 'smarter' systems that encourage more accurate and complete data capture, alongside the advancement of machine learning approaches that enable the analysis of free-text on a wider scale, could overcome many of the current limitations of records-based data capture on presenting symptoms and intervals.

Frameworks from different disciplinary traditions can additionally contribute to the evidence base underlying early diagnosis interventions (as mentioned in Chapter 1). For example, drawing from research in diagnostic errors, it will be useful to examine clinical decisions taken following presentations with particular symptoms; examining guideline-concordant and guideline-discordant diagnostic investigation or referral decisions for different symptomatic presentations may reveal likely mechanisms for minimising missed diagnostic opportunities.

Expanding the lines of enquiry pursued in this thesis to other distinct populations of cancer patients may also be useful. For example, teenager and young adults (TYA) are diagnosed with a very different case-mix of cancer sites compared to adults; this also means a distinct range of symptomatic presentations. I have recently begun preliminary research on the epidemiology of self-reported presenting symptoms among a cohort of TYA cancer patients (Herbert *et al*, 2018). This indicates that the average 'symptom burden' and reported symptom frequencies tend to be greater than those previously reported in studies of electronic databases of primary care records, likely reflecting methodological differences as discussed in Chapter 3. Nevertheless, much of the insight and approaches presented in this thesis may be translated to this unique population.

Chapter 5 identified a number of patients who were diagnosed with cancer during follow-up for a pre-existing cancer diagnosis; such diagnoses are likely to become more common as cancer

survivorship increases (Murphy *et al*, 2017). Among such cases, prior experience of cancer (and the diagnostic pathway) may influence associations between symptoms and intervals to diagnosis and warrants further examination.

A final consideration is that interventions that aim to reduce time to diagnosis for cancer are also likely to improve the diagnostic experience of other diseases (Brooker *et al*, 2014; Battaglia *et al*, 2015; Burton *et al*, 2017; Thormann *et al*, 2017). Coordinating such efforts by taking a holistic patient- (and therefore symptom-) centred approach represents a promising avenue for future research, particularly given that many symptoms indicative of cancer can also represent other serious diseases (Jones *et al*, 2009; Dobson *et al*, 2014; Stapley *et al*, 2017).

10.7 Conclusions

Interventions promoting early diagnosis are an increasingly common component of cancer control strategies. Symptomatic presentations are often a defining feature of such interventions but to date, epidemiological evidence in this regard has been limited. I aimed to examine the presenting symptoms of cancer patients and associated diagnostic intervals, using cross-sectional data collected through a national clinical audit on a representative incident cohort of cancer patients.

My enquiries included the symptom signature of breast cancer; the cancer signatures of abdominal symptoms; patterns of variation in diagnostic timeliness; and associations between alarm symptoms and stage at diagnosis. The findings contribute to the evidence base underlying public health and health system interventions promoting early diagnosis, and can motivate future studies using similar data sources and related methodologies. Strengthening the evidence base of early diagnosis interventions in this way will ultimately lead to improvements in the diagnostic experience and outcomes for individuals diagnosed with cancer.

References

Abdel-Rahman M, Stockton D, Rachet B, Hakulinen T, Coleman MP (2009) What if cancer survival in Britain were the same as in Europe: how many deaths are avoidable? *Br J Cancer* **101**: S115–S124, doi:10.1038/sj.bjc.6605401.

Abel GA, Saunders CL, Lyratzopoulos G (2016) Post-sampling mortality and non-response patterns in the English Cancer Patient Experience Survey: Implications for epidemiological studies based on surveys of cancer patients. *Cancer Epidemiol* **41**: 34–41, doi:10.1016/j.canep.2015.12.010.

Ades AE, Biswas M, Welton NJ, Hamilton W (2014) Symptom lead time distribution in lung cancer: natural history and prospects for early diagnosis. *Int J Epidemiol* **43**: 1865–1873, doi:10.1093/ije/dyu174.

Ahn HS, Kim HJ, Welch HG (2014) Korea's Thyroid-Cancer "Epidemic" — Screening and Overdiagnosis. *N Engl J Med* **371**: 1765–1767, doi:10.1056/NEJMp1409841.

Alexiusdottir KK, Möller PH, Snaebjornsson P, Jonasson L, Olafsdottir EJ, Björnsson ES, Tryggvadottir L, Jonasson JG (2012) Association of symptoms of colon cancer patients with tumor location and TNM tumor stage. *Scand J Gastroenterol* **47**: 795–801, doi:10.3109/00365521.2012.672589.

Allgar VL, Neal RD (2005) Delays in the diagnosis of six cancers: analysis of data from the National Survey of NHS Patients: Cancer. *Br J Cancer* **92**: 1959–1970, doi:10.1038/sj.bjc.6602587.

Andersen BL, Cacioppo JT, Roberts DC (1995) Delay in seeking a cancer diagnosis: Delay stages and psychophysiological comparison processes. *Br J Soc Psychol* **34**: 33–52, doi:10.1111/j.2044-8309.1995.tb01047.x.

Astin M, Griffin T, Neal RD, Rose P, Hamilton W (2011) The diagnostic value of symptoms for colorectal cancer in primary care: a systematic review. *Br J Gen Pract* **61**: 231–243, doi:10.3399/bjgp11X572427.

Astin MP, Martins T, Welton N, Neal RD, Rose PW, Hamilton W (2015) Diagnostic value of symptoms of oesophagogastric cancers in primary care: a systematic review and meta-analysis. *Br J Gen Pract* **65**: e677–e691, doi:10.3399/bjgp15X686941.

Austoker J, Bankhead C, Forbes LJL, Atkins L, Martin F, Robb K, Wardle J, Ramirez AJ (2009) Interventions to promote cancer awareness and early presentation: systematic review. *Br J Cancer* **101 Suppl**: S31-9, doi:10.1038/sj.bjc.6605388.

Avilés-Izquierdo JA, Molina-López I, Rodríguez-Lomba E, Marquez-Rodas I, Suarez-Fernandez R, Lazaro-Ochaita P (2016) Who detects melanoma? Impact of detection patterns on characteristics and prognosis of patients with melanoma. *J Am Acad Dermatol* 1–8, doi:10.1016/j.jaad.2016.07.009.

Bailey SE, Ukoumunne OC, Shephard EA, Hamilton W (2017) Clinical relevance of thrombocytosis in primary care: a prospective cohort study of cancer incidence using English electronic medical records and cancer registry data. *Br J Gen Pract* **67**: e405–e413, doi:10.3399/bjgp17X691109.

Bankhead CR, Kehoe ST, Austoker J (2005) Symptoms associated with diagnosis of ovarian cancer: a systematic review. *BJOG An Int J Obstet Gynaecol* **112**: 857–865, doi:10.1111/j.1471-0528.2005.00572.x.

Battaglia M, Nigi L, Dotta F (2015) Towards an Earlier and Timely Diagnosis of Type 1 Diabetes: Is it Time to Change Criteria to Define Disease Onset? *Curr Diab Rep* **15**: 115, doi:10.1007/s11892-015-0690-6.

Baughan P, O'Neill B, Fletcher E (2009) Auditing the diagnosis of cancer in primary care: the experience in Scotland. *Br J Cancer* **101 Suppl**: S87-91, doi:10.1038/sj.bjc.6605397.

Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM (2015) The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLOS Med* **12**: e1001885, doi:10.1371/journal.pmed.1001885.

Bird-Lieberman EL, Fitzgerald RC (2009) Early diagnosis of oesophageal cancer. *Br J Cancer* **101**: 1–6, doi:10.1038/sj.bjc.6605126.

Biswas M, Ades AE, Hamilton W (2015) Symptom lead times in lung and colorectal cancers: what are the benefits of symptom-based approaches to early diagnosis? *Br J Cancer* **112**: 271–277, doi:10.1038/bjc.2014.597.

Blak B, Thompson M, Dattani H, Bourke A (2011) Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *J Innov Heal Informatics* **19**: 251–255, doi:10.14236/jhi.v19i4.820.

Blumen H, Fitch K, Polkus V (2016) Comparison of Treatment Costs for Breast Cancer, by Tumor Stage and Type of Service. *Am Heal drug benefits* **9**: 23–32.

Booth TC, Jackson A, Wardlaw JM, Taylor SA, Waldman AD (2010) Incidental findings found in 'healthy' volunteers during imaging performed for research: Current legal and ethical implications. *Br J Radiol* **83**: 456–465, doi:10.1259/bjr/15877332.

Brindle L, Pope C, Corner J, Leydon G, Banerjee A (2012) Eliciting symptoms interpreted as normal by patients with early-stage lung cancer: Could GP elicitation of normalised symptoms reduce delay in diagnosis? Cross-sectional interview study. *BMJ Open* **2**: e001977, doi:10.1136/bmjopen-2012-001977.

Brocken P, Prins JB, Dekhuijzen PNR, Van Der Heijden HFM (2012) The faster the better? - A systematic review on distress in the diagnostic phase of suspected cancer, and the influence of rapid diagnostic pathways. *Psychooncology* **21**: 1–10, doi:10.1002/pon.1929.

Brooker D, Fontaine J La, Evans S, Bray J, Saad K (2014) Public health guidance to facilitate timely diagnosis of dementia: ALzheimer's COoperative Valuation in Europe recommendations. *Int J Geriatr Psychiatry* **29**: 682–693, doi:10.1002/gps.4066.

Bruyninckx R, Buntinx F, Aertgeerts B, Van Casteren V (2003) The diagnostic value of macroscopic haematuria for the diagnosis of urological cancer in general practice. *Br J Gen Pract* **53**: 31–35.

Burgess C, Ramirez A, Richards M, Love S (1998) Who and what influences delayed presentation in breast cancer? *Br J Cancer* **77**: 1343–1348, doi:10.1038/bjc.1998.224.

Burgess CC, Potts HWW, Hamed H, Bish AM, Hunter MS, Richards MA, Ramirez AJ (2006) Why do older women delay presentation with breast cancer symptoms? *Psychooncology* **15**: 962–968, doi:10.1002/pon.1030.

Burton C, Iversen L, Bhattacharya S, Ayansina D, Saraswat L, Sleeman D (2017) Pointers to earlier diagnosis of endometriosis: a nested case-control study using primary care electronic health records. *Br J Gen Pract* bjgp17X693497, doi:10.3399/bjgp17X693497.

Calanzani N, Weller D, Campbell C (2017) The characteristics of national health initiatives promoting earlier cancer diagnosis among adult populations : a systematic review protocol. 1–10, doi:10.1136/bmjopen-2017-015922.

Campbell J, Pyer M, Rogers S, Jones J, Ramirez AJJ, Forbes LJLJL (2015) Promoting early presentation of breast cancer in women over 70 years old in general practice. *J Public Health (Bangkok)* **38**: 1–8, doi:10.1093/pubmed/fdv125.

Campbell NCN, Murray E, Darbyshire J, Emery J, Farmer A, Griffiths F, Guthrie B, Lester H, Wilson P, Kinmonth AL Al (2007) Designing and evaluating complex interventions to improve health care. *Br Med J* **334**: 455–459, doi:10.1136/bmj.39108.379965.BE.

Cancer Australia (2013) Lung Cancer Awareness Month https://canceraustralia.gov.au/healthy-living/campaigns-events/lung-cancer-awareness-month.

Cancer Australia (2014) Cancer Australia Strategic Plan 2014-2019 (Surrey Hills).

Cancer Australia (2016) Campaigns and Events https://canceraustralia.gov.au/healthy-living/campaigns-events.

Cancer Council Austalia (2017) Cancer Guidelines Wiki platform https://wiki.cancer.org.au/australia/Guidelines.

Cancer Research UK (2014) About Be Clear on Cancer http://www.cancerresearchuk.org/health-professional/awareness-and-prevention/be-clear-on-cancer/about-be-clear-on-cancer.

Cancer Research UK (2016a) Be Clear on Cancer Programme evaluation http://www.cancerresearchuk.org/health-professional/early-diagnosis-activities/be-clear-on-cancer/programme-evaluation.

Cancer Research UK (2016b) Cancer Awareness Measure (CAM) Key Findings Report; 2014 & Trends Analysis (2008-2014).

Cancer Research UK (2016c) How MDCs could improve early diagnosis https://www.cancerresearchuk.org/sites/default/files/how_mdcs_could_improve_early_diagnosis.pdf.

Cancer Research UK (2017) Breast Cancer Statistics http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer.

CDC (2017) Get the facts about gynecological cancer https://www.cdc.gov/cancer/knowledge/publications/brochures.htm.

Coleman M, Forman D, Bryant H, Butler J, Rachet B, Maringe C, Nur U, Tracey E, Coory M, Hatcher J, McGahan C, Turner D, Marrett L, Gjerstorff M, Johannesen T, Adolfsson J, Lambe M, Lawrence G, Meechan D, Morris E, Middleton R, Steward J, Richards M (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. *Lancet* **377**: 127–138, doi:10.1016/S0140-6736(10)62231-3.

Collins GS, Altman DG (2012a) Identifying patients with undetected colorectal cancer: an independent validation of QCancer (Colorectal). *Br J Cancer* **107**: 260–265, doi:10.1038/bjc.2012.266.

Collins GS, Altman DG (2012b) Identifying patients with undetected gastro-oesophageal cancer in primary care: External validation of QCancer® (Gastro-Oesophageal). *Eur J Cancer* **49**: 1040–1048, doi:10.1016/j.ejca.2012.10.023.

Collins GS, Altman DG (2012c) Identifying women with undetected ovarian cancer: independent and external validation of QCancer ® (Ovarian) prediction model. *Eur J Cancer Care (Engl)* **22**: 423–429, doi:10.1111/ecc.12015.

Collins GS, Altman DG (2013a) Identifying patients with undetected pancreatic cancer in primary care: an independent and external validation of QCancer® (Pancreas). *Br J Gen Pract* **63**: 636–642, doi:10.3399/bjgp13X671623.

Collins GS, Altman DG (2013b) Identifying patients with undetected renal tract cancer in primary care: an independent and external validation of QCancer® (Renal) prediction model. *Cancer*

Epidemiol **37**: 115–120, doi:10.1016/j.canep.2012.11.005.

Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M (2008) Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ* **337**: a1655, doi:10.1136/bmj.a1655.

Crawford SC (2002) The waiting time paradox: population based retrospective study of treatment delay and survival of women with endometrial cancer in Scotland. *BMJ* **325**: 196–196, doi:10.1136/bmj.325.7357.196.

Cromme SK, Whitaker KL, Winstanley K, Renzi C, Smith CF, Wardle J (2016) Worrying about wasting GP time as a barrier to help-seeking: a community-based, qualitative study. *Br J Gen Pract* **66**: e474–e482, doi:10.3399/bjgp16X685621.

Cufari ME, Proli C, Phull M, Nicholson A, Al-Sahaf M, Raubenheimer H, Asadi N, Perikleous P, Jordan S, Dusmet M, Ladas G, Anikin V, Beddow E, Mcgonigle N, Shedden L, Chavan H, Leung M, Lim E (2016) Increasing incidence of non-smoking lung cancer: presentation of patients with early disease to a tertiary institution in the UK. *Lung Cancer* **91**: S17–S18, doi:10.1016/S0169-5002(16)30066-6.

Dahl TL, Vedsted P, Jensen H (2017) The effect of standardised cancer pathways on Danish cancer patients' dissatisfaction with waiting time. *Dan Med J* **64**:.

Dalberg K, Hellborg H, Wärnberg F (2008) Paget's disease of the nipple in a population based cohort. *Breast Cancer Res Treat* **111**: 313–319, doi:10.1007/s10549-007-9783-5.

Danish National Board of Health (2005) National Cancer Plan II: National Board of Health recommendations for improving cancer healthcare services (Copenhagen).

Davies L, Hendrickson CD, Hanson GS, L W, L C, F J (2017) Experience of US Patients Who Self-identify as Having an Overdiagnosed Thyroid Cancer. *JAMA Otolaryngol Neck Surg* **25**: 147–156, doi:10.1001/jamaoto.2016.4749.

Davies L, Ouellette M, Hunter M, Welch HG (2010) The increasing incidence of small thyroid cancers: Where are the cases coming from? *Laryngoscope* **120**: 2446–2451, doi:10.1002/lary.21076.

Department of Health (2000) HSC2000/013 Referral Guidelines for suspected cancer (London).

Din NU, Ukoumunne OC, Rubin G, Hamilton W, Carter B, Stapley S, Neal RD (2015) Age and Gender Variations in Cancer Diagnostic Intervals in 15 Cancers: Analysis of Data from the UK Clinical Practice Research Datalink. *PLoS One* **10**: e0127717, doi:10.1371/journal.pone.0127717.

Dobson C, Russell A, Rubin G (2014) Patient delay in cancer diagnosis: what do we really mean and can we be more specific? *BMC Health Serv Res* **14**: 387, doi:10.1186/1472-6963-14-387.

Dobson CM, Russell A, Brown SR, Rubin GP (2018) The role of social context in symptom appraisal and help-seeking among people with lung or colorectal symptoms: A qualitative interview studyy. *Eur J Cancer Care (Engl)* 1–13, doi:10.1111/ecc.12815.

Donker GA, Wiersma E, Van Der Hoek L, Heins M (2016) Determinants of general practitioner's cancer-related gut feelings-a prospective cohort study. *BMJ Open* **6**: 1–8, doi:10.1136/bmjopen-2016-012511.

Dove-Edwin I (2005) Prevention of colorectal cancer by colonoscopic surveillance in individuals with a family history of colorectal cancer: 16 year, prospective, follow-up study. *Bmj* **331**: 1047–0, doi:10.1136/bmj.38606.794560.EB.

Dyrop HB, Safwat A, Vedsted P, Maretty-Kongstad K, Hansen BH, Jørgensen PH, Baad-Hansen T, Keller J (2016) Characteristics of 64 sarcoma patients referred to a sarcoma center after unplanned excision. *J Surg Oncol* **113**: 235–239, doi:10.1002/jso.24137.

Ebell MH, Culp MB, Radke TJ (2016) A Systematic Review of Symptoms for the Diagnosis of Ovarian Cancer. *Am J Prev Med* **50**: 384–394, doi:10.1016/j.amepre.2015.09.023.

Eberl MM, Phillips RL, Lamberts H, Okkes I, Mahoney MC (2008) Characterizing Breast Symptoms in Family Practice. *Ann Fam Med* **6**: 528–533, doi:10.1370/afm.905.

Elliss-Brookes L, McPhail S, Ives A, Greenslade M, Shelton J, Hiom S, Richards M (2012) Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. *Br J Cancer* **107**: 1220–1226, doi:10.1038/bjc.2012.408.

Elnegaard S, Andersen RS, Pedersen AF, Larsen PV, Søndergaard J, Rasmussen S, Balasubramaniam K, Svendsen RP, Vedsted P, Jarbøl DE (2015) Self-reported symptoms and healthcare seeking in the general population -exploring "The Symptom Iceberg". *BMC Public Health* **15**: 685, doi:10.1186/s12889-015-2034-5.

Elnegaard S, Pedersen AF, Sand Andersen R, Christensen R de-P, Jarbøl DE (2017) What triggers healthcare-seeking behaviour when experiencing a symptom? Results from a population-based survey. *BJGP Open* **1**: BJGP-2016-0775, doi:10.3399/bjgpopen17X100761.

Emery JD, Gray V, Walter FM, Cheetham S, Croager EJ, Slevin T, Saunders C, Threlfall T, Auret K, Nowak AK, Geelhoed E, Bulsara M, Holman CDJ (2017) The Improving Rural Cancer Outcomes Trial: a cluster-randomised controlled trial of a complex intervention to reduce time to diagnosis in rural cancer patients in Western Australia. *Br J Cancer* **117**: 1459–1469, doi:10.1038/bjc.2017.310.

Emery JD, Walter FM, Gray V, Sinclair C, Howting D, Bulsara M, Bulsara C, Webster A, Auret K, Saunders C, Nowak A, D'Arcy Holmand C (2013) Diagnosing cancer in the bush: A mixed-methods study of symptom appraisal and help-seeking behaviour in people with cancer from rural Western Australia. *Fam Pract* **30**: 294–301, doi:10.1093/fampra/cms087.

Esserman LJ, Thompson IM, Reid B, Nelson P, Ransohoff DF, Welch HG, Hwang S, Berry D a., Kinzler KW, Black WC, Bissell M, Parnes H, Srivastava S (2014) Addressing overdiagnosis and overtreatment in cancer: A prescription for change. *Lancet Oncol* **15**: e234–e242, doi:10.1016/S1470-2045(13)70598-9.

Esteva M, Leiva A, Ramos M, Pita-Fernández S, González-Luján L, Casamitjana M, Sánchez MA, Pértega-Díaz S, Ruiz A, Gonzalez-Santamaría P, Martín-Rabadán M, Costa-Alcaraz AM, Espí A, Macià F, Segura JM, Lafita S, Arnal-Monreal F, Amengual I, Boscá-Watts MM, Hospital A, Manzano H, Magallón R (2013) Factors related with symptom duration until diagnosis and treatment of symptomatic colorectal cancer. *BMC Cancer* **13**: 87, doi:10.1186/1471-2407-13-87.

Esteva M, Ramos M, Cabeza E, Llobera J, Ruiz A, Pita S, Segura JM, Cortes JM, Gonzalez-Lujan L (2007) Factors influencing delay in the diagnosis of colorectal cancer: a study protocol. *BMC Cancer* **7**: 86, doi:10.1186/1471-2407-7-86.

Evans J, Chapple A, Salisbury H, Corrie P, Ziebland S (2014) 'It can't be very important because it comes and goes'--patients' accounts of intermittent symptoms preceding a pancreatic cancer diagnosis: a qualitative study. *BMJ Open* **4**: e004215, doi:10.1136/bmjopen-2013-004215.

Ewing M, Naredi P, Zhang C, Lindsköld L, Månsson J (2018) Clinical features of patients with non-metastatic lung cancer in primary care: a case-control study. *BJGP Open* bjgpopen18X101397, doi:10.3399/bjgpopen18X101397.

Fang TC, Oh YS, Szabo A, Khan A, Dua KS (2016) Utility of dysphagia grade in predicting endoscopic ultrasound T-stage of non-metastatic esophageal cancer. *Dis Esophagus* **29**: 642–648, doi:10.1111/dote.12394.

Forbes LJL, Simon AE, Warburton F, Boniface D, Brain KE, Dessaix A, Donnelly C, Haynes K, Hvidberg L, Lagerlund M, Lockwood G, Tishelman C, Vedsted P, Vigmostad MN, Ramirez AJ, 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? *Br J Cancer* **108**: 292–300, doi:10.1038/bjc.2012.542.

Forbes LJL, Warburton F, Richards MA, Ramirez AJ (2014) Risk factors for delay in symptomatic presentation: a survey of cancer patients. *Br J Cancer* **111**: 581–588, doi:10.1038/bjc.2014.304.

Forrest LF, Adams J, White M, Rubin G (2014) Factors associated with timeliness of post-primary care referral, diagnosis and treatment for lung cancer: population-based, data-linkage study. *Br J Cancer* **111**: 1843–1851, doi:10.1038/bjc.2014.472.

Forster AS, Renzi C, Lyratzopoulos G (2018) Diagnosing cancer in patients with ' non-alarm ' symptoms : Learning from diagnostic care innovations in Denmark. *Cancer Epidemiol* **54**: 101–103, doi:10.1016/j.canep.2018.03.011.

Friedman LC, Kalidas M, Elledge R, Dulay MF, Romero C, Chang J, Liscum KR (2006) Medical and psychosocial predictors of delay in seeking medical consultation for breast symptoms in women in a public sector setting. *J Behav Med* **29**: 327–334, doi:10.1007/s10865-006-9059-2.

Friese CR, Abel GA, Magazu LS, Neville BA, Richardson LC, Earle CC (2009) Diagnostic delay and complications for older adults with multiple myeloma. *Leuk Lymphoma* **50**: 392–400, doi:10.1080/10428190902741471.

Fuller E, Fitzgerald K, Hiom S (2016) Accelerate, Coordinate, Evaluate Programme: a new approach to cancer diagnosis. *Br J Gen Pract* **66**: 176–177, doi:10.3399/bjgp16X684457.

George A, Grimer R (2012) Early symptoms of bone and soft tissue sarcomas: could they be diagnosed earlier? *Ann R Coll Surg Engl* **94**: 261–266, doi:10.1308/003588412X13171221590016.

Gibson LM, Littlejohns TJ, Adamska L, Garratt S, Doherty N, Wardlaw JM, Maskell G, Parker M, Brownsword R, Matthews PM, Collins R, Allen NE, Sellors J, Sudlow CL (2017) Impact of detecting potentially serious incidental findings during multi-modal imaging. *Wellcome Open Res* **2**: 114, doi:10.12688/wellcomeopenres.13181.1.

Goff BA, Mandel L, Muntz HG, Melancon CH (2000) Ovarian carcinoma diagnosis. *Cancer* **89**: 2068–2075, doi:10.1002/1097-0142(20001115)89:10<2068::AID-CNCR6>3.0.CO;2-Z.

Grooss K, Hjertholm P, Carlsen AH, Vedsted P (2016) Patients with cancer and change of general practice: A Danish population-based cohort study. *Br J Gen Pract* **66**: e491–e498, doi:10.3399/bjgp16X685633.

Guldbrandt L, Fenger-Grøn M, Rasmussen T, Jensen H, Vedsted P (2015) The role of general practice in routes to diagnosis of lung cancer in Denmark: a population-based study of general practice involvement, diagnostic activity and diagnostic intervals. *BMC Health Serv Res* **15**: 21, doi:10.1186/s12913-014-0656-4.

Halpern MT, Ward EM, Pavluck AL, Schrag NM, Bian J, Chen AY (2008) Association of insurance status and ethnicity with cancer stage at diagnosis for 12 cancer sites : a retrospective analysis. **9**: doi:10.1016/S1470-2045(08)70032-9.

Hamilton W (2009a) Five misconceptions in cancer diagnosis. *Br J Gen Pract* **59**: 441–447, doi:10.3399/bjgp09X420860.

Hamilton W (2009b) The CAPER studies: five case-control studies aimed at identifying and quantifying the risk of cancer in symptomatic primary care patients. *Br J Cancer* **101 Suppl**: S80-6, doi:10.1038/sj.bjc.6605396.

Hamilton W, Kernick D (2007) Clinical features of primary brain tumours: a case-control study using electronic primary care records. *Br J Gen Pract* **57**: 695–699.

Hamilton W, Lancashire R, Sharp D, Peters TJ, Cheng K, Marshall T (2009a) The risk of colorectal

cancer with symptoms at different ages and between the sexes: a case-control study. *BMC Med* **7**: 17, doi:10.1186/1741-7015-7-17.

Hamilton W, Peters TJ, Bankhead C, Sharp D (2009b) Risk of ovarian cancer in women with symptoms in primary care: population based case-control study. *BMJ* **339**: b2998, doi:10.1136/bmj.b2998.

Hamilton W, Peters TJ, 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, doi:10.1136/thx.2005.045880.

Hamilton W, Round A, Sharp D, Peters TJ (2005b) Clinical features of colorectal cancer before diagnosis: a population-based case-control study. *Br J Cancer* **93**: 399–405, doi:10.1038/sj.bjc.6602714.

Hansen RP, Vedsted P, Sokolowski I, Søndergaard J, Olesen F (2011) Time intervals from first symptom to treatment of cancer: a cohort study of 2,212 newly diagnosed cancer patients. *BMC Health Serv Res* **11**: 284, doi:10.1186/1472-6963-11-284.

Healthcare Improvement Scotland (2014) Scottish Referral Guidelines for Suspected Cancer.

Helsper C, van Erp N, Peeters P, de Wit N (2017) Time to diagnosis and treatment for cancer patients in the Netherlands: Room for improvement? *Eur J Cancer* **87**: 113–121, doi:10.1016/j.ejca.2017.10.003.

Herbert A, Lyratzopoulos G, Whelan J, Taylor RM, Barber J, Gibson F, Fern LA (2018) Diagnostic timeliness in adolescents and young adults with cancer: a cross-sectional analysis of the BRIGHTLIGHT cohort. *Lancet Child Adolesc Heal* **2**: 180–190, doi:10.1016/S2352-4642(18)30004-X.

Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, Smeeth L (2015) Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* **44**: 827–836, doi:10.1093/ije/dyv098.

Higashi T, Wenger NS, Adams JL, Fung C, Roland M, McGlynn EA, Reeves D, Asch SM, Kerr EA, Shekelle PG (2007) Relationship between Number of Medical Conditions and Quality of Care. *N Engl J Med* **356**: 2496–2504, doi:10.1056/NEJMsa066253.

Hiom SC (2015) Diagnosing cancer earlier: reviewing the evidence for improving cancer survival. *Br J Cancer* **112**: S1–S5, doi:10.1038/bjc.2015.23.

Hippisley-Cox J, Coupland C (2011a) Identifying patients with suspected gastro-oesophageal cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract* **61**: e707-14, doi:10.3399/bjgp11X606609.

Hippisley-Cox J, Coupland C (2011b) Identifying patients with suspected lung cancer in primary care : derivation and validation of an algorithm. *Br J Gen Pract* **61**: 715–723, doi:10.3399/bjgp11X606627.e715.

Hippisley-Cox J, Coupland C (2012a) Identifying patients with suspected colorectal cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract* **62**: e29-37, doi:10.3399/bjgp12X616355.

Hippisley-Cox J, Coupland C (2012b) Identifying patients with suspected pancreatic cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract* **62**: e29-37, doi:10.3399/bjgp12X616355.

Hippisley-Cox J, Coupland C (2012c) Identifying patients with suspected renal tract cancer in primary care : derivation and validation of an algorithm. 251–260, doi:10.3399/bjgp12X636074.e251.

Hippisley-Cox J, Coupland C (2012d) Identifying women with suspected ovarian cancer in

primary care: derivation and validation of algorithm. *BMJ* **344**: d8009–d8009, doi:10.1136/bmj.d8009.

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. *Br J Gen Pract* **63**: 1–10, doi:10.3399/bjgp13X660724.

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. *Br J Gen Pract* **63**: 11–21, doi:10.3399/bjgp13X660733.

Hippisley-Cox J, Stables D, Pringle M (2004) QRESEARCH: a new general practice database for research. *Inform Prim Care* **12**: 49–50.

Holtedahl K, Vedsted P, Borgquist L, Donker GA, Buntinx F, Weller D, Braaten T, Hjertholm P, Månsson J, Strandberg EL, Campbell C, Ellegaard L, Parajuli R (2017) Abdominal symptoms in general practice: Frequency, cancer suspicions raised, and actions taken by GPs in six European countries. Cohort study with prospective registration of cancer. *Heliyon* **3**: e00328, doi:10.1016/j.heliyon.2017.e00328.

Howell DA, Warburton F, Ramirez A-J, Roman E, Smith AG, Forbes LJL (2015) Risk factors and time to symptomatic presentation in leukaemia, lymphoma and myeloma. *Br J Cancer* **113**: 1114–1120, doi:10.1038/bjc.2015.311.

Hripcsak G, Albers DJ (2013) Next-generation phenotyping of electronic health records. *J Am Med Informatics Assoc* **20**: 117–121, doi:10.1136/amiajnl-2012-001145.

Huggenberger IK, Andersen JS (2015) Predictive value of the official cancer alarm symptoms in general practice--a systematic review. *Dan Med J* **62**: 1–9.

Hughes DL, Neal RD, Lyratzopoulos G, Rubin G (2015) Profiling for primary-care presentation, investigation and referral for liver cancers. *Eur J Gastroenterol Hepatol* **28**: 1, doi:10.1097/MEG.0000000000000555.

IKNL [Netherlands Comprehensive Cancer Organisation] (2012) Breast cancer guideline.

Independent Cancer Taskforce (2015) Achieving World-Class Cancer Outcomes: A strategy for England 2015-2020 (London).

Information Services Division (2017) Detect Cancer Early Staging Data report.

Ingeman ML, Christensen MB, Bro F, Knudsen ST, Vedsted P (2015) The Danish cancer pathway for patients with serious non-specific symptoms and signs of cancer-a cross-sectional study of patient characteristics and cancer probability. *BMC Cancer* **15**: 421, doi:10.1186/s12885-015-1424-5.

Innos K, Padrik P, Valvere V, Eelma E, Kütner R, Lehtsaar J, Tekkel M (2013) Identifying women at risk for delayed presentation of breast cancer: a cross-sectional study in Estonia. *BMC Public Health* **13**: 947, doi:10.1186/1471-2458-13-947.

Ironmonger L, Ohuma E, Ormiston-Smith N, Gildea C, Thomson CS, Peake MD (2014) An evaluation of the impact of large-scale interventions to raise public awareness of a lung cancer symptom. *Br J Cancer* **112**: 207–216, doi:10.1038/bjc.2014.596.

Janz NK, Becker MH, Anderson LA, Marcoux BC (1989) Interventions to enhance breast self-examination practice: a review. *Public Health Rev* **17**: 89–163.

Jenniskens K, de Groot JAH, Reitsma JB, Moons KGM, Hooft L, Naaktgeboren CA (2017) Overdiagnosis across medical disciplines: a scoping review. *BMJ Open* **7**: e018448, doi:10.1136/bmjopen-2017-018448.

Jensen H, Torring ML, Fenger-Gron M, Olesen F, Overgaard J, Vedsted P (2016) Tumour stage

and implementation of standardised cancer patient pathways: a comparative cohort study. *Br J Gen Pract* **66**: e434–e443, doi:10.3399/bjgp16X684805.

Jensen H, Torring ML, Olesen F, Overgaard J, Fenger-Gron M, Vedsted P, Tørring ML, Olesen F, Overgaard J, Fenger-Grøn M, Vedsted P, Torring ML, Olesen F, Overgaard J, Fenger-Gron M, Vedsted P, Tørring ML, Olesen F, Overgaard J, Fenger-Grøn M, Vedsted P, Torring ML, Olesen F, Overgaard J, Fenger-Gron M, Vedsted P (2015) Diagnostic intervals before and after implementation of cancer patient pathways - a GP survey and registry based comparison of three cohorts of cancer patients. *BMC Cancer* **15**: 308, doi:10.1186/s12885-015-1317-7.

Jensen H, Tørring ML, Olesen F, Overgaard J, Vedsted P (2014) Cancer suspicion in general practice, urgent referral and time to diagnosis: a population-based GP survey and registry study. *BMC Cancer* **14**: 636, doi:10.1186/1471-2407-14-636.

Jones R, Charlton J, Latinovic R, Gulliford MC (2009) Alarm symptoms and identification of non-cancer diagnoses in primary care: Cohort study. *BMJ* **339**: 491–493, doi:10.1136/bmj.b3094.

Jones R, Latinovic R, Charlton J, Gulliford MC (2007) Alarm symptoms in early diagnosis of cancer in primary care: cohort study using General Practice Research Database. *BMJ* **334**: 1040, doi:10.1136/bmj.39171.637106.AE.

Jones R, Rubin GP, Hungin P (2001) Is the two week rule for cancer referrals working? *BMJ* **322**: 1555–1556, doi:10.1136/bmj.322.7302.1555.

Kaku M, Mathew A, Rajan B (2008) Impact of socio-economic factors in delayed reporting and late-stage presentation among patients with cervix cancer in a major cancer hospital in South India. *Asian Pacific J Cancer Prev* **9**: 589–594.

Kaushal A, Ramirez AJ, Warburton F, Forster AS, Linsell L, Burgess C, Tucker L, Omar L, Forbes LJ (2016) 'Promoting Early Presentation' intervention sustains increased breast cancer awareness in older women for three years: A randomized controlled trial. *J Med Screen* **0**: 1–3, doi:10.1177/0969141316667408.

Keane MG, Horsfall L, Rait G, Pereira SP (2014) A case-control study comparing the incidence of early symptoms in pancreatic and biliary tract cancer. *BMJ Open* **4**: e005720–e005720, doi:10.1136/bmjopen-2014-005720.

Keeble S, Abel GA, Saunders CL, McPhail S, Walter FM, Neal RD, Rubin GP, Lyratzopoulos G (2014) Variation in promptness of presentation among 10,297 patients subsequently diagnosed with one of 18 cancers: Evidence from a National Audit of Cancer Diagnosis in Primary Care. *Int J Cancer* **135**: 1220–1228, doi:10.1002/ijc.28763.

Khakbazan Z, Taghipour A, Latifnejad Roudsari R, Mohammadi E, Roudsari RL, Mohammadi E (2014) Help seeking behavior of women with self-discovered breast cancer symptoms: A meta-ethnographic synthesis of patient delay. *PLoS One* **9**: 1–24, doi:10.1371/journal.pone.0110262.

Kocher F, Lunger F, Seeber A, Amann A, Pircher A, Hilbe W, Fiegl M (2016) Incidental Diagnosis of Asymptomatic Non-Small-Cell Lung Cancer: A Registry-Based Analysis. *Clin Lung Cancer* **17**: 62–67, doi:10.1016/j.cllc.2015.08.006.

Koeling R, Tate AR, Carroll JA (2011) Automatically estimating the incidence of symptoms recorded in GP free text notes. In Proceedings of the First International Workshop on Managing Interoperability and Complexity in Health Systems - MIXHS '11, (New York, New York, USA: ACM Press), p. 43.

Koshiaris C, Oke J, Abel L, Nicholson BD, Ramasamy K, Bruel A Van Den, Abel L (2018) Quantifying intervals to diagnosis in myeloma : a systematic review and meta-analysis. 1–10, doi:10.1136/bmjopen-2017-019758.

Kroczek EK, Wieners G, Steffen I, Lindner T, Streitparth F, Hamm B, Maurer MH (2017) Non-traumatic incidental findings in patients undergoing whole-body computed tomography at

initial emergency admission. *Emerg Med J* emermed-2016-205722, doi:10.1136/emermed-2016-205722.

Larsen MB, Hansen RP, Sokolowski I, Vedsted P (2014) Agreement between patient-reported and doctor-reported patient intervals and date of first symptom presentation in cancer diagnosis – A population-based questionnaire study. *Cancer Epidemiol* **38**: 100–105, doi:10.1016/j.canep.2013.10.006.

Lawrenson R, Logie J, Marks C (2006) Risk of colorectal cancer in general practice patients presenting with rectal bleeding, change in bowel habit or anaemia. *Eur J Cancer Care (Engl)* **15**: 267–271, doi:10.1111/j.1365-2354.2005.00637.x.

Leiva A, Esteva M, Llobera J, Macià F, Pita-Fernández S, González-Luján L, Sánchez-Calavera MA, Ramos M (2017) Time to diagnosis and stage of symptomatic colorectal cancer determined by three different sources of information: A population based retrospective study. *Cancer Epidemiol* **47**: 48–55, doi:10.1016/j.canep.2016.10.021.

Lewis L, Marcu A, Whitaker K, Maguire R (2017) Patient factors influencing symptom appraisal and subsequent adjustment to oesophageal cancer: A qualitative interview study. *Eur J Cancer Care (Engl)* e12745, doi:10.1111/ecc.12745.

Liberman AL, Newman-Toker DE (2018) Symptom-Disease Pair Analysis of Diagnostic Error (SPADE): a conceptual framework and methodological approach for unearthing misdiagnosis-related harms using big data. *BMJ Qual Saf* bmjqs-2017-007032, doi:10.1136/bmjqs-2017-007032.

Lim A, Mesher D, Gentry-Maharaj A, Balogun N, Widschwendter M, Jacobs I, Sasieni P, Menon U (2016) Time to diagnosis of Type I or II invasive epithelial ovarian cancers: a multicentre observational study using patient questionnaire and primary care records. *BJOG An Int J Obstet Gynaecol* **123**: 1012–1020, doi:10.1111/1471-0528.13447.

Lim AW, Ramirez AJ, Hamilton W, Sasieni P, Patnick J, Forbes LJ (2014) Delays in diagnosis of young females with symptomatic cervical cancer in England: an interview-based study. *Br J Gen Pract* **64**: e602-10, doi:10.3399/bjgp14X681757.

London Cancer (2017) ACE Vague Symptoms Cluster Project Report (London).

Lumbreras B, Donat L, Hernández-Aguado I (2010) Incidental findings in imaging diagnostic tests: a systematic review. *Br J Radiol* **83**: 276–289, doi:10.1259/bjr/98067945.

Lurie G, Wilkens LR, Thompson PJ, Matsuno RK, Carney ME, Goodman MT (2010) Symptom presentation in invasive ovarian carcinoma by tumor histological type and grade in a multiethnic population: A case analysis. *Gynecol Oncol* **119**: 278–284, doi:10.1016/j.ygyno.2010.05.028.

Lynch BM, Youlden D, Fritschi L, Newman B, Pakenham KI, Leggett B, Owen N, Aitken JF (2008) Self-reported information on the diagnosis of colorectal cancer was reliable but not necessarily valid. *J Clin Epidemiol* **61**: 498–504, doi:10.1016/j.jclinepi.2007.05.018.

Lyratzopoulos G (2014) Markers and measures of timeliness of cancer diagnosis after symptom onset: A conceptual framework and its implications. *Cancer Epidemiol* **38**: 211–213, doi:10.1016/j.canep.2014.03.009.

Lyratzopoulos G, Abel G (2013) Earlier diagnosis of breast cancer: focusing on symptomatic women. *Nat Rev Clin Oncol* **10**: 544, doi:10.1038/nrclinonc.2012.126-c1.

Lyratzopoulos G, Abel GA, McPhail S, Neal RD, Rubin GP (2013a) Gender inequalities in the promptness of diagnosis of bladder and renal cancer after symptomatic presentation: evidence from secondary analysis of an English primary care audit survey. *BMJ Open* **3**: e002861, doi:10.1136/bmjopen-2013-002861.

Lyratzopoulos G, Abel GA, McPhail S, Neal RD, Rubin GP (2013b) Measures of promptness of cancer diagnosis in primary care: secondary analysis of national audit data on patients with 18 common and rarer cancers. *Br J Cancer* **108**: 686–690, doi:10.1038/bjc.2013.1.

Lyratzopoulos G, Neal RD, Barbiere JM, Rubin GP, Abel GA (2012) Variation in number of general practitioner consultations before hospital referral for cancer: Findings from the 2010 National Cancer Patient Experience Survey in England. *Lancet Oncol* **13**: 353–365, doi:10.1016/S1470-2045(12)70041-4.

Lyratzopoulos G, Saunders CL, Abel GA, McPhail S, Neal RD, Wardle J, Rubin GP (2015) The relative length of the patient and the primary care interval in patients with 28 common and rarer cancers. *Br J Cancer* **112**: S35–S40, doi:10.1038/bjc.2015.40.

Lyratzopoulos G, Wardle J, Rubin G (2014) Rethinking diagnostic delay in cancer: how difficult is the diagnosis? *BMJ* **349**: g7400–g7400, doi:10.1136/bmj.g7400.

Macleod U, Mitchell ED, Burgess C, Macdonald S, Ramirez A (2009) Risk factors for delayed presentation and referral of symptomatic cancer: evidence for common cancers. *Br J Cancer* **101 Suppl**: S92–S101, doi:10.1038/sj.bjc.6605398.

Maconi G, Manes G, Porro GB (2008) Role of symptoms in diagnosis and outcome of gastric cancer. *World J Gastroenterol* **14**: 1149–1155.

Malats N, Belloc J, Gallén M, Porta M (1995) Disagreement between hospital medical records and a structured patient interview on the type and date of the first symptom in cancers of the digestive tract. *Rev Epidemiol Sante Publique* **43**: 533–540.

Marcu A, Black G, Vedsted P, Lyratzopoulos G, Whitaker KL (2017) Educational differences in responses to breast cancer symptoms: A qualitative comparative study. *Br J Health Psychol* **22**: 26–41, doi:10.1111/bjhp.12215.

Marcu A, Lyratzopoulos G, Black G, Vedsted P, Whitaker KL (2016) Educational differences in likelihood of attributing breast symptoms to cancer: a vignette-based study. *Psychooncology* **25**: 1191–1197, doi:10.1002/pon.4177.

Marlow LA V, McGregor LM, Nazroo JY, Wardle J (2014) Facilitators and barriers to help-seeking for breast and cervical cancer symptoms: a qualitative study with an ethnically diverse sample in London. *Psychooncology* **23**: 749–757, doi:10.1002/pon.3464.

McCutchan GM, Wood F, Edwards A, Richards R, Brain KE (2015) Influences of cancer symptom knowledge, beliefs and barriers on cancer symptom presentation in relation to socioeconomic deprivation: a systematic review. *BMC Cancer* **15**: 1000, doi:10.1186/s12885-015-1972-8.

McLachlan S, Mansell G, Sanders T, Yardley S, Van Der Windt D, Brindle L, Chew-Graham C, Little P (2015) Symptom perceptions and help-seeking behaviour prior to lung and colorectal cancer diagnoses: A qualitative study. *Fam Pract* **32**: 568–577, doi:10.1093/fampra/cmv048.

Mendonca SC, Abel G., Saunders CL, Wardle J, Lyratzopoulos G (2016) Pre-referral general practitioner consultations and subsequent experience of cancer care: evidence from the English Cancer Patient Experience Survey. *Eur J Cancer Care (Engl)* **25**: 478–490, doi:10.1111/ecc.12353.

Mills K, Emery J, Cheung C, Hall N, Birt L, Walter FM (2014) A qualitative exploration of the use of calendar landmarking instruments in cancer symptom research. *BMC Fam Pract* **15**: 167, doi:10.1186/s12875-014-0167-8.

Moffat J, Bentley A, Ironmonger L, Boughey A, Radford G, Duffy S (2015) The impact of national cancer awareness campaigns for bowel and lung cancer symptoms on sociodemographic inequalities in immediate key symptom awareness and GP attendances. *Br J Cancer* **112**: S14–S21, doi:10.1038/bjc.2015.31.

Mooney SJ (2017) Systems Thinking in Population Health Research and Policy. In Systems Science and Population Health, A.M. El-Sayed, and S. Galea, eds. (Oxford University Press), pp. 49–58.

Moseholm E, Lindhardt B (2017) Patient characteristics and cancer prevalence in the Danish cancer patient pathway for patients with serious non-specific symptoms and signs of cancer-A nationwide, population-based cohort study. *Cancer Epidemiol* doi:10.1016/j.canep.2017.08.003.

Mounce LTA, Price S, Valderas JM, Hamilton W (2017) Comorbid conditions delay diagnosis of colorectal cancer: a cohort study using electronic primary care records. *Br J Cancer* **116**: 1536–1543, doi:10.1038/bjc.2017.127.

Moynihan R, Doust J, Henry D (2012) Preventing overdiagnosis: how to stop harming the healthy. *Bmj* **344**: e3502–e3502, doi:10.1136/bmj.e3502.

Munck A, Damsgaard J, Hansen D, Bjerrum L, Søndergaard J (2003) The Nordic method for quality improvement in general practice. *Qual Prim Care* **11**: 73–78.

Murphy CC, Gerber DE, Pruitt SL (2017) Prevalence of Prior Cancer Among Persons Newly Diagnosed With Cancer. *JAMA Oncol* **75390**: 1–4, doi:10.1001/jamaoncol.2017.3605.

Nadkarni PM, Ohno-Machado L, Chapman WW (2011) Natural language processing: an introduction. *J Am Med Informatics Assoc* **18**: 544–551, doi:10.1136/amiajnl-2011-000464.

Nadpara P, Madhavan SS, Tworek C, International T, Epidemiology C, Nadpara P, Madhavan SS, Tworek C (2015) Guideline-concordant timely lung cancer care and prognosis among elderly patients in the United States: A population-based study. *Cancer Epidemiol* **39**: 1136–1144, doi:10.1016/j.canep.2015.06.005.

Næser E, Fredberg U, Møller H, Vedsted P (2017) Clinical characteristics and risk of serious disease in patients referred to a diagnostic centre: A cohort study. *Cancer Epidemiol* doi:10.1016/j.canep.2017.07.014.

National Academies of Medicine (2015) Improving Diagnosis in Healthcare (Washington, DC: The National Academies Press).

National Cancer Registration and Analysis Service (2017) Routes to Diagnosis 2006-2015 workbook (a).

NCRAS and ONS (2017) Table 1: Number of patient diagnoses and proportion diagnosed at each stage, adults (aged 15 to 99), England, diagnosed in 2015 https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancersurvivalinengland/adultstageatdiagnosisandchildhoodpatientsfollowedupto2016.

Neal RD, Din NU, Hamilton W, Ukoumunne OC, Carter B, Stapley S, Rubin G (2014) Comparison of cancer diagnostic intervals before and after implementation of NICE guidelines: analysis of data from the UK General Practice Research Database. *Br J Cancer* **110**: 584–592, doi:10.1038/bjc.2013.791.

Neal RD, Tharmanathan P, France B, Din NU, Cotton S, Fallon-Ferguson J, Hamilton W, Hendry A, Hendry M, Lewis R, Macleod U, Mitchell ED, Pickett M, Rai T, Shaw K, Stuart N, Tørring ML, 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. *Br J Cancer* **112**: S92–S107, doi:10.1038/bjc.2015.48.

NHS Digital (2016) NHS e-Referral Service Managing Urgent Referrals for Suspected Cancer: Best Practice Guide.

NHS England (2016) Achieving World-Class Cancer Outcomes : Taking the strategy forward.

NHS Scotland (2012) Get Checked Early http://www.getcheckedearly.org/.

NICE (2005) Referral guidelines for suspected cancer (London).

NICE (2015) Suspected cancer: recognition and referral (Cardiff).

Nicholson BD (2017) Detecting cancer in primary care: Where does early diagnosis stop and overdiagnosis begin? *Eur J Cancer Care (Engl)* e12692, doi:10.1111/ecc.12692.

Nielsen N, Vedsted P, Jensen H (2018) Risk of cancer and repeated urgent referral after negative investigation for cancer. *Fam Pract* 1–7, doi:10.1093/fampra/cmx138.

Niksic M, Rachet B, Warburton FG, Forbes LJL (2016) Ethnic differences in cancer symptom awareness and barriers to seeking medical help in England. *Br J Cancer* **115**: 136–144, doi:10.1038/bjc.2016.158.

Niksic M, Rachet B, Warburton FG, Wardle J, Ramirez AJ, Forbes LJL (2015) Cancer symptom awareness and barriers to symptomatic presentation in England—are we clear on cancer? *Br J Cancer* **113**: 533–542, doi:10.1038/bjc.2015.164.

O'Mahony M, McCarthy G, Corcoran P, Hegarty J (2013) Shedding light on women's help seeking behaviour for self discovered breast symptoms. *Eur J Oncol Nurs* **17**: 632–639, doi:10.1016/j.ejon.2013.03.012.

Office for National Statistics (2016) Cancer Registration Statistics, England https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancerregistrationstatisticscancerregistrationstatisticsengland.

Olesen F, Hansen RP, Vedsted P (2009) Delay in diagnosis: the experience in Denmark. *Br J Cancer* **101**: S5–S8, doi:10.1038/sj.bjc.6605383.

Ong ELH, Goldacre R, Hoang U, Sinclair R, Goldacre M (2014) Subsequent primary malignancies in patients with nonmelanoma skin cancer in england: A national record-linkage study. *Cancer Epidemiol Biomarkers Prev* **23**: 490–498, doi:10.1158/1055-9965.EPI-13-0902.

Ouasmani F, Hanchi Z, Haddou Rahou B, Bekkali R, Ahid S, Mesfioui A (2016) Determinants of Patient Delay in Seeking Diagnosis and Treatment among Moroccan Women with Cervical Cancer. *Obstet Gynecol Int* **2016**: 1–9, doi:10.1155/2016/4840762.

Ozawa M, Brennan PM, Zienius K, Kurian KM, Hollingworth W, Weller D, Hamilton W, Grant R, Ben-Shlomo Y (2018) Symptoms in primary care with time to diagnosis of brain tumours. *Fam Pract* 1–8, doi:10.1093/fampra/cmx139.

Öztürk Ç, Fleer J, Hoekstra HJ, Hoekstra-Weebers JEHM (2015) Delay in diagnosis of testicular cancer; A need for awareness programs. *PLoS One* **10**: 1–10, doi:10.1371/journal.pone.0141244.

Pedersen AF, Hansen RP, Vedsted P (2013) Patient Delay in Colorectal Cancer Patients: Associations with Rectal Bleeding and Thoughts about Cancer. *PLoS One* **8**: e69700, doi:10.1371/journal.pone.0069700.

Pérez G, Porta M, Borrell C, Casamitjana M, Bonfill X, Bolibar I, Fernández E (2008) Interval from diagnosis to treatment onset for six major cancers in Catalonia, Spain. *Cancer Detect Prev* **32**: 267–275, doi:10.1016/j.cdp.2008.05.006.

Pham MT, Rajić A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA (2014) A scoping review of scoping reviews: advancing the approach and enhancing the consistency. *Res Synth Methods* **5**: 371–385, doi:10.1002/jrsm.1123.

Pollack CE, Soulos PR, Herrin J, Xu X, Christakis NA, Forman HP, Yu JB, Killelea BK, Wang SY, Gross CP (2017) The Impact of Social Contagion on Physician Adoption of Advanced Imaging Tests in Breast Cancer. *J Natl Cancer Inst* **109**: 1–8, doi:10.1093/jnci/djw330.

Poum A, Promthet S, Duffy SW, Parkin DM (2014) Factors associated with delayed diagnosis of breast cancer in northeast Thailand. *J Epidemiol* **24**: 102–108.

Powe BD, Finnie R (2003) Cancer fatalism: the state of the science. *Cancer Nurs* **26**: 454-65; quiz 466-7.

Power E, Wardle J (2015) Change in public awareness of symptoms and perceived barriers to seeing a doctor following Be Clear on Cancer campaigns in England. *Br J Cancer* **112 Suppl**: S22-6, doi:10.1038/bjc.2015.32.

Prades J, Espinàs J a, Font R, Argimon JM, Borràs JM (2011) Implementing a Cancer Fast-track Programme between primary and specialised care in Catalonia (Spain): a mixed methods study. *Br J Cancer* **105**: 753–759, doi:10.1038/bjc.2011.308.

Price SJ, Shephard EA, Stapley SA, Barraclough K, Hamilton WT (2014) Non-visible versus visible haematuria and bladder cancer risk: a study of electronic records in primary care. *Br J Gen Pract* **64**: e584-9, doi:10.3399/bjgp14X681409.

Price SJ, Stapley SA, Shephard E, Barraclough K, Hamilton WT (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**: e011664, doi:10.1136/bmjopen-2016-011664.

Probst HB, Hussain ZB, Andersen O (2012) Cancer patient pathways in Denmark as a joint effort between bureaucrats, health professionals and politicians-A national Danish project. *Health Policy (New York)* **105**: 65–70, doi:10.1016/j.healthpol.2011.11.001.

Pruitt SL, Harzke AJ, Davidson NO, Schootman M (2013) Do diagnostic and treatment delays for colorectal cancer increase risk of death? *Cancer Causes Control* **24**: 961–977, doi:10.1007/s10552-013-0172-6.

Public Health England (2016a) Be Clear on Cancer http://www.nhs.uk/be-clear-on-cancer/.

Public Health England (2016b) Be Clear on Cancer - Current Campaigns http://www.cancerresearchuk.org/health-professional/early-diagnosis-activities/be-clear-on-cancer.

Public Health England (2017a) 100 years of public health marketing: enduring public health challenges and revolutions in communication https://publichealthengland.exposure.co/100-years-of-public-health-marketing.

Public Health England (2017b) Abdominal Symptoms Regional Pilot https://campaignresources.phe.gov.uk/resources/campaigns/16-be-clear-on-cancer/Abdominal Symptoms Regional Pilot.

Public Health England, Department of Health (2015) Evaluation of the Be Clear on Cancer ovarian cancer awareness campaign (London).

Queenan JA, Gottlieb BH, Feldman-Stewart D, Hall SF, Irish J, Groome PA (2017) Symptom appraisal, help seeking, and lay consultancy for symptoms of head and neck cancer. *Psychooncology* 1–9, doi:10.1002/pon.4458.

Ramirez AJ, Westcombe AM, Burgess CC, Sutton S, Littlejohns P, Richards MA (1999) Factors predicting delayed presentation of symptomatic breast cancer: A systematic review. *Lancet* **353**: 1127–1131, doi:10.1016/S0140-6736(99)02142-X.

Ramsay L (2017) Sharp shoulder pain, getting worse, after a fall. *BMJ* **356**: j571, doi:10.1136/bmj.j571.

Rasmussen S, Søndergaard J, Larsen PV, Balasubramaniam K, Elnegaard S, Svendsen RP, Andersen RS, Pedersen AF, Vedsted P, Jarbøl DE (2014) The Danish Symptom Cohort: Questionnaire and Feasibility in the Nationwide Study on Symptom Experience and Healthcare-Seeking among 100 000 Individuals. *Int J Family Med* **2014**: 187280,

doi:10.1155/2014/187280.

Redaniel MT, Martin RM, Ridd MJ, Wade J, Jeffreys M (2015) Diagnostic Intervals and Its Association with Breast, Prostate, Lung and Colorectal Cancer Survival in England: Historical Cohort Study Using the Clinical Practice Research Datalink. *PLoS One* **10**: e0126608, doi:10.1371/journal.pone.0126608.

Redondo M, Rodrigo I, Pereda T, Funez R, Acebal M, Perea-Milla E, Jimenez E (2009) Prognostic implications of emergency admission and delays in patients with breast cancer. *Support Care Cancer* **17**: 595–599, doi:10.1007/s00520-008-0513-2.

Renzi C, Lyratzopoulos G, Card T, Chu TPC, Macleod U, Rachet B (2016) Do colorectal cancer patients diagnosed as an emergency differ from non-emergency patients in their consultation patterns and symptoms? A longitudinal data-linkage study in England. *Br J Cancer* 1–10, doi:10.1038/bjc.2016.250.

Reynen E, Robson R, Ivory J, Hwee J, Straus SE, Pham B, Tricco AC (2017) A retrospective comparison of systematic reviews with same-topic rapid reviews. *J Clin Epidemiol* doi:10.1016/j.jclinepi.2017.12.001.

Richards MA (2009a) The National Awareness and Early Diagnosis Initiative in England: assembling the evidence. *Br J Cancer* **101**: S1–S4, doi:10.1038/sj.bjc.6605382.

Richards MA (2009b) The size of the prize for earlier diagnosis of cancer in England. *Br J Cancer* **101**: S125–S129, doi:10.1038/sj.bjc.6605402.

Richards MA, Westcombe AM, Love SB, Littlejohns P, Ramirez AJ (1999) Influence of delay on survival in patients with breast cancer: a systematic review. *Lancet* **353**: 1119–1126, doi:10.1016/S0140-6736(99)02143-1.

Robb K, Stubbings S, Ramirez AJ, Macleod U, Austoker J, Waller J, Hiom S, Wardle J (2009) Public awareness of cancer in Britain: a population-based survey of adults. *Br J Cancer* **101**: S18–S23, doi:10.1038/sj.bjc.6605386.

Roberts MM, French K, Duffy J (1984) Breast cancer and breast self-examination: what do Scottish women know? *Soc Sci Med* **18**: 791–797.

Robinson KM, Christensen KB, Ottesen B, Krasnik A (2012) Diagnostic delay, quality of life and patient satisfaction among women diagnosed with endometrial or ovarian cancer: A nationwide Danish study. *Qual Life Res* **21**: 1519–1525, doi:10.1007/s11136-011-0077-3.

Romanoff A, Constant TH, Johnson KM, Guadiamos MC, María A, Vega B, Zunt J, Anderson BO, Vega AMB, Zunt J, Anderson BO (2017) Association of Previous Clinical Breast Examination With Reduced Delays and Earlier-Stage Breast Cancer Diagnosis Among Women in Peru. *JAMA Oncol* **98195**: doi:10.1001/jamaoncol.2017.1023.

Rubin G, Walter F, Emery J, de Wit N (2018) Reimagining the diagnostic pathway for gastrointestinal cancer. *Nat Rev Gastroenterol Hepatol* doi:10.1038/nrgastro.2018.1.

Rubin GP, McPhail S, Elliot K, McPhail S, Royal College of General Practitioners, Royal College of GPs, Rubin GP, McPhail S, Elliot K, McPhail S, Royal College of General Practitioners, GPs RC of (2011) National Audit of Cancer Diagnosis in Primary Care (London).

Rubin GP, Saunders CL, Abel GA, McPhail S, Lyratzopoulos G, Neal RD (2015) Impact of investigations in general practice on timeliness of referral for patients subsequently diagnosed with cancer: analysis of national primary care audit data. *Br J Cancer* **112**: 676–687, doi:10.1038/bjc.2014.634.

Rutherford MJ, Hinchliffe SR, Abel GA, Lyratzopoulos G, Lambert PC, Greenberg DC (2013) How much of the deprivation gap in cancer survival can be explained by variation in stage at diagnosis: an example from breast cancer in the East of England. *Int J Cancer* **133**: 2192–2200,

doi:10.1002/ijc.28221.

Ryerson AB, Eheman C, Burton J, McCall N, Blackman D, Subramanian S, Richardson LC (2007) Symptoms, Diagnoses, and Time to Key Diagnostic Procedures Among Older U.S. Women With Ovarian Cancer. *Obstet Gynecol* **109**: 1053–1061, doi:10.1097/01.AOG.0000260392.70365.5e.

Salika T, Lyratzopoulos G, Whitaker KL, Waller J, Renzi C (2017) Do comorbidities influence help-seeking for cancer alarm symptoms? A population-based survey in England. *J Public Health (Bangkok)* 1–10, doi:10.1093/pubmed/fdx072.

Savkov A, Carroll J, Koeling R, Cassell J (2016) Annotating patient clinical records with syntactic chunks and named entities: the Harvey Corpus. *Lang Resour Eval* **50**: 523–548, doi:10.1007/s10579-015-9330-7.

Scott SE, Walter FM, Webster A, Sutton S, Emery J (2013) The Model of Pathways to Treatment: Conceptualization and integration with existing theory. *Br J Health Psychol* **18**: 45–65, doi:10.1111/j.2044-8287.2012.02077.x.

Scottish Government (2012) Detect Cancer Early http://www.gov.scot/Topics/Health/Services/Cancer/Detect-Cancer-Early.

Shephard EA, Neal RD, Rose P, Walter FM, Litt EJ, Hamilton WT (2015a) Quantifying the risk of multiple myeloma from symptoms reported in primary care patients: a large case-control study using electronic records. *Br J Gen Pract* **65**: e106–e113, doi:10.3399/bjgp15X683545.

Shephard EA, Neal RD, Rose PW, Walter FM, Hamilton W (2016) Symptoms of adult chronic and acute leukaemia before diagnosis: large primary care case-control studies using electronic records. *Br J Gen Pract* **66**: e182–e188, doi:10.3399/bjgp16X683989.

Shephard EA, Neal RD, Rose PW, Walter FM, Hamilton WT (2015b) Quantifying the risk of Hodgkin lymphoma in symptomatic primary care patients aged >=40 years: a case-control study using electronic records. *Br J Gen Pract* **65**: e289–e294, doi:10.3399/bjgp15X684805.

Shephard EA, Stapley S, Neal RD, Rose P, Walter FM, Hamilton WT (2012) Clinical features of bladder cancer in primary care. *Br J Gen Pract* **62**: 598–604, doi:10.3399/bjgp12X654560.

Shim J, Brindle L, Simon M, George S (2014) A systematic review of symptomatic diagnosis of lung cancer. *Fam Pract* **31**: 137–148, doi:10.1093/fampra/cmt076.

Singh H, Sittig DF (2015) Advancing the science of measurement of diagnostic errors in healthcare: the Safer Dx framework. *BMJ Qual Saf* **24**: 103–110, doi:10.1136/bmjqs-2014-003675.

Smith LK, Pope C, Botha JL (2005) Patients' help-seeking experiences and delay in cancer presentation: A qualitative synthesis. *Lancet* **366**: 825–831, doi:10.1016/S0140-6736(05)67030-4.

Smith TJ, Hillner BE (2011) Bending the Cost Curve in Cancer Care. *N Engl J Med* **364**: 2060–2065, doi:10.1056/NEJMsb1013826.

Stapley S, Peters TJ, Neal RD, Rose PW, Walter FM, Hamilton W (2012) The risk of pancreatic cancer in symptomatic patients in primary care: a large case-control study using electronic records. *Br J Cancer* **106**: 1940–1944, doi:10.1038/bjc.2012.190.

Stapley S, Peters TJ, Neal RD, Rose PW, Walter FM, Hamilton W (2013) The risk of oesophago-gastric cancer in symptomatic patients in primary care: a large case-control study using electronic records. *Br J Cancer* **108**: 25–31, doi:10.1038/bjc.2012.551.

Stapley S, Peters TJ, 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. *Br J Cancer* **95**: 1321–1325, doi:10.1038/sj.bjc.6603439.

Stapley SA, Rubin GP, Alsina D, Shephard EA, Rutter MD, Hamilton WT (2017) Clinical features of bowel disease in patients aged <50 years in primary care: a large case-control study. *Br J Gen Pract* **67**: e336–e344, doi:10.3399/bjgp17X690425.

Stephens MR, Lewis WG, White S, Blackshaw GRJC, Edwards P, Barry JD, Allison MC (2005) Prognostic significance of alarm symptoms in patients with gastric cancer. *Br J Surg* **92**: 840–846, doi:10.1002/bjs.4984.

Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR (2009) Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* **338**: b2393–b2393, doi:10.1136/bmj.b2393.

Stolper E, Van de Wiel M, Van Royen P, Van Bokhoven M, Van der Weijden T, Dinant GJ (2011) Gut Feelings as a Third Track in General Practitioners' Diagnostic Reasoning. *J Gen Intern Med* **26**: 197–203, doi:10.1007/s11606-010-1524-5.

Swann R, McPhail S, Witt J, Shand B, Abel GA, Hiom S, Rashbass J, Lyratzopoulos G, Rubin G (2018) Diagnosing cancer in primary care: results from the National Cancer Diagnosis Audit. *Br J Gen Pract* **68**: e63–e72, doi:10.3399/bjgp17X694169.

The Royal College of Radiologists (2011) Management of Incidental Findings Detected During Research Imaging.

Thormann A, Sørensen PS, Koch-Henriksen N, Laursen B, Magyari M (2017) Comorbidity in multiple sclerosis is associated with diagnostic delays and increased mortality. *Neurology* doi:10.1212/WNL.0000000000004508.

Tørring ML, Frydenberg M, Hansen RP, Olesen F, Hamilton W, Vedsted P (2011) Time to diagnosis and mortality in colorectal cancer: a cohort study in primary care. *Br J Cancer* **104**: 934–940, doi:10.1038/bjc.2011.60.

Tørring ML, Murchie P, Hamilton W, Vedsted P, Esteva M, Lautrup M, Winget M, Rubin G (2017) Evidence of advanced stage colorectal cancer with longer diagnostic intervals: a pooled analysis of seven primary care cohorts comprising 11 720 patients in five countries. *Br J Cancer* 1–10, doi:10.1038/bjc.2017.236.

UCLH (2015) Multidisciplinary diagnosis for patients with abdominal symptoms https://www.uclh.nhs.uk/HP/GPNEWS/Pages/Multidisciplinarydiagnosisforpatientswithabdominalsymptoms.aspx.

Vaismoradi M, Turunen H, Bondas T (2013) Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nurs Heal Sci* **15**: 398–405, doi:10.1111/nhs.12048.

Vedsted P, Olesen F (2015) A differentiated approach to referrals from general practice to support early cancer diagnosis – the Danish three-legged strategy. *Br J Cancer* **112**: S65–S69, doi:10.1038/bjc.2015.44.

Verheij RA, Curcin V, Delaney BC, McGilchrist MM (2018) Possible Sources of Bias in Primary Care Electronic Health Record Data Use and Reuse. *J Med Internet Res* **20**: e185, doi:10.2196/jmir.9134.

Vine MF, Calingaert B, Berchuck A, Schildkraut JM (2003) Characterization of prediagnostic symptoms among primary epithelial ovarian cancer cases and controls. *Gynecol Oncol* **90**: 75–82, doi:10.1016/S0090-8258(03)00175-6.

Vrinten C, McGregor LM, Heinrich M, von Wagner C, Waller J, Wardle J, Black GB (2017) What do people fear about cancer? A systematic review and meta-synthesis of cancer fears in the general population. *Psychooncology* **26**: 1070–1079, doi:10.1002/pon.4287.

Vrinten C, Wardle J, Marlow LA (2016) Cancer fear and fatalism among ethnic minority women

in the United Kingdom. *Br J Cancer* **114**: 597–604, doi:10.1038/bjc.2016.15.

Wagland R, Brindle L, Ewings S, James E, Moore M, Rivas C, Esqueda AI, Corner J (2016) Promoting Help-Seeking in Response to Symptoms amongst Primary Care Patients at High Risk of Lung Cancer: A Mixed Method Study. *PLoS One* **11**: e0165677, doi:10.1371/journal.pone.0165677.

Walker S, Hyde C, Hamilton W (2014) Risk of breast cancer in symptomatic women in primary care: a case-control study using electronic records. *Br J Gen Pract* **64**: e788–e793, doi:10.3399/bjgp14X682873.

Waller J, Robb K, Stubbings S, Ramirez A, Macleod U, Austoker J, Hiom S, Wardle J (2009) Awareness of cancer symptoms and anticipated help seeking among ethnic minority groups in England. *Br J Cancer* **101**: S24–S30, doi:10.1038/sj.bjc.6605387.

Walter F, Webster A, Scott S, Emery J (2012) The Andersen Model of Total Patient Delay: A Systematic Review of Its Application in Cancer Diagnosis. *J Health Serv Res Policy* **17**: 110–118, doi:10.1258/jhsrp.2011.010113.

Walter FM, Birt L, Cavers D, Scott S, Emery J, Burrows N, Cavanagh G, MacKie R, Weller D, Campbell C (2014) 'This isn't what mine looked like': a qualitative study of symptom appraisal and help seeking in people recently diagnosed with melanoma. *BMJ Open* **4**: e005566–e005566, doi:10.1136/bmjopen-2014-005566.

Walter FM, Emery JD, Mendonca S, Hall N, Morris HC, Mills K, Dobson C, Bankhead C, Johnson M, Abel GA, Rutter MD, Hamilton W, Rubin GP (2016a) Symptoms and patient factors associated with longer time to diagnosis for colorectal cancer: results from a prospective cohort study. *Br J Cancer* **115**: 533–541, doi:10.1038/bjc.2016.221.

Walter FM, Mills K, Mendonça SC, Abel GA, Basu B, Carroll N, Ballard S, Lancaster J, Hamilton W, Rubin GP, Emery JD (2016b) Symptoms and patient factors associated with diagnostic intervals for pancreatic cancer (SYMPTOM pancreatic study): a prospective cohort study. *Lancet Gastroenterol Hepatol* **1**: 298–306, doi:10.1016/S2468-1253(16)30079-6.

Walter FM, Rubin G, Bankhead C, Morris HC, Hall N, Mills K, Dobson C, Rintoul RC, Hamilton W, Emery J (2015) Symptoms and other factors associated with time to diagnosis and stage of lung cancer: a prospective cohort study. *Br J Cancer* **112**: S6–S13, doi:10.1038/bjc.2015.30.

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 ML, Johannesen TB, Lawrence G, McGahan CE, Middleton R, Steward J, Tracey E, Turner D, Richards MA, Coleman MP (2013) Breast cancer survival and stage at diagnosis in Australia, Canada, Denmark, Norway, Sweden and the UK, 2000-2007: a population-based study. *Br J Cancer* **108**: 1195–1208, doi:10.1038/bjc.2013.6.

Ward E, Jemal A, Cokkinides V, Singh GK, Cardinez C, Ghafoor A, Thun M (2004) Cancer Disparities by Race/Ethnicity and Socioeconomic Status. *CA Cancer J Clin* **54**: 78–93, doi:10.3322/canjclin.54.2.78.

Webber C, Jiang L, Grunfeld E, Groome PA (2017) Identifying predictors of delayed diagnoses in symptomatic breast cancer: a scoping review. *Eur J Cancer Care (Engl)* **26**: e12483, doi:10.1111/ecc.12483.

Welch HG, Black WC (2010) Overdiagnosis in cancer. *J Natl Cancer Inst* **102**: 605–613, doi:10.1093/jnci/djq099.

Wellcome Collection (2017) Can Graphic Design Save Your Life? (London).

Weller D, Vedsted P, Rubin G, Walter FM, Emery J, Scott S, Campbell C, Andersen RS, Hamilton W, Olesen F, Rose P, Nafees S, van Rijswijk E, Hiom S, Muth C, Beyer M, Neal RD (2012) The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. *Br J Cancer* **106**: 1262–1267, doi:10.1038/bjc.2012.68.

Weyers W (2012) The 'epidemic' of melanoma between under- and overdiagnosis. *J Cutan Pathol* **39**: 9–16, doi:10.1111/j.1600-0560.2011.01831.x.

Whitaker KL, Macleod U, Winstanley K, Scott SE, Wardle J (2015a) Help seeking for cancer 'alarm' symptoms: a qualitative interview study of primary care patients in the UK. *Br J Gen Pract* **65**: e96–e105, doi:10.3399/bjgp15X683533.

Whitaker KL, Scott SE, Wardle J (2015b) Applying symptom appraisal models to understand sociodemographic differences in responses to possible cancer symptoms: a research agenda. *Br J Cancer* **112 Suppl**: S27-34, doi:10.1038/bjc.2015.39.

Whitaker KL, Scott SE, Winstanley K, Macleod U, Wardle J (2014) Attributions of Cancer 'Alarm' Symptoms in a Community Sample. *PLoS One* **9**: e114028, doi:10.1371/journal.pone.0114028.

Whiting PF (2011) QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Ann Intern Med* **155**: 529, doi:10.7326/0003-4819-155-8-201110180-00009.

WHO (2016) International Classification of Primary Care, Second edition (ICPC-2) http://www.who.int/classifications/icd/adaptations/icpc2/en/.

WHO (2017) Guide To Cancer Early Diagnosis.

Wilson JMG, Jungner G (1968) Principles and Practice of Screening for Disease. *World Heal Organ* 163.

World Health Organisation (2017) Cancer Fact sheet http://www.who.int/mediacentre/factsheets/fs297/en/ (accessed: 01/11/2017).

World Health Organization (2002) National Cancer Control Programmes (Geneva).

World Wide Breast Cancer (2016) Know Your Lemons https://www.worldwidebreastcancer.org/about/.

Young JLJ, Roffers SD, Ries LA, Fritz AG, Hurlbut AA (2001) SEER Summary Staging Manual - 2000 Codes and Coding Instructions (Bethesda, MD, MD).

Appendices

Appendices 1 and 2 relate to my research profile and aspects of professional development during this PhD. The remainder of Appendices contain supplementary materials for each corresponding Chapter.

Appendix 1. Academic research profile

A1.1 Publications relating to this thesis

There have been three peer-reviewed publications relating to research presented in this thesis:

- Koo MM, Wagner C von, Abel G, McPhail S, Rubin G, Lyratzopoulos G (2017) **Typical and atypical symptoms in women with breast cancer: Evidence of variation in diagnostic intervals from a national audit of cancer diagnosis.** *Cancer Epidemiology* 48 140–146, http://dx.doi.org/10.1016/j.canep.2017.04.010
  Based on work presented in Chapter 6, see Appendix 5 for full paper
  Online dissemination (metrics by PlumX): 35 Twitter interactions across 3 URLs (6 tweets, 29 retweets)

- Koo MM, Hamilton W, Walter F, Rubin G, Lyratzopoulos G (2018) **Symptom signatures and diagnostic timeliness in cancer patients: a review of current evidence.** *Neoplasia* 20 (2) 165–174, http://dx.doi.org/10.1016/j.neo.2017.11.005
  Based on work presented in Chapter 3, see Appendix 3 for full paper
  Online dissemination (metrics by PlumX): 18 Twitter interactions across 3 URLs (5 tweets, 13 retweets)

- Koo MM, Wagner C von, Abel G, McPhail S, Rubin G, Lyratzopoulos G (2018) **The nature and frequency of abdominal symptoms in cancer patients and their associations with time to help-seeking: evidence from a national audit of cancer diagnosis.** *Journal of Public Health* [Epub ahead of print], https://doi.org/10.1093/pubmed/fdx188
  Based on work presented in Chapter 7, see Appendix 6 for full paper
  Online dissemination (metrics by Altmetric): 29 tweets from 26 users; 42 downloads

A1.2 Research dissemination

A1.2.1 Conference/meeting contributions

Typical and atypical symptoms in women with breast cancer: Evidence of variation in diagnostic intervals from a national audit of cancer diagnosis *(based on work presented in Chapter 6)*

- Oral presentation (15 minutes) at the PHE Cancer Data and Outcomes Conference (CDOC) 13–14 June 2017, Manchester
- Poster presentation at the National Cancer Research Institute (NCRI) conference 6–8 November 2016, Liverpool
- Poster presentation at the Cancer Research UK Early Diagnosis (ED) Research Conference 23–24 February 2017, London

Frequency and nature of presenting abdominal symptoms in primary care before a cancer diagnosis (b*ased on work presented in Chapter 7)*

- Oral presentation (15 minutes) at the PHE Cancer Data and Outcomes Conference (CDOC) 13–14 June 2017, Manchester
- Oral (E-poster, 3 minutes) presentation at the Cancer & Primary Care Research Network (Ca-PRI) conference 18–20 April 2017, Edinburgh
- Poster presentation at the National Cancer Research Institute (NCRI) conference 6–8 November 2016, Liverpool
- Poster presentation at the Cancer Research UK Early Diagnosis (ED) Research Conference 23–24 February 2017, London

The spectrum of presenting symptoms of cancer patients and symptom-specific diagnostic intervals: a review of current evidence to help guide the targeting of early diagnosis initiatives *(based on work presented in Chapter 3)*

- Oral presentation (10 minutes) at the Cancer & Primary Care Research Network (Ca-PRI) conference, 18–20 April 2017, Edinburgh

Common pathways to incidental diagnosis of cancer beyond screening: insights from a national audit of cancer patients in England *(based on work presented in Chapter 5)*

- Oral presentation (15 minutes) at the Preventing Overdiagnosis (POD) conference, 17–19 August 2017, Quebec

A1.2.2 Conference/meeting attendances

PHE BCOC intelligence meeting "What next for Be Clear on Cancer intelligence?"
11th November 2016, London
This was a meeting attended by 40 people including colleagues from Public Health England National Cancer Registration and Analysis Service (PHE-NCRAS), external academics, PHE wellbeing (social marketing) science colleagues, NHS England, and local government.
National Cancer Intelligence Network (NCIN) Cancer Outcomes Conference, 8–10 June 2015, Belfast

3rd National Awareness and Early Diagnosis Initiative (NAEDI) conference, 26–27 March 2015, London (attended before starting PhD)

A1.2.3 Institute 3-minute thesis (3MT) competition

Participation and awarded first place at Institute-level heats (academic year 2016–17)

A1.2.4 Media coverage

Cancer Research UK and NCRI press release, 8th Nov 2016 "One in six women diagnosed with breast cancer have a symptom other than a lump"
Available via: http://www.cancerresearchuk.org/about-us/cancer-news/press-release/2016-11-08-one-in-six-women-diagnosed-with-breast-cancer-have-a-symptom-other-than-a-lump

Interview with Razia Iqbal on BBC World Service radio programme, 8th Nov 2016
Publicly available via: http://bit.ly/2G7JloM

Related online press coverage:
BBC "Warning over non-lump breast cancers" Available via: http://www.bbc.co.uk/news/health-37894360
Huffington Post "Breast cancer symptoms many not include lump and women must learn other signs, researchers warn" Available via: http://www.huffingtonpost.co.uk/entry/women-with-breast-cancer-dont-always-have-lump_uk_5820abeee4b09d57a9a97d3b
Daily Express "Breast cancer symptoms: Women warned to spot PAIN as well as unusual lumps" Available via: https://www.express.co.uk/life-style/health/730057/breast-cancer-symptoms-awareness-signs
The Sun "Can you spot all the signs? 1 in 6 breast cancer patients DON'T have a lump but what are the 6 other signs to watch out for?" https://www.thesun.co.uk/living/2138343/a-lump-is-not-the-only-sign-of-breast-cancer-as-1-in-6-women-suffer-other-symptoms-we-reveal-what-to-watch-out-for/

Appendix 2. Professional development during PhD

A2.1 Other publications published during the PhD (2015–18)

Koo MM, Zhou Y, Lyratzopoulos G (2015) **Delays in diagnosis and treatment of lung cancer: Lessons from US healthcare settings.** *Cancer Epidemiol* 39: 1145–1147, http://dx.doi.org/10.1016/j.canep.2015.08.008.

A2.2 Courses undertaken as part of doctoral training

Statistical Analysis Methods for Epidemiology and Social Sciences
April/May 2015

Stata NetCourse 120: Introduction to Statistical Graphics Using Stata
July 2015

Several academic writing courses
Feb 2016 – May 2017

Farr Short Course: Using Primary Care EHRs for Research
Farr Short Course: Analysing Free Text in Medical Records
Farr Short Course: SQL for Biomedical Researchers
April 2016

Practical use of multiple imputation to handle missing data in Stata course
February 2017

A2.3 Transferrable skills development

Institute student representative (2015–18) and Faculty student representative (2017–18)
Representation of student matters across the Institute of Epidemiology and Health Care since 2015 (75 students), and more recently, the Faculty of Population Health Sciences (seven Institutes; 347 students). Achievements include leading Institute-wide consultation on teaching activities by PhD students which led to the development of guidance regarding payment and regulation of teaching activities for students and staff at the Institute (academic year 2016–17).

Lead student on UCL ChangeMakers project "The A to Z of a PhD: Improving Informational Needs for PGR Students at the Institute of Epidemiology and Health Care (IEHC)" (2017–18)
Awarded an intramural competitive small grant (£840) to evaluate informational needs of PhD students and develop new resources within the Institute of Epidemiology and Health Care.

Institute of Epidemiology and Health Care Early Careers Forum committee member (2015–18)
As a member of the committee, I led the organisation of multiple events and seminars for early career researchers within IEHC. In addition, I developed an online workspace for the committee, and initiated the commissioning and design of a new group logo.

Appendix 3. Appendices relating to Chapter 3

A3.1 Related publication in Neoplasia

Open Access text available at: https://doi.org/10.1016/j.neo.2017.11.005

NEOPLASIA
www.neoplasia.com

Volume 20 Number xx Month 2018 pp. 165–174 165

Symptom Signatures and Diagnostic Timeliness in Cancer Patients: A Review of Current Evidence

Minjoung M. Koo[*], William Hamilton[†], Fiona M. Walter[*], Greg P. Rubin[§] and Georgios Lyratzopoulos[*,‡]

[*]University College London, 1-19 Torrington Place, London WC1E 6BT, UK; [†]University of Exeter Medical School, St Luke's Campus, Heavitree Road, Exeter, EX1 2LU, UK; [‡]University of Cambridge, Primary Care Unit, Strangeways Research Laboratory, Cambridge, CB2 0SR, UK; [§]Institute of Health and Society, Newcastle University, Sir James Spence Institute, Royal Victoria Infirmary, Newcastle upon Tyne, NE1 4LP, UK

Abstract

Early diagnosis is an important aspect of contemporary cancer prevention and control strategies, as the majority of patients are diagnosed following symptomatic presentation. The nature of presenting symptoms can critically influence the length of the diagnostic intervals from symptom onset to presentation (the patient interval), and from first presentation to specialist referral (the primary care interval). Understanding which symptoms are associated with longer diagnostic intervals to help the targeting of early diagnosis initiatives is an area of emerging research. In this Review, we consider the methodological challenges in studying the presenting symptoms and intervals to diagnosis of cancer patients, and summarize current evidence on presenting symptoms associated with a range of In this Review, we consider the methodological challenges in studying the presenting symptoms and intervals to diagnosis of cancer patients, and summarize current evidence on presenting symptoms associated with a range of common and rarer cancer sites. We propose a taxonomy of cancer sites considering their symptom signature and the predictive value of common presenting symptoms. Finally, we consider evidence on associations between symptomatic presentations and intervals to diagnosis before discussing implications for the design, implementation, and evaluation of public health or health system interventions to achieve the earlier detection of cancer.

Neoplasia (2018) 20, 165–174

Introduction

Diagnosing cancer earlier is a critical aim of contemporary cancer control policies. Screening interventions can achieve asymptomatic detection but are currently only available for a limited number of cancer sites, and their effectiveness is further constrained by limited sensitivity and both suboptimal and unequal uptake. This means that the majority of cancer patients continue to be diagnosed following symptomatic presentation, for whom timely diagnosis is associated with better clinical and patient-reported outcomes [1–5]. Diagnosing cancer at an earlier stage is also likely to be cost-effective given the increasing costs of novel drug therapies for advanced stage disease [6]. These considerations highlight the need for efforts aimed at shortening intervals to diagnosis in patients who present with symptoms.

Substantial variation in measures of diagnostic timeliness exists between patients with different cancers [7–10]. Much of this variation has been attributed to the differing nature, frequency, and combinations of presenting symptoms (the 'symptom signature') of each cancer site (as defined in Box 1), though empirical evidence supporting this explanation is sparse. Presenting symptoms can influence the time from symptom onset to first presentation (the patient interval) and the time from first presentation to subsequent referral to specialist care (the primary care interval) [11]. Studying how different symptoms are associated with the length of these two intervals is therefore a priority for early diagnosis research.

Address all correspondence to: Minjoung Monica Koo, Epidemiology of Cancer Healthcare and Outcomes (ECHO) Research Group, Department of Behavioural Science and Health, University College London, 1-19 Torrington Place, London, WC1E 6BT, UK.
E-mail: Monica.koo.14@ucl.ac.uk

Received 22 September 2017; Revised 13 November 2017; Accepted 13 November 2017

1476-5586
https://doi.org/10.1016/j.neo.2017.11.005

A3.2 RECORD-QUADAS Risk of bias tool

Section of paper	Criteria to judge (bold print indicates summary of section)	Comments Low/High/ Unclear
Methods – setting	5. Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection *In other words, could the setting of the study have introduced bias?*	
Methods – Participants	6. (a) Cohort study–Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up Case-control study–Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls Cross-sectional study–Give the eligibility criteria, and the sources and methods of selection of participants --- 6. (b) Cohort study–For matched studies, give matching criteria and number of exposed and unexposed Case-control study–For matched studies, give matching criteria and the number of controls per case 6.1: The methods of study population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. --- Domain 1: patient selection A. Risk of bias 1A1: Was a consecutive or random sample of patients enrolled? (Yes/No/Unclear) 1A3: Did the study avoid inappropriate exclusions? (Yes/No/Unclear) *Could the selection of patients have introduced bias? Also see Results section*	
Data sources/ measurement	8. For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group *Could the measurement of symptoms have introduced bias?*	
Study size	10. Explain how the study size was arrived at *Could the study size have affected the external validity of results? Also see generalisability*	
Data access and cleaning methods	12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population. 12.2: Authors should provide information on the data cleaning methods used in the study.	

	Could data cleaning have introduced bias?	
Results – Participants	13. (a) Report the numbers of individuals at each stage of the study (e.g., numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) 13. (b) Give reasons for non-participation at each stage. 13. (c) Consider use of a flow diagram --- 13.1: Describe in detail the selection of the persons included in the study (i.e., study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram. *Could the selection of patients have introduced bias? Also see Methods section*	
Discussion	19. Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias --- 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported. *Were there other potential sources of bias mentioned by the authors?*	
Generalisability	21. Discuss the generalisability (external validity) of the study results *Could the study size have affected the external validity of results? Also see study size*	

A3.3 Assessed risk of bias of studies included in the literature review

Cancer	Paper	Sample size	Dimensions of risk of bias†					
			Setting?	Study population?	Symptoms?	External validity?	Data cleaning?	Other sources of bias?
Bladder*	Shephard et al., 2012	4915	+	+	+	+	+	+
	Price et al., 2014	4915	+	+	+	+	+	+
	Price et al., 2016[1]	4915	+	+	–	+	+	+
Brain	Hamilton et al., 2007	3505	+	+	+	+	+	+
Breast	Walker et al., 2014	3166	+	+	+	+	+	+
	Redaniel et al., 2015[2]	8544	+	+	?	+	?	+
Cervical	Walker et al., 2017	885	+	+	+	+	+	+
Colorectal	Hamilton et al., 2005a	349	+	+	+	+	+	+
	Stapley et al., 2006	349	+	+	–	+	+	+
	Hamilton et al., 2009a	5477	+	+	+	+	?	+
	Hippisley-Cox & Coupland 2012a	2603	+	–	–	+	+	+
	Collins & Altman 2012a	3712	+	–	–	+	+	+
	Redaniel et al., 2015[2]	5912	+	+	?	+	?	+
	Walter et al., 2016a	152	+	–	+	+	+	+
	Renzi et al., 2016	1606	+	+	+	+	+	+
	Pruitt et al., 2013	9669	+	+	+	?	+	+
Endometrial	Walker et al., 2013	3166	+	+	+	+	+	+
Leukaemia	Shephard et al., 2016[3]	3814	+	+	–	+	+	+
Liver	Hughes et al., 2015	130	+	+	+	+	+	+
Lung	Hamilton et al., 2005b	247	+	+	+	+	+	+
	Hippisley-Cox & Coupland, 2011a	2196	+	–	–	+	+	+
	Ades et al., 2014	247	+	+	+	+	?	+
	Redaniel et al., 2015[2]	5737	+	+	?	+	?	+
	Walter et al., 2015	153	+	–	+	+	+	+
	Nadpara et al., 2015	43833	+	+	?	?	+	+
Lymphoma	Shephard et al., 2015a	283	+	+	+	+	+	+
	Shephard et al., 2015b	4362	+	+	+	+	+	+
Myeloma	Shephard et al., 2015c	2703	+	+	+	+	+	+

Cancer	Paper	Sample size	Dimensions of risk of bias†					
			Setting?	Study population?	Symptoms?	External validity?	Data cleaning?	Other sources of bias?
Oesophago-gastric	Stephens et al., 2005	300	+	–	+	+	+	+
	Hippisley-Cox & Coupland 2011b	781	+	–	–	+	+	+
	Collins & Altman 2012b	287	+	–	–	+	+	+
	Stapley et al., 2013	7471	+	+	+	+	+	+
Ovarian	Ryerson et al., 2007	3250	+	+	+	?	+	+
	Hamilton et al., 2009b	212	+	+	+	+	+	+
	Hippisley-Cox & Coupland 2012b	538	+	–	–	+	+	+
	Collins & Altman 2012c	735	+	–	–	+	+	+
	Lim et al., 2015	182	+	+	+	+	+	+
Pancreatic	Stapley et al., 2012	3635	+	+	+	+	+	+
	Hippisley-Cox & Coupland 2012c	781	+	–	–	+	+	+
	Collins & Altman 2013a	287	+	–	–	+	+	+
	Keane et al., 2014	296	+	+	+	+	+	+
	Walter et al., 2016b	391	+	–	+	+	+	+
	Price et al., 2016[1]	561	+	+	–	+	+	+
Prostate	Hamilton et al., 2006	217	+	+	+	+	+	+
	Redaniel et al., 2015[2]	1763	+	+	?	+	?	+
Renal	Shephard et al., 2013	3149	+	+	+	+	+	+
	Hippisley-Cox & Coupland 2012d[4]	1622	+	–	–	+	+	+
	Collins & Altman 2013b[4]	2283	+	–	–	+	+	+

† '+' denotes low risk of bias; '?' denotes unclear risk of bias; '-' denotes high risk of bias
*the three studies examine the same patient population (n=4915) but with different methodologies
1 includes independent symptom frequencies for bladder cancer and pancreatic cancer patients
2 includes independent symptom frequencies for breast, colorectal, lung, and prostate cancer patients
2 includes patients with acute leukaemia (n=937) + chronic leukaemia (n=2877)
4 reported information on renal tract cancer patients (including bladder cancer)

A3.4 Symptom signature tables

A3.4.2 Cancers with a broad symptom signature, varying predictive value

Cervical cancer

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Walker et al., 2017	Primary care, CPRD data (Read coded)	2000–09	885	40+ years	Post-menopausal bleeding 20.7% Abdominal pain 8.1% Vaginal discharge or vaginitis 7.7% Urinary tract infection 7.6% Haematuria 2.7% Irregular menstruation 2.3% Inter-menstrual bleeding 1.2%

Endometrial cancer

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Walker et al., 2013	Primary care, CPRD data (Read coded)	2000–09	3166	40+ years	General abnormal vaginal bleeding 63% Post-menopausal bleeding 33.9% [1] 12.8% [2] Excessive bleeding 4.0% Irregular menstruation 15.6% [1] 5.6% [2] Vaginal discharge 7.5% Haematuria 4.4% Abdominal pain 5.1% [1] 2.3% [2]

1 frequency based on symptom reported on one visit before diagnosis
2 frequency based on symptom reported on two or more visits before diagnosis

Liver cancer

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Hughes et al., 2015	Primary care, free-text data from NACDPC	2009–10	130 [1]	All ages	Right upper quadrant pain 16% Decompensated liver disease 14% Weight loss 14% Jaundice 11% Epigastric pain 11% Symptoms from metastatic disease[2] 9% Abdominal mass 7% Nausea 6% Anaemia 5% Fatigue 5% Loss of appetite 2% Pruritus 1%

1 symptom frequencies are based on n=88, excluding incidental diagnoses (n=16), rising α-fetoprotein level (n=1), and follow-up for a previously treated colon cancer (n=1), and patients with missing information (n=24)
2 no further information on what these symptoms included

Appendix

Lung cancer

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Hamilton et al., 2005b	Primary care, data from 21 general practices in Exeter	1998–02	247	40+ years	Haemoptysis 20% Weight loss 27% Loss of appetite 19% Dyspnoea 56% Chest or rib pain 42% Fatigue 35% Finger clubbing 4.5% Thrombocytosis 14% Abnormal spirometry 9.7%
Hippisley-Cox & Coupland, 2011a	Primary care, QResearch data (Read coded)	2000–10	2196	30–84 years	Haemoptysis 23.0% [1]
Ades et al., 2014 [2]	Primary care, data from 21 general practices in Exeter	1998–02	247	40+ years	Cough 64.8% Chest pain 40.5%
Redaniel et al., 2015	Primary care, CPRD data (Read coded)	1998–09	5737	15+ years	Haemoptysis 8.8% SVC obstruction 0.4% Stridor 0.1% Anorexia 1.7% Cervical lymphadenopathy 0.5% Chest signs 2.8% Chest/rib pain 14.9% Cough 40.9% Dyspnoea 18.5% Fatigue 4.1% Finger clubbing 0.5% Hoarseness 1.9% Shoulder pain 5.0%
Walter et al., 2015	Primary & secondary care data; self-reported symptoms before diagnosis	2010–12	153	40+ years	Coughing up blood 21.6% Cough or worsening cough >3 weeks 56.2% Breathlessness or worsening breathlessness 41.2% Chest/ shoulder pain 35.3% Hoarseness 12.4% Decreased appetite 22.2% Unexplained weight loss 15% Fatigue or tiredness 45.1% Feeling different “in yourself” 34.6%
Nadpara et al., 2015	SEER-Medicare	2003–06	43,833	66+ years	Cough 14.0% Weight loss 4.8% Dyspnoea 15.6% Chest pain 20.4% Bone pain 2.5% Fever 2.1% Weakness 14.9% SVC obstruction 0.2%

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
					Dysphagia 1.6% Wheezing and stridor 1.1%

SVC: superior vena cava
1 Frequencies of other symptoms included in study were not reported
2 Same study population as Hamilton et al., 2005a; frequencies of additional/different symptoms displayed only

Oesophago-Gastric cancer

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Stephens et al., 2005 [1]	Secondary care, Self-reported and verified with medical records	1995–03	300	17–93 years	Weight loss 44.0% Vomiting 35.7% Anaemia 28.7% Dysphagia 27.7% GI bleed 18.3%
Hippisley-Cox & Coupland, 2011b	Primary care, QResearch data (Read coded)	2000–10	2527	30–84 years	Dysphagia 32.3% Haematemesis 7.5% Abdominal pain 23.0% Appetite loss 2.6% Weight loss 8.0%
Collins & Altman, 2012b [2]	Primary care, THIN data (Read coded)	2000–08	1343	30–84 years	Dysphagia 45.9% Haematemesis 6.2% Abdominal pain 24.7% Appetite loss 2.1% Weight loss 12.3%
Stapley et al., 2013	Primary care, CPRD data (Read coded)	2000–09	7471	40+ years	Dysphagia 32.4% Dyspepsia 17.3% Nausea or vomiting 13.1% Abdominal pain 12.1% [3] Epigastric pain 8.3% Reflux 11.3% Chest pain 9.7% Weight loss 8.2% Constipation 8.1%

GI bleed: gastro-intestinal bleed
1 only described frequency of alarm symptoms before diagnosis of gastric cancer, and not non-alarm symptoms
2 all symptom frequencies calculated manually based on published data; frequencies based on symptoms reported separately for men and women diagnosed with oesophago-gastric cancers
3 specified as all unspecified abdominal pain excluding epigastric pain

Ovarian cancer

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Ryerson et al., 2007[1]	SEER-Medicare	1995–99	2652[1]	65+ years	Abdominal/pelvic swelling 53% Abnormal bleeding 9% Female genital organ pain 19% Abdominal pain/tenderness 60%

Appendix

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
					Constipation, diarrhoea, other digestive disorders 22% Abnormal weight gain/loss 7% Nausea and vomiting 9% Intestinal obstruction 10% Early satiety 2% Abdominal distention (flatulence, gas, pain) 6% Ascites 20% Breathlessness 1% Malaise and fatigue 15% Symptoms of urinary system 15%
Hamilton et al., 2009b	Primary care, Data from 39 general practices in Exeter/ Devon area	2000–07	212	40+ years	Abdominal distension 36% Abdominal bloating 17% Abdominal pain 53% Post-menopausal bleeding 13% Loss of appetite 21% Constipation 20% Diarrhoea 27% Rectal bleeding 8.5% Urinary frequency 14%
Hippisley-Cox & Coupland, 2012b	Primary care, QResearch data (coded data only)	2000–10	538	30–84 years	Abdominal pain 49.4% Post-menopausal bleeding 9.1% Abdominal distension 7.8% Weight loss 4.1% Loss of appetite 2% Rectal bleeding 2%
Collins & Altman, 2012c	Primary care, THIN data (Read coded)	2000–08	735	30–84 years	Abdominal pain 50.5% Post-menopausal bleeding 9.0% Abdominal distension 11.0% Weight loss 4.8% Loss of appetite 1.2% Rectal bleeding 3.3%
Lim et al., 2015[2]	Subsample of women enrolled in UKOPS self-reported symptoms before diagnosis & data from primary care records	2006–08	182 [2]	45+ years	Pelvic or abdominal pain/discomfort 34% 25.1% Increased abdominal size 37.9% 10.8% Bloating 41.4% 7.4% Lump in abdomen 8.9% 5.4% Indigestion 14.3% 7.9% Constipation 13.8% 6.9% diarrhoea 7.4% 3.0% Change in bowel habit 1.0% 6.4% Nausea or vomiting 8.9% 5.9%

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
					Irregular vaginal bleeding 6.9% 5.9% Urinary frequency or urgency 16.7% 10.8% Loss of appetite 13.8% 2.5% Weight loss 13.8% 3.9% Fatigue 24.1% 6.9% Back pain 9.4% 8.9%

UKOPS: UK Ovarian Cancer Population Study
1 symptom frequencies calculated manually based on published data split by LRD staging system, and excluding 598 asymptomatic patients
2 symptom frequencies calculated manually based on published data; frequencies based on patient questionnaire and GP notes respectively
3 number of patients who had invasive epithelial (type I and II) (n=158) and borderline ovarian cancer (n=24)

Pancreatic cancer

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Stapley et al., 2012	Primary care, CPRD data (Read coded)	2000–09	3635	40+ years	Abdominal pain 42.4% Jaundice 30.5% New onset diabetes 22.1% Nausea/vomiting 16.2% Back pain 12.4% Constipation 11.8% Diarrhoea 10.6% Weight loss 9.7% Malaise 5.1%
Hippisley-Cox & Coupland, 2012c	Primary care, QResearch data (Read coded)	2000–10	781	30–84 years	Abdominal pain 39.8% Weight loss 7.8% Appetite loss 3.5% Dysphagia 1.4% Abdominal distension 1.2%
Collins & Altman, 2013a [1]	Primary care, THIN data (Read coded)	2000–08	287	30–84 years	Abdominal pain 57.3% Appetite loss 3.9% Weight loss 13.3%
Keane et al., 2014	Primary care, THIN data (Read coded)	2000–10	296	18+ years	Abdominal pain 44% Back pain 30% Non-cardiac chest pain 13% Shoulder pain 7% Dyspepsia/reflux 26% Nausea and vomiting 20% Abdominal mass 4% Bloating 3% Upper GI bleeding 3% Dysphagia 2% Hepatomegaly 1% Jaundice 35% Pruritus 8% Change in bowel habit 35% Pancreatitis 4% Steatorrhea 1% Weight loss 10% Lethargy 8%

Appendix

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
					Anorexia 5% DVT/PE 4% Insomnia 2% Fracture 1% Change in taste/smell 0.7%
Walter et al., 2016b	Primary & secondary care data; self-reported symptoms before diagnosis	2010–12	391	40+ years	Indigestion 27% Decreased appetite 28% Fatigue 20% Feeling different 21% Change in bowel habit 27% Weight loss 16% Back pain 15% Jaundice 12% Change in urine/stool colour 11%
Price et al., 2016 [2]	Primary care, CPRD data (Read coded & uncoded data)	2000–09	3647	40+ years	Jaundice 42.9% Abdominal pain 49.1%

DVT/PE: deep vein thrombosis/ pulmonary embolism
1 all symptom frequencies calculated manually based on published data; frequencies based on symptoms reported separately for men and women diagnosed with pancreatic cancer
2 majority of patients derived from same study population as Stapley et al., 2012 but uncoded data was used to examine several purposefully selected symptoms

Prostate cancer

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Hamilton et al., 2006	Primary care, data from 21 general practices in Exeter	1998–02	217	40+ years	Retention 15% Hesitancy 17% Impotence 31% Frequency 47% Nocturia 29% Haematuria 15%
Redaniel et al., 2015	Primary care, CPRD data (Read coded)	1998–09	1763	15+ years	Enlarged prostate 20.2% Haematuria 33.6% Hesitancy 4.8% Nocturia 37.2% Poor stream 2.1% Terminal dribbling 2.0%

Renal cancer

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Shephard et al., 2013	Primary care, CPRD data (Read coded)	2000–09	3149	40+ years	Visible haematuria 18% Back pain 11% Abdominal pain 11% Fatigue 7% Constipation 6% Nausea 5%

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
					Lower urinary tract infection 11%
Hippisley-Cox & Coupland, 2012d [1]	Primary care, QResearch data (Read coded)	2000–10	1622	30–84 years	Haematuria 74.0% Abdominal pain 11% Appetite loss 0.4% Weight loss 2.3% Anaemia 4.2%
Collins & Altman, 2013b [1]	Primary care, THIN data (Read coded)	2000–08	2283	30–84 years	Haematuria 72.1% Abdominal pain 11% Appetite loss 0.7% (women only) Weight loss 3.5% (women only) Anaemia 4.5%

1 includes bladder cancer cases

A3.4.3 Cancers with a broad symptom signature, low predictive value

Brain/CNS cancer

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Hamilton et al., 2007	Primary care, CPRD data (Read coded)	1988–06	3505	18+ years	Headache 10.2% Motor loss 8.8% New onset seizure 4.4% Confusion 3.1% Weakness 2.7% Memory loss 1.1% Visual disorder 1.0%

Leukaemia

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Shephard et al., 2016 [1]	Primary care, CPRD data (Read coded)	2000–09	937 [1]	40+ years	Infection 25% [3] Breathlessness 15% Fatigue 12% Chest pain 10% Abdominal pain 10% diarrhoea 7% Vomiting and nausea 6% Bruising 4% [4] Fever 3% Nosebleeds and bleeding gums 3% Flu 2% Weight loss 2%
Shephard et al., 2016 [2]	Primary care, CPRD data (Read coded)	2000–09	2877 [2]	40+ years	Infection 21% [3] Cough 14% Hypertension 14% Breathlessness 7%

Appendix

1 patients with acute leukaemia
2 patients with chronic leukaemia
3 infection consists of urinary tract infection, upper respiratory tract infection, skin infection, and chest infection symptoms

Lymphoma

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Shephard et al., 2015a	Primary care, CPRD data (Read coded)	2000–09	283 [1]	40+ years	Lymphadenopathy 18% Head and neck mass 11% [3] Lump 7%
Shephard et al., 2015b [4]	Primary care, CPRD data (Read coded)	2000–09	4362 [2]	40+ years	Infection 21% Lymphadenopathy 14% Abdominal pain 14% Mass 11% Shortness of breath 9% Head and neck mass 8% Fatigue 7% Constipation 6% Vomiting and nausea 6% Indigestion 5% Weight loss 4% Back pain (re-occurrence) 4% Malaise 4%

1 Hodgkin's Lymphoma
2 non-Hodgkin's Lymphoma
3 includes cervical lymphadenopathy (enlarged neck lymph nodes)
4 infection consists of urinary tract infection, upper respiratory tract infection, skin infection, and chest infection symptoms

Myeloma

Study	Setting/ source of data	Study period	Sample size	Study population age/range	Symptoms
Shephard et al., 2015c	Primary care, CPRD data (Read coded)	2000–09	2703	40+ years	Back pain 28% Chest pain 15% Chest infection 12% Breathlessness 10% Nausea 6% Fracture 6% Joint pain 4% Combined bone pain 4% Weight loss 4% Rib pain 3% Nose bleeds 3%

Appendix 4. Appendices relating to Chapter 4

A4.1 Logic rules used for symptom coding

Logic rule	Explanation and justification, with examples given as appropriate
Unspecified symptoms attributed using knowledge of subsequent cancer diagnosis NB similar logic rules were applied for individual cases (see Appendix 4.2)	
Unspecified bleeding assumed to be a particular type using knowledge of subsequent cancer diagnosis	Given the known symptom signature of each cancer, other sources of bleeding beyond the assumed types are unlikely to have been part of presenting symptoms. E.g. unspecified 'bleed' in: Cervical and endometrial cancer patients assumed to be vaginal bleeding Colorectal cancer patients assumed to be rectal bleeding Gastric cancer patients assumed to be gastro-intestinal bleeding Vulval cancer patients assumed to be vulval bleeding
Unspecified lump assumed to be in a particular location using knowledge of subsequent cancer diagnosis	Given the known symptom signature of each cancer, other lumps beyond the assumed location are unlikely to have been part of presenting symptoms. E.g. unspecified 'lump' in: Breast cancer patients assumed to be a breast lump Testicular cancer patients assumed to be testicular lump Thyroid cancer patients assumed to be a thyroid lump Prostate cancer patients assumed to refer to the prostate
Unspecified pain assumed to be in a particular location using knowledge of subsequent cancer diagnosis	Given the known symptom signature of each cancer, pain in locations other than the assumed body part are unlikely to have been part of presenting symptoms. E.g. unspecified 'pain' in: Breast cancer patients assumed to be breast pain Testicular cancer patients assumed to be testicular pain
Unspecified frequency assumed to be of a particular type, using knowledge of subsequent cancer diagnosis	Given the known symptom signature of each cancer, more common (and therefore more likely) manifestations of 'frequency' were assumed for different cancers. E.g. unspecified 'frequency' in: Colorectal cancer patients assumed to be faecal frequency, and therefore diarrhoea Prostate or bladder cancer patients assumed to be urinary frequency
Unspecified contour of a lump or shape of lump in breast cancer patients assumed to be a breast abnormality	Changes in the contour or shape of a lump located outside the breast is likely to be a very rare manifestation of breast cancer; therefore, it is more likely to be regarding a breast lump.
Unspecified infection in lung cancer patients assumed to be chest infection	Respiratory infections may be part of the symptom signature of lung cancer while other types of infections (e.g. urinary) are not.
Unspecified "urgency" in prostate cancer patients assumed to be urinary urgency rather than bowel urgency	Urinary urgency is a much more common symptom of prostate cancer than bowel urgency.
Unspecified symptoms attributed using information on other symptoms	Given the presence of other symptoms, it is highly likely that the unspecified symptom is of a particular location or type. E.g. unspecified 'itching, bleeding' in a colorectal cancer patient assumed to be rectal abnormalities E.g. unspecified 'pain and bloating' in pancreatic cancer patients assumed to be abdominal pain E.g. unspecified abnormalities that
Unspecified bloating assumed to mean abdominal bloating	Bloating is a term specifically used for abdominal bloating; other uses are unlikely.
Unspecified flooding assumed to mean PMB flooding	The term "flooding" is most commonly associated with post-menopausal bleeding (PMB) in the context of cancer symptoms
Screen-detected patients	
Mention of abnormal FOBT, mammograms, and smears were assumed to be part of the national screening programme even if this was not explicitly mentioned	In the absence of symptom information, it was assumed that such investigations were triggered through cancer screening.

Appendix

Logic rule	Explanation and justification, with examples given as appropriate
Individuals with clinical signs rather than patient elicited symptoms	
Individuals with pleural effusion (PE) were coded under "breathlessness"	This assumes that the doctor's examination of the patient which led to the discovery of PE was prompted by the patient presenting with breathlessness. Of 22 individuals, 4 had pleural effusion as the sole 'symptom'; the remainder had other symptoms, most commonly breathlessness E.g. "Breathlessness due to large pleural effusion" E.g. "SOB–PE went to out of hours"
Mention of an unspecified lump, mass, or swelling found on examination, without any other information were coded as "Lump (NOS)"	This assumes that the doctor's examination of the patient which led to the discovery of a lump, mass, or swelling was prompted by the patient presenting with a lump, mass, or swelling, respectively.

A4.2 Full list of symptom categories and definitions

Symptom frequencies are based on NACDPC patient population before data cleaning (n=18,879); see section 4.5 in Chapter 4. Patients could have more than one assigned symptom category.

Entry No.	Symptom category (frequency)	Definition (examples are provided partially to ensure anonymity of data)
1	Breast lump (n=2337)	Mention of breast lump or mass Includes all lumps described as fungating / hard / indurated / mobile / tender Includes unspecified lump or mass in breast cancer patients Excludes breast swelling or breast oedema unless both are mentioned (both coded as Breast swelling)
2	Lower urinary tract symptoms (LUTS) in men (n=1377)	A composite category created by lumping together the following symptom categories in men: Lower urinary tract symptoms (LUTS) / urinary frequency / voiding urinary symptoms / incontinence/ nocturia / urgency / acute urinary retention Excludes haematospermia (separate category) Individual categories were kept in case of separate analyses and also for use in women; see respective entries for more details on category definition and examples
3	Atypical diagnosis (n=1424)	Mention of incidental diagnosis (explicit or implicit through circumstances) Mention of "no symptoms" / "asymptomatic" with or without further contextual information Mention of an investigation (e.g. blood test, urine analysis, imaging) without further information on symptoms (see Chapter 5)
4	Abdominal pain (n=1054)	Mention of ache, colicky pain, discomfort, or tenderness in the abdominal region Includes abdominal pain in all locations, including left or right iliac fossa (LIF/RIF), left or right upper quadrant (LUQ/RUQ), biliary, subcostal Includes unspecified "pain and bloating" in a pancreatic patient Excludes acute abdominal pain, pelvic pain, groin pain, loin pain & renal colic, and epigastric pain (separate categories)
5	Haematuria (n=942)	Mention of visible blood in urine (macroscopic haematuria) or unspecified haematuria Synonyms: frank haematuria, gross haematuria, visible haematuria Includes unspecified "blood and urine" in bladder cancer patient Excludes microscopic haematuria (included in Atypical category) Assumes unspecified haematuria was macroscopic
6	Per rectal (PR) bleed (chronic or not otherwise specified) (n=798)	Includes all mention of blood in stool or rectal bleeding Includes mention of unspecified bleeding in colorectal cancer patients Excludes acute PR bleeds (separate category)
7	Weight loss (n=797)	Mention of weight loss assumed to be abnormal or unexpected Includes mention of early satiety, e.g. "off food" Excludes cachexia (separate category)
8	Cough (n=730)	Mention of a cough (productive or unproductive) Excludes haemoptysis (separate category)
9	Breathlessness (n=724)	Mention of breathless (dyspnoea) Synonym: shortness of breath (SOB) Includes breathlessness on mild exertion or exertion Includes tightness or pressure in the chest Includes pleural effusion (PE)
10	Invalid (n=631)	No mention of a specific symptom or any other details that could enable classification; includes blank entries in 392 patients Excludes patient records that had sufficient details to ascertain incidental nature, see Atypical diagnosis
11	Abnormal prostate specific antigen (PSA) test (n=627)	Mention of PSA screening or PSA test results Includes unspecified 'abnormal blood tests' in prostate cancer patients Includes mention of an abnormal or high PSA result, or specific PSA values (assuming they are abnormal/borderline)

Appendix

Entry No.	Symptom category (frequency)	Definition (examples are provided partially to ensure anonymity of data)
12	Mole or pigmentation abnormalities (n=542)	Mention of an abnormal mole, naevus, or pigmented/dark lesion Includes abnormalities such as growing, enlarging, bleeding, itchy, or unspecified 'change' Includes mention of an unspecified abnormality in melanoma patients Excludes specific mention of ulceration, if mole is not mentioned (separate category) Excludes other skin abnormalities such as unspecified or non-pigmented skin lesions, rashes, spots, warts, lumps (separate category) Excludes unspecified skin lesions in breast cancer patients; penis; or vulva (separate categories)
13	Fatigue & general weakness (n=482)	Mention of fatigue, lethargy, tiredness, or weakness (generalised rather than specific) Includes fatigue (chronic, extreme) and exhaustion Excludes specific unilateral weakness (separate category)
14	Anaemia (n=461)	Mention of anaemia (severe or otherwise, macrocytic or microcytic) Cases that are explicitly incidental were coded as an atypical diagnosis
15	Change in bowel habit (CIBH) (n=454)	Mention of a change in, or abnormal, bowel habit If constipation or diarrhoea were specified alongside CIBH, these were not coded, assuming that this was part of the CIBH symptom Excludes mention of constipation alone, or diarrhoea alone (separate categories)
16	Voiding urinary symptoms (n=444)	Mention of any voiding and obstructive urinary symptoms Includes unspecified or specified chronic retention Excludes acute urinary retention (separate category) Includes obstruction / hesitancy/ difficulty passing urine / dribbling / poor stream or flow
17	Dysphagia (n=417)	Mention of difficulty swallowing (dysphagia) Includes "dysphapia" and "dysphasia" in Oesophageal or Stomach cancer patients (n=4) assuming typographical errors Includes sensation of a lump in throat (globus) Excludes pain on swallowing (odynophagia) (separate category, not mutually exclusive) Excludes choking (coded as breathlessness)
18	Back pain (n=414)	Mention of back pain (sciatica) Includes various locations: lower back / lumbar / thoracic / spine / lumbosacral /sacroiliac / sacral / interscapular Includes different types of pain: acute or chronic, and ache or pain
19	Nocturia (n=393)	Mention of urge to urinate in the night (nocturia) Includes nocturnal enuresis (instead of coding as urinary incontinence)
20	Urinary frequency (n=390)	Mention of frequency of urination Includes polyuria and diuresis Includes unspecified "frequency" in prostate and bladder cancer patients
21	Diarrhoea (n=383)	Mention of diarrhoea, or frequent/loose motions Includes unspecified "increased frequency" in colorectal cancer patients (assumed to be faecal) Excludes mention of diarrhoea alongside change in bowel habit (separate category, see Change in bowel habit) Excludes faecal or bowel urgency which doesn't state bowel opening (separate category – see Tenesmus)
22	Explicit lower urinary tract symptoms (LUTS) (n=369)	Mention of "LUTS" or "prostatism" verbatim Includes long term or gradual onset LUTS Excludes haematospermia (separate category)
23	Neck lump or lymphadenopathy (n=349)	Mention of a lump(s), mass, or swelling in the neck or mention of neck lymph nodes (LN) Includes cervical lymph nodes or lymphadenopathy Includes supraclavicular lymphadenopathy/ supraclavicular node/supraclavicular mass
24	Musculoskeletal pain (n=345)	Any mention of pain in muscle (myalgia), bone pain, joint pain (arthralgia), or generalised aches and pains Includes joint inflammation (arthritis), stiffness, cramps

Entry No.	Symptom category (frequency)	Definition (examples are provided partially to ensure anonymity of data)
25	Other skin abnormalities (n=338)	Mention of abnormal skin changes that cannot be classified as a "Mole or pigmented lesion abnormality" or "Ulceration" Includes unspecified or non-pigmented skin lesions Included rashes, spots, warts unless unspecified unless in breast cancer patients, penis, or vulva (separate categories – see Breast abnormalities; Penis abnormalities; and Vuval abnormalities respectively) Included lumps and itches if they were described as being localised/small, see [1] below Excludes ulcers or non-healing skin lesions (separate category) Excludes petechial or purpuric rashes (separate category – see Bruising)
26	Post-menopausal bleeding (PMB) (n=315)	Mention of post-menopausal bleeding or spotting Includes peri-menopausal bleeding Includes mention of flooding (unspecified) PMB flooding; this is a term specifically used for PMB Excludes post-coital bleeding (PCB), per vaginal (PV) bleeding, and Menstrual abnormalities (separate categories, mutually exclusive except where both are explicitly mentioned)
27	Chest pain (n=314)	Mention of pain in the chest Includes all areas of chest: chest, chest wall, retrosternal, retrosternal pain on eating, pleuritic, left costal pain, left lower chest, posterior chest pain Includes all types of pain: ache, atypical, discomfort, dull pain, sharp pain, acute, vague Includes chest wall pain Includes other related chest symptoms such as "spontaneous pneumothorax" and "rumbling/vibration feeling…when lying down"
28	Dyspepsia and related epigastric symptoms (n=269)	Mention of indigestion (dyspepsia) and other upper gastrointestinal (GI) symptoms Includes burping, belching, 'upper GI symptoms', waterbrash, GI upset, gastritis Excludes hiccups (separate category – see Other) Excludes reflux and regurgitation (separate category – see Reflux) Excludes epigastric pain or discomfort (separate category – see Epigastric pain)
29	Constipation (n=232)	Mention of constipation Excludes mention of constipation alongside change in bowel habit (separate category, see Change in bowel habit)
30	Haemoptysis (n=226)	Mention of haemoptysis aka coughing up blood Includes blood in sputum or blood stained sputum
31	Abnormal nipple (n=193)	Mention of any nipple abnormalities Includes bleeding or discharge from nipple, and changes in shape or skin of nipple
32	Vaginal bleeding (inter-menstrual or not otherwise specified) (n=193)	Explicit mention of inter-menstrual bleeding (IMB) or unspecified per vaginal (PV) bleeding Includes unspecified bleeding in Endometrial and Cervical cancer patients Also included mentions of pink or brown discharge Excludes post-coital bleeding (PCB), post-menopausal bleeding (PMB) and Menstrual abnormalities (separate categories, mutually exclusive except where both are explicitly mentioned)
33	Abdominal bloating (n=183)	Mention of abdominal bloating or swelling Includes both upper and lower regions Includes unspecified bloating, assuming all bloating is abdominal Includes mention of fullness if totally unspecified Includes mention of epigastric or upper GI fullness Excludes mention of ascites or distension (separate categories – see Ascites and Abdominal distension)
34	Epigastric pain (n=182)	Mention of epigastric pain or discomfort Synonyms: upper gastrointestinal (GI) pain Excludes other epigastric symptoms (separate category, see Dyspepsia and related epigastric symptoms)
35	Breast pain (n=181)	Mention of pain, discomfort, or tenderness in the breast Synonyms: mastalgia, mastodynia, mammalgia Includes unspecified pain, discomfort, or tender lumps in breast cancer patients

Appendix

Entry No.	Symptom category (frequency)	Definition (examples are provided partially to ensure anonymity of data)
36	Chest infection (n=181)	Mention of a respiratory tract infection (upper/lower) or unspecified chest infection Includes chest infections that do not resolve or are chronic/persistent/recurrent Includes chest infections with specified details e.g. "bronchitis" / "bronchopneumonia" / "…bronchiolitis and inflammatory lung disease" / "laryngitis" / "…recurrent viral infections" / "Community Acquired pneumonia" / "…complete pneumothorax" / "Lobar Pneumonia" Includes unspecified "infection" in a lung cancer patient Excludes infective exacerbation of COPD or asthma (separate category – see COPD/Asthma)
37	Vomiting (n=179)	Mention of vomiting Excludes regurgitation (separate category – see Reflux) Includes mention of diarrhoea and vomiting (D&V), also coded as Nausea
38	Jaundice (n=171)	Mention of jaundice Including painless jaundice /obstructive jaundice/recurrent obstructive jaundice
39	Malaise (n=170)	Includes malaise and generalised states of unwell Includes cases where this was noticed by doctor Excluded symptoms that changed over time, such as general decline/ general deterioration (separate category – see Other) Examples "general malaise…" / "feeling unwell" / "feeling awful" / "feeling terrible" / "feeling down with recurrent colds"/ "wt feels terrible"
40	Other (n=163)	Any other symptom that was deemed too rare to warrant the creation of a separate category Includes frailty, debility, and mentions of deteriorating or worsening health Includes abnormal smell sensation, alopecia, azoospermia, bowels open at night, disturbed sleep, debility, dental abscess, frailty, gynaecomastia, pelvic dysfunction, radiculopathy, sepsis, toothache, weight gain
41	Anorexia (n=158)	Mention of poor eating (anorexia) Synonyms: loss of appetite (LOA); low appetite Excludes feeling full (separate category – see Dyspepsia and related epigastric symptoms and Abdominal bloating)
42	Hoarseness & voice related symptoms (n=150)	Mention of hoarseness (dysphonia) or other voice related symptoms Includes changes in voice and loss of voice
43	Urinary tract infection (UTI) (n=150)	Mention of a UTI or cystitis Includes suspected cases of UTI e.g. "symptoms of UTI" / "urinary track symptoms" Includes recurrent cystitis, cystitis symptoms
44	Ulceration (n=148)	Mention of an ulcer or a sore Includes lesions, wounds, or sores described as non-healing or fungating Includes ulceration of the breast Includes mouth ulcers (NB also categorised as "Sore/ulcer in mouth") Includes "lesion on penis" assumed to be ulcers, ALSO categorised as "Penis abnormalities" Excludes internal ulcers i.e. stomach ulcers Excludes mention of skin lesions that aren't described as non-healing, unless they are mouth ulcers – assumed all mouth lesions are non-healing
45	Lump not otherwise specified (NOS) (n=147)	Mention of a lump/ bloating/ swelling that doesn't fit in other existing categories Includes cyst/cystic lump/cyst increasing in size Includes unspecified localised swellings, and swelling of areas other than the limbs/extremities e.g. "chest wall tenderness/swelling" Excludes swelling of limps and extremities (separate category – see Oedema) Excludes localised lumps on skin e.g. scalp or toenails (separate category – see Other skin abnormalities) Excludes unspecified "lump" in breast cancer patients (assumed to be breast lump) Excludes unspecified "lump" in prostate cancer patients (assumed to be prostatic lump)

Entry No.	Symptom category (frequency)	Definition (examples are provided partially to ensure anonymity of data)
		Excludes unspecified "lump" in testicular cancer patients (assumed to be testicular lump)
46	Dysuria (n=139)	Mention of pain or discomfort on urinating (dysuria) Includes unspecified pain or discomfort in the urinary tract, assuming this was felt on urinating e.g. "Bladder discomfort" / "non-specific urinary discomfort" / "Urethral pain" Includes other abnormal feeling of pressure on urination, e.g. "… urine frequency and pressure sensation" / "pressure in bladder on urination"
47	Testicular lump or mass (n=139)	Mention of testicular lump or mass Includes unspecified "lump" in testicular cancer patients
48	Nausea (n=123)	Mention of nausea alone Includes mention of "sickness" Excludes mention of vomiting (separate category – see Vomiting)
49	Sore throat (n=118)	Includes discomfort/pain in throat Includes other sensations in the throat, excluding difficulty swallowing (separate category) e.g. "abnormal feeling in throat" Excluded laryngitis (separate category – see Chest infections)
50	Abdominal lump or mass (n=115)	Mention of a lump or mass in abdomen Includes specified locations including left/right iliac fossa (RIF/LIF), left/right upper quadrant (LUQ/RUQ), abdominal wall, epigastric Includes unspecified mass in a colorectal cancer patient
51	Mouth abnormalities (n=102)	Mention of an abnormalities in mouth or tongue Includes sores, ulcers, lumps, swellings, growths and other non-lesional abnormalities Sores and ulcers were also categorised as "Ulceration"
52	Urinary symptoms not otherwise specified (NOS) (n=101)	Mention of unspecified "urinary symptoms" Also includes other urinary symptoms that didn't fit into existing categories (haematuria, dysuria, urinary frequency, urinary incontinence, urine abnormalities, voiding urinary Sx, urinary urgency)
53	Headache (n=93)	Mention of headache Includes specified areas e.g. frontal, occipital Includes type or frequency e.g. dull, severe, increasing
54	Loin pain & renal colic (n=92)	Mention of loin or flank pain Includes mention of renal colic or pain
55	Head and neck symptoms (n=89)	Mention of a lump, swelling, or pain in the head and neck area Includes enlarged tonsil/ tonsillitis/ tonsillar swelling Excludes trigeminal neuralgia (separate category – see Nervous system symptoms)
56	Loss of consciousness (LOC) (n=88)	Mention of any loss of consciousness Synonyms: collapse, acute collapse, fainting, faint event Excluded unspecified "Faint" and sensations of faintness (categorised as Dizziness)
57	Gynaecological abnormalities (n=87)	Mention of gynaecological abnormalities Includes PV discharge Includes cyst/ lump/ mass/ swelling in vagina Includes vulva related symptoms (which are also coded as Vulval symptoms) Includes vaginal pain and painful intercourse (dyspareunia) Excludes any menstrual problems (separate category – see Menstrual abnormalities) Excludes pink/brown vaginal discharge (separate category – see PV bleed)
58	Axillary lump (n=86)	Mention of a lump or swelling in underarm and armpit region (axillary area) Includes specific mention of arm lymphadenopathies and axillary lymph nodes (LN) Excludes axillary rash (separate category – see "Other skin abnormalities")
59	Dizziness (n=85)	Mention of dizziness or lightheaded Synonyms: funny turn, feeling faint, giddiness Excludes actual incidents of faints (separate category – see Loss of consciousness)

Appendix

Entry No.	Symptom category (frequency)	Definition (examples are provided partially to ensure anonymity of data)
60	Anal/rectal abnormalities (n=82)	Mention of anal, perianal, or rectal lump/mass/swelling or other abnormalities Includes unspecified "Itching,bleeding" in a CRC patient Includes any mention of haemorrhoids aka piles or fibroids Includes abnormalities felt on examination (described alongside another symptom) e.g. "groin pain, abnormal DRE" / "Raised PSA with hard irregular mass on PR"
61	Urinary urgency (n=82)	Mention of urgency to urinate Includes unspecified "urgency" if other urinary symptoms were also described e.g. "dysuria, frequency, urgency" / "frequency urgency" / "Poor stream, nocturia, urgency." Includes unspecified "urgency" in prostate cancer patients
62	Screen-detected (n=78)	Mention of any screening that is part of the national screening programmes Synonyms: mammogram, faecal occult blood test (FOBT), or cervical smears Includes unspecified screening tests if they were in patients diagnosed with breast, cervical, or CRC Excludes individuals diagnosed with other cancers (e.g. Renal, Other, endometrial, ovarian) even if a screening programme is mentioned (separate category – see Atypical diagnosis) Excluded screening tests done privately
63	Melaena & gastrointestinal (GI) bleed (n=78)	Mention of black stools (melaena) or GI bleeding Includes unspecified bleeding in stomach cancer patients
64	Pain not otherwise specified (NOS) (n=78)	Mention of unspecified pain, or localised pain that couldn't be categorised elsewhere Includes unspecified discomfort, aches, and pain
65	Oedema (n=78)	Mention of oedema or swelling Includes oedema or swelling in the legs or extremities (ankle, foot, hand/fingers) Excludes breast oedema (separate category – see Breast swelling) Excludes upper limb oedema (separate category – see Oedema of the upper limb)
66	Reflux (n=75)	Mention of reflux Synonyms: hyperacidity, reflux oesophagitis, gastro-oesophageal reflux, heartburn Includes hiatus hernia and regurgitation Excludes dyspepsia or other epigastric symptoms (separate category)
67	Prostate abnormalities (n=74)	Mention of prostate abnormalities Includes unspecified "lump" and rectal lumps in prostate cancer patients Includes mention of abnormal per rectal (PR) examinations or digital rectal examination (DRE)
68	Lymphadenopathy not otherwise specified (NOS) (n=73)	Mention of unspecified lymph node (LN) enlargement Includes lymph node /swollen glands/enlarged gland Excludes inguinal/groin LN (separate category – see Groin lump or lymphadenopathy) Excludes cervical/supraclavicular/neck LN (separate category – see Neck lump or lymphadenopathy) Excludes axillary LN (separate category – see Axillary lump)
69	Groin lump or lymphadenopathy (n=73)	Mention of a lump, swelling, or lymphadenopathy (LN) in the groin area Includes inguinal or groin LN
70	Confusion (n=72)	Mention of confusion Includes acute or increasing confusion
71	Abdominal distension (n=71)	Mention of abdominal distension Excludes ascites (separate category – see Ascites) This category is not mutually exclusive with abdominal bloating
72	Falls (n=71)	Mention of a fall with no other symptoms, or symptoms caused by a fall e.g. "pain in knee following fall" Includes unspecified falls Excludes individuals diagnosed via the investigation of a fall e.g. "found to be in retention when admitted following fall" or "CXR finding after fall" (separate category – see Atypical diagnosis)

Entry No.	Symptom category (frequency)	Definition (examples are provided partially to ensure anonymity of data)
73	Speech & other cognitive abnormalities (n=65)	Mention of any speech or any other cognitive abnormalities Includes difficulty making intelligible speech or understanding speech (expressive or receptive dysphasia) Includes difficulty speaking due to muscular dysfunction (dysarthria) e.g. slurred speech Includes memory issues Includes other cognitive changes
74	Abnormal vision (n=62)	Mention of vision related abnormalities Includes blurred vision, double vision (diplopia), altered vision Includes loss of vision
75	Breast skin abnormalities (n=61)	Mention of any changes in breast skin Includes skin dimpling, thickening, tethering, puckering, indentation, itching Includes skin lesions that aren't described to be fungating (separate category – see Ulceration) Includes generic breast skin changes and unspecified skin changes in breast cancer patients Excludes breast contour abnormalities or other breast abnormalities (separate categories – see Breast contour abnormalities and Breast abnormalities (NOS) respectively)
76	Testicular abnormalities (n=60)	Mention of any testicular abnormalities including Includes pain and synonyms: discomfort, ache, pain, tenderness Includes scrotal, testicular, epididimal areas Includes unspecified "discomfort", "lump", and "pain" in testicular cancer patients
77	Erectile dysfunction (n=57)	Mention of erectile dysfunction
78	Unsteadiness or impaired mobility (n=56)	Mention of unsteadiness or general impairment of mobility Includes vertigo Includes symptoms that seem to describe loss of motor control Includes symptoms that seem to reflect poor mobility due to progressing malignancy E.g. "unsteadiness" / "unsteady on feet" / "feeling wobbly"
79	Unilateral weakness (n=55)	Mention of weakness in one side (hemiparesis/hemiplegia) or in one limb (ataxia) Excludes any mention of numbness (separate category – see PNS symptoms) Excludes mention of general weakness (separate category – see Fatigue & general weakness)
80	COPD/Asthma (n=53)	Mention of COPD or asthma as the only symptom, or mentioned alongside other symptoms unrelated to COPD/asthma Includes cases where a symptom is described alongside COPD, and it is likely to be exacerbation of COPD e.g. "Increasing SOB (COPD pt)" / "Known chronic COPD. Breathing became more difficult" Includes mention of COPD or asthma exacerbation in "lung" or "other" cancer patients but not for other cancers, assuming that asthma exacerbation in any other cancer is indicative of an Atypical diagnosis e.g. "Asthma attack" / "Difficult poorly-controlled asthma" / "infective exacerbation of COPD"
81	Vulval abnormalities (n=53)	Mention of any vulval abnormalities including pain Includes lump, irregularity, mass, pain, soreness, itch, irritation Includes unspecified lesions, cysts, warts, lumps, and pain in vulval cancer patients Includes unspecified "bleeding" in Vulval patients [assumed to be bleeding lesions of the vulva and not PV bleed] e.g. "Lump found and bleeding"
81	Urinary incontinence (n=51)	Mention of incontinence specified to be urinary Includes unspecified incontinence mentioned alongside other urinary symptoms, e.g. "nocturia,leakage" Includes unspecified incontinence in bladder, prostate, and ovarian cancer patients (n=10) Includes bed wetting Excludes nocturnal enuresis which was included in Nocturia
83	Fever (n=49)	Mention of fever or a temperature (pyrexia)

Appendix

Entry No.	Symptom category (frequency)	Definition (examples are provided partially to ensure anonymity of data)
		Synonyms: pyrexia or pyrexia of unknown origin (PUO) Includes rigors
84	Anal/rectal pain (n=47)	Mention of anal, perianal, or rectal pain Includes unspecified pain mentioned in context of rectal bleeding NB this category is not mutually exclusive with Anal/rectal abnormalities
85	Peripheral nervous system (PNS) symptoms (n=47)	Mention of tingling, numbness, and other symptoms associated with the peripheral nervous system (PNS) Includes paraesthesia –tingling or pricking, caused by pressure on or damage to peripheral nerves Includes explicit mentions of neuropathy or neuropathic pain Includes tremors and shaking
86	Acute urinary retention (n=47)	Mention of acute urinary retention Included unspecified "retention" where the diagnosis was prostate or bladder cancer, or if it was in the context of other urinary symptoms Included cases where the acute nature of retention was implicit, e.g. "…then went into retention"
87	Bowel obstruction (n=42)	Mention of bowel obstruction Synonyms: small bowel obstruction, subacute obstruction, intestinal obstruction Includes "pencil-thin stools" in a colorectal cancer patient Includes unspecified "obstruction" in colorectal cancer patients
88	Thyroid lump/swelling (n=42)	Mention of thyroid enlargement, swelling, or lump Synonyms: goitre Includes unspecified "lump" in thyroid cancer patients
89	Haematospermia (n=41)	Mention of blood in semen (haematospermia)
90	Bruising (n=40)	Mention of bruising (skin haematoma) in general Includes petechiae or petechial rashes, and purpuric rashes
91	Pallor (n=40)	Mention of being pale or pallor Assumed to be elicited by the doctor in majority of cases
92	Flatulence (n=39)	Mention of flatulence/ wind
93	Wheeze (n=39)	Mention of a wheeze or wheezing
94	Fits (n=38)	Mention of a fit Synonyms: convulsion, seizure Includes absence seizures, grand mal seizures
95	Groin pain (n=38)	Mention of pain, discomfort, or ache in the groin area
96	Haematemesis (n=37)	Mention of vomiting blood (hematemesis) Synonym: coffee ground vomit and black vomit
97	Night sweats (n=35)	Mention of night sweats Excludes sweating where night-time is not specified (separate category – see Sweating)
98	Menstrual abnormalities (n=35)	Explicit mention of irregularities in the frequency, cycle, or volume of menstrual bleeding Excludes post-coital bleeding (PCB), post-menopausal bleeding (PMB) and per vaginal (PV) bleeding (separate categories, mutually exclusive except where both are explicitly mentioned)
99	Thromboembolism (n=35)	Mention of deep vein thrombosis (DVT) or pulmonary embolism (PE) in isolation, or mentioned alongside other symptoms unrelated to DVT/PE For cases where DVT/PE is mentioned alongside symptoms related to DVT/PE, they were coded as both the symptom and Thromboembolism Excludes relevant DVT/PE symptoms if DVT or PE is not explicitly mentioned e.g. leg swelling (separate categories – see various)
100	Cerebrovascular abnormalities (n=34)	Mention of any cerebrovascular conditions as the sole symptom e.g. "collapsed with brain bleed…" / "…possible stroke…" / "TIA" / "Stroke like symptoms"
101	Cardiovascular abnormalities (n=34)	Mention of heart palpitations and other cardiovascular abnormalities Includes cardiac pain/ bradycardia/tachycardia/ irregular heart beat/ cardiac tamponade/ irregular heart beat
102	Tenesmus (n=33)	Mention of bowel or faecal urgency (tenesmus) Includes unspecified urgency where there was no mention of actual bowel movement

Entry No.	Symptom category (frequency)	Definition (examples are provided partially to ensure anonymity of data)
103	Nasal obstruction (n=32)	Mention of any nasal obstruction Includes sinusitis, rhinitis, and common colds
104	Mucus per rectum (PR) or in stools (n=29)	Mention of mucus per rectum or in stools Includes undefined mucus mentioned alongside other bowel symptoms e.g. "pr bleeding and mucus" Includes undefined discharge from rectum or with bowels
105	Acute abdominal pain (n=28)	Mention of acute or severe abdominal pain Assumes severe pain was acute in nature Excludes right iliac fossa (RIF) pain unless explicitly acute in nature (separate category – see Right iliac fossa pain below) Excludes renal colic (categorised as Loin pain and renal colic)
106	Pelvic pain (n=28)	Mention of pain in the pelvic area Includes dragging sensation or heavy feeling in pelvis Excludes suprapubic pain or discomfort (separate category – see Suprapubic pain)
107	Sweating (n=26)	Mention of sweating or sweats Excludes night sweats (separate category – see Night sweats)
108	Right iliac fossa (RIF) pain (n=26)	Mention of pain or pain or tenderness in the right iliac fossa
109	Odynophagia (n=26)	Mention of pain on swallowing (odynophagia) Not mutually exclusive with dysphagia
110	Post-coital bleeding (PCB) (n=26)	Mention of post-coital bleeding Excludes post-menopausal bleeding (PMB), per vaginal (PV) bleeding, and menstrual abnormalities (separate categories, mutually exclusive except where both are explicitly mentioned)
111	Fractures (n=25)	Mention of fracture or pathological fracture Includes head injury or subdural haematoma following a fall Includes unspecified fractures (assumes that the fracture caused cancer suspicion and diagnosis was not incidental) Excludes explicit cases where cancer diagnosis is the incidental result of investigating a fracture (separate category – see Atypical diagnosis)
112	Breast contour abnormalities (n=24)	Mention of changes in breast contour or shape Synonyms: breast distortion, breast nodularity Includes change in contour or shape of an unspecified lump in breast cancer patients Excludes breast skin abnormalities or other breast abnormalities (separate categories – see Breast skin abnormalities and Breast abnormalities (NOS) respectively)
113	Itch (n=23)	Mention of generalised or unspecified itch (pruritus) Includes itchiness of locations not otherwise categorised such as scalp and limbs Excluded itchy mole/lesion or other specific locations e.g. penis, vulva, breast, or anus (separate categories)
114	Urine abnormalities (n=22)	Mention of any urine abnormalities Includes abnormal smell, abnormal colour e.g. "dark"/ "odd" / "rust" Includes abnormal consistency e.g. "urine" / "cloudy" / "gritty"
115	Penis abnormalities (n=21)	Mention of abnormalities of the penis Including pain and bleeding related symptoms
116	Ear symptoms (n=20)	Mention of any ear related symptom Including hearing loss
117	Renal symptoms (n=20)	Mention of renal failure or deterioration in isolation, or alongside other symptoms unrelated to renal failure/deterioration For cases where renal symptoms are mentioned alongside symptoms related to the renal abnormality, they were coded as both the symptom and 'Renal Symptoms' Excludes individuals who were incidentally diagnosed with cancer through renal investigations (separate category – see Atypical diagnosis)
118	Oedema of upper limb (n=19)	Mention of oedema or swelling in the upper limb
119	Ascites (n=17)	Mention of ascites (an abnormal accumulation of fluid in the abdomen

Appendix

Entry No.	Symptom category (frequency)	Definition (examples are provided partially to ensure anonymity of data)
		Excludes abdominal distension and abdominal bloating (separate categories, unless both were mentioned)
120	Epistaxis (n=17)	Mention of epistaxis Synonym: nose bleeding
121	Mental conditions (n=17)	Mention of mental conditions as the main or only presenting symptom, without it being a clear coexisting condition Includes anxiety, depression, stress Excludes cognitive and behaviour changes (separate category – see Speech & Other cognitive abnormalities) Excludes confusion (separate category – see Confusion)
122	Breast swelling (n=16)	Mention of swelling of the breast Includes unspecified swelling in breast cancer patients Mutually exclusive with breast lump unless both are specified
123	Breast infection/inflammation (n=16)	Mention of localised breast infection or inflammation Synonyms: mastitis Includes cases with breast abscess Includes other clinical signs that are indicative of infection or inflammation, e.g. "pain and erythema" / "Red swollen breast – no lump" / "Sudden onset red hot tender lump…" / "temperature and mastalgia"
124	Faecal incontinence (n=15)	Mention of incontinence specified to be faecal Includes unspecified soiling or incontinence if mentioned with bowel symptoms or in CRC patients e.g. "diarrhoea with incontinence" / "Explosive incontinence"
125	Pelvic mass (n=15)	Mention of mass or swelling in the pelvic area
126	Breast rash (n=13)	Mention of localised rash in the breast Includes unspecified localised rashes in breast cancer patients Excludes breast infection/inflammation (separate category) Excludes breast skin abnormalities or other breast abnormalities (separate categories – see Breast skin abnormalities and Breast abnormalities (NOS) respectively)
127	Axillary pain (n=13)	Mention of pain or discomfort in the axillary area
128	Eye symptom (n=12)	Mention of any eye related symptoms e.g. "red eye"/ "epiphora" / "eye pain"/ "Squint"
129	Prolapse (n=11)	Mention of prolapse of the rectum, haemorrhoids, uterus
130	Breathing abnormalities (n=10)	Breathing abnormalities such as stridor and creps Excluded hoarseness (separate category – see Hoarseness)
131	Steatorrhoea (n=10)	Mention of steatorrhoea (floating faeces due to fat) Includes light coloured or pale stools
132	Thirst/dry mouth (n=10)	Mention of thirst or dry mouth
133	Breast abnormalities (NOS) (n=9)	Mention of breast related abnormalities not categorised as Breast pain; Breast skin abnormalities; Breast swelling' Breast contour abnormalities; Breast infection/inflammation; Breast rash; Breast bruising; Ulceration) Excludes mention of breast cysts (separate category – see Breast lump) Excludes mention of discharge (separate category – see Nipple abnormalities)
134	Hepatomegaly (n=9)	Mention of enlarged or palpable liver (hepatomegaly) Includes cases where this was noticed on examination
135	Nervous system symptoms (n=8)	Mention of nervous system symptoms e.g. "facial nerve palsy"/ "Trigeminal Neuralgia"
136	Suprapubic pain (n=8)	Mention of pain or discomfort in the suprapubic area Excludes mentions of pelvic pain (separate category – see Pelvic pain)
137	Breast bruising (n=7)	Mention of localised bruising in the breast
138	Hot flushes (n=7)	Mention of flushes and hot flushes
139	Cachexia (n=6)	Mention of cachexia (abnormal weight loss associated with malignancy or other disease) Mutually exclusive with Weight Loss category unless both were mentioned
140	Clubbing (n=6)	Includes finger or nail clubbing
141	Hernia (n=6)	Includes hernia/ strangulated hernia Excludes hiatus hernia (separate category – see Reflux)

Entry No.	Symptom category (frequency)	Definition (examples are provided partially to ensure anonymity of data)
142	Acute per rectal (PR) bleed (n=5)	Mention of PR bleed which is acute in nature Defined as acute through explicit description or as inferred from circumstances
143	Colitis (n=5)	Mention of colitis or exacerbation of existing colitis in isolation For cases where colitis is mentioned alongside symptoms associated with colitis, they were coded as both the respective symptom and 'Colitis'
144	Bleeding not otherwise specified (NOS) (n=4)	Mention of bleeding that could not be categorised as any other category Excludes unspecified "bleeding" in ovarian/endometrial cancer patients (assumed to be PV bleed) Excludes unspecified "bleeding" in CRC patients (assumed to be rectal bleeding) Excludes unspecified "bleeding" in stomach cancer patients (assumed to be gastro-intestinal bleeding)
145	Splenomegaly (n=4)	Mention of splenomegaly (enlarged spleen)

Appendix 5. Appendices relating to Chapter 5

A5.1 Including ethnicity as a covariate

In order to explore potential confounding of the association between sex, age, and cancer site and odds of atypical diagnosis of cancer by ethnicity, I ran two multivariate logistic regression models that calculated crude and adjusted odds ratios of atypical diagnosis (versus 'typical symptomatic diagnosis').

The first included sex, age group, and cancer site as covariates and conducted on a sample population with complete (non-missing) information on ethnicity, while the second included ethnicity as an additional covariate (thus both models n=12446). Odds ratio values of the covariates sex, age group, and cancer site were largely unaffected by the additional adjustment for ethnicity, while the odds ratio for ethnicity indicated a null association with incidental status.

[Table presented on next page]

	Model without adjustment for ethnicity (n=12,446)		Model with adjustment for ethnicity (n=12,446)	
	OR(95% CI)	P-value[1]	OR (95% CI)	P-value[1]
Sex				
Men	Ref.	**0.004**	Ref.	**0.005**
Women	0.8 (0.6–0.9)		0.8 (0.6–0.9)	
Age group				
20–49 years	0.4 (0.3–0.5)		0.4 (0.3–0.5)	
50–59 years	0.8 (0.6–1.1)		0.8 (0.6–1.1)	
60–69 years	Ref.	**<0.001**	Ref.	**<0.001**
70–79 years	1.0 (0.8–1.3)		1.0 (0.8–1.3)	
80+ years	1.1 (0.9–1.4)		1.1 (0.9–1.4)	
Ethnicity				
White	-	-	Ref.	0.519
Non-white	-		0.9 (0.6–1.2)	
Cancer site				
Leukaemia	11.1 (8.4–14.8)		11.2 (8.4–14.9)	
Liver	6.4 (3.9–10.6)		6.5 (3.9–10.6)	
Renal	4.3 (3.0–6.1)		4.3 (3.0–6.1)	
Myeloma	3.5 (2.2–5.4)		3.5 (2.3–5.4)	
Thyroid	4.9 (2.7–9.0)		5.0 (2.7–9.0)	
Gallbladder	1.5 (0.5–4.2)		1.5 (0.5–4.3)	
Bladder	1.4 (1.0–2.0)		1.4 (1.0–2.0)	
Lung	1.3 (1.0–1.8)		1.3 (1.0–1.8)	
Lymphoma	1.5 (1.0–2.2)		1.5 (1.0–2.2)	
Vulval	1.6 (0.6–4.5)		1.6 (0.6–4.5)	
Mesothelioma	1.1 (0.4–3.0)		1.1 (0.4–3.0)	
Colorectal	Ref.		Ref.	
Melanoma	1.1 (0.7–1.7)		1.1 (0.7–1.7)	
Stomach	0.8 (0.4–1.5)		0.8 (0.4–1.5)	
Ovarian	0.7 (0.3–1.4)		0.7 (0.3–1.4)	
Breast	1.0 (0.8–1.4)		1.0 (0.8–1.4)	
Pancreatic	0.7 (0.3–1.3)		0.7 (0.3–1.3)	
Endometrial	0.7 (0.4–1.3)		0.7 (0.4–1.4)	
Laryngeal	0.4 (0.1–1.6)		0.4 (0.1–1.6)	
Oropharyngeal	0.5 (0.2–1.4)		0.5 (0.2–1.4)	
Small Intestine	0.5 (0.1–3.7)		0.5 (0.1–3.7)	
Sarcoma	0.6 (0.1–2.3)		0.6 (0.1–2.3)	
Oesophageal	0.3 (0.1–.6)		0.3 (0.1–0.6)	
Brain	0.2 (0.1–.9)		0.2 (0.1–0.9)	
Cervical	0.3 (0.04–2.2)		0.3 (0.04–2.2)	
Testicular	0.3 (0.04–2.2)		0.3 (0.04–2.2)	

1 joint Wald test p-value; bold indicates <0.05

A5.2 Cancer case-mix of atypically diagnosed cancer

The relative proportions of cancer sites among patients with atypically diagnosed cancer are presented in the table below.

Cancer	No of patients	%
Leukaemia	169	20% (17–23%)
Lung	118	14% (12–16%)
Colorectal	111	13% (11–15%)
Breast	89	10% (9–13%)
Renal	65	8% (6–10%)
Bladder	58	7% (5–9%)
Lymphoma	40	5% (3–6%)
Myeloma	39	5% (3–6%)
Melanoma	35	4% (3–6%)
Liver	28	3% (2–5%)
Thyroid	16	2% (1–3%)
Ovarian	14	2% (1–3%)
Endometrial	11	1.3% (0.7–2.3%)
Pancreatic	11	1.3% (0.7–2.3%)
Stomach	11	1.3% (0.7–2.3%)
Oesophageal	8	0.9% (0.5–1.8%)
Gallbladder	5	0.6% (0.3–1.4%)
Oropharyngeal	5	0.6% (0.3–1.4%)
Mesothelioma	4	0.5% (0.2–1.2%)
Vulval	4	0.5% (0.2–1.2%)
Laryngeal	3	0.4% (0.1–1.0%)
Brain	2	0.2% (0.1–0.9%)
Sarcoma	2	0.2% (0.1–0.9%)
Cervical	1	0.1% (0.02–0.7%)
Small Intestine	1	0.1% (0.02–0.7%)
Testicular	1	0.1% (0.02–0.7%)

Appendix 6. Appendices relating to Chapter 6

A6.1 Related publication in Cancer Epidemiology

Open Access text available at: https://doi.org/10.1016/j.canep.2017.04.010

Cancer Epidemiology 48 (2017) 140–146

Contents lists available at ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

Research article

Typical and atypical presenting symptoms of breast cancer and their associations with diagnostic intervals: Evidence from a national audit of cancer diagnosis

Minjoung Monica Koo[a,*], Christian von Wagner[a], Gary A. Abel[b], Sean McPhail[a,c], Greg P. Rubin[d], Georgios Lyratzopoulos[a,c,e]

[a] University College London, 1-19 Torrington Place, London WC1E 6BT, UK
[b] University of Exeter, St Luke's Campus, Heavitree Road, Exeter EX1 2LU, UK
[c] National Cancer Registration and Analysis Service, Public Health England Zone A, 2nd Floor, Skipton House, 80 London Road, London SE1 6LH, UK
[d] School of Medicine, Pharmacy and Health, Durham University, Stockton on Tees TS17 6BH, UK
[e] Cambridge Centre for Health Services Research, University of Cambridge, Cambridge CB2 0SR, UK

ARTICLE INFO

Article history:
Received 7 February 2017
Received in revised form 8 April 2017
Accepted 18 April 2017
Available online xxx

Keywords:
Breast neoplasms
Early detection of cancer
Signs and symptoms
Primary health care
Female
Delayed diagnosis
Early diagnosis

ABSTRACT

Introduction: Most symptomatic women with breast cancer have relatively short diagnostic intervals but a substantial minority experience prolonged journeys to diagnosis. Atypical presentations (with symptoms other than breast lump) may be responsible.
Methods: We examined the presenting symptoms of breast cancer in women using data from a national audit initiative (n = 2316). Symptoms were categorised topographically. We investigated variation in the length of the patient interval (time from symptom onset to presentation) and the primary care interval (time from presentation to specialist referral) across symptom groups using descriptive analyses and quantile regression.
Results: A total of 56 presenting symptoms were described: breast lump was the most frequent (83%) followed by non-lump breast symptoms, (e.g. nipple abnormalities (7%) and breast pain (6%)); and non-breast symptoms (e.g. back pain (1%) and weight loss (0.3%)).
Greater proportions of women with 'non-lump only' and 'both lump and non-lump' symptoms waited 90 days or longer before seeking help compared to those with 'breast lump only' (15% and 20% vs. 7% respectively). Quantile regression indicated that the differences in the patient interval persisted after adjusting for age and ethnicity, but there was little variation in primary care interval for the majority of women.
Conclusions: About 1 in 6 women with breast cancer present with a large spectrum of symptoms other than breast lump. Women who present with non-lump breast symptoms tend to delay seeking help. Further emphasis of breast symptoms other than breast lump in symptom awareness campaigns is warranted.

1. Introduction

Breast lump is the most common presenting symptom among women with breast cancer and has relatively high predictive value for malignancy [1,2]. Consequently, it has long been the focus of public health education campaigns about cancer symptom awareness [3,4]. Although women with breast cancer typically experience short diagnostic intervals compared to other cancer patients, some women continue to experience long diagnostic intervals [2,5–8]. This is concerning as longer intervals to diagnosis have been shown to be associated with lower five-year survival of breast cancer patients, and additionally, a prolonged diagnostic experience may lead to poorer experience of subsequent cancer care [9–11]. Further, inequalities in stage at diagnosis and survival of breast cancer patients have been linked to variation in the length of the patient interval [12–14].

Prior literature exploring reasons for delayed help-seeking suggests that women subsequently diagnosed with breast cancer

* Corresponding author at: Epidemiology of Cancer Healthcare and Outcomes (ECHO) Group, Research Dept. of Behavioural Science and Health, University College London, 1-19 Torrington Place, London WC1E 7HB, UK.
E-mail address: monica.koo14@ucl.ac.uk (M.M. Koo).

http://dx.doi.org/10.1016/j.canep.2017.04.010

A6.2 Missing outcome data

Variation in the odds of missing outcome data was examined using multivariate logistic regression models.

Breast cancer patients who were missing information on the patient and primary care interval and number of pre-referral consultations were more likely to be older (70 years or over), and first present in places other than general practice.

A6.2.1 Patient interval

	Missing		Non-missing		Adjusted Odds Ratio	Joint Wald test p-value
	N	%	N	%		
Total	433	19%	1883	81%	-	
Age group						
15–49 years	82	19%	555	29%	0.8 (0.6–1.1)	
50–69 years	127	29%	654	35%	Ref.	**<0.001**
70+ years	224	52%	674	36%	1.6 (1.2–2.0)	
Ethnicity						
White	404	93%	1746	93%	Ref.	
Non-white	29	7%	137	7%	1.0 (0.6–1.5)	0.853
No of symptoms						
1	386	89%	1724	92%	Ref.	
2 or more	47	11%	159	8%	1.5 (0.8–3.2)	0.240
Symptom group						
Breast lump only	312	72%	1458	77%	Ref.	
Non-lump only	52	12%	210	11%	1.1 (0.7–1.5)	
Both lump and non-lump	26	6%	113	6%	0.6 (0.3–1.5)	
Non-breast symptoms only	36	8%	89	5%	1.3 (0.8–2.1)	0.491
Lump and non-breast symptoms	4	0.9%	8	0.4%	1.4 (0.3–6.1)	
Non-lump and non-breast symptoms	2	0.5%	5	0.3%	0.5 (0.1–4.1)	
Breast lump, non-lump, non-breast symptoms	1	0.2%	0	0.0%	N/A	
Place of presentation						
General Practice	355	82%	1833	97%	Ref.	
Outpatients	17	4%	15	0.8%	5.3 (2.6–10.9)	
A&E	7	2%	5	0.3%	6.7 (2.0–22.0)	**<0.001**
Out of Hours	1	0.2%	1	0.1%	4.9 (0.3–81.4)	
Other	46	11%	25	1.3%	8.3 (5.0–13.8)	
Unknown	7	2%	4	0.2%	9.4 (2.7–33.2)	

A6.2.2 Primary care interval

	Missing		Non-missing		Adjusted Odds Ratio	Joint Wald test p-value
	N	%	N	%		
Total	115	5%	2201	95%	-	
Age group						
15–49 years	18	16%	619	28%	0.7 (0.4–1.4)	
50–69 years	36	31%	745	34%	Ref.	0.224
70+ years	61	53%	837	38%	1.3 (0.7–2.1)	
Ethnicity						
White	105	91%	2045	93%		
Non-white	10	9%	156	7%	1.0 (0.4–2.4)	0.940
No of symptoms						
1	106	92%	2004	91%	Ref.	
2 or more	9	8%	197	9%	1.0 (0.3–3.8)	0.981
Symptom group						
Breast lump only	73	63%	1697	77%	Ref.	
Non-lump only	19	17%	243	11%	1.8 (1.0–3.5)	
Both lump and non-lump	4	3%	135	6%	0.5 (0.1–3.0)	
Non-breast symptoms only	18	16%	107	5%	1.9 (0.9-4.1)	0.116
Lump and non-breast symptoms	0	0.0%	12	0.5%	N/A	
Non-lump and non-breast symptoms	1	0.9%	6	0.3%	0.5 (0.02–11.8)	
Breast lump, non-lump, non-breast symptoms	0	0.0%	1	0.0%	N/A	
Place of presentation						
General Practice	48	42%	2140	97%	Ref.	
Outpatients	19	17%	13	0.6%	54.4 (24.8–119.1)	
A&E	6	5%	6	0.3%	42.9 (11.7–157.8)	**<0.001**
Out of Hours	1	0.9%	1	0.0%	37.4 (2.2–640.8)	
Other	33	29%	38	1.7%	37.0 (20.9–65.5)	
Unknown	8	7%	3	0.1%	122.2 (29.7–503.3)	

A6.2.3 Number of pre-referral consultations

	Missing		Non-missing		Adjusted Odds Ratio	Joint Wald test p-value
	N	%	N	%		
Total	314	14%	2002	86%	-	
Age group						
15–49 years	64	20%	573	29%	0.8 (0.6–1.2)	
50–69 years	99	32%	682	34%	(ref)	**0.021**
70+ years	151	48%	747	37%	1.3 (1.0–1.7)	
Ethnicity						
White	287	91%	1863	93%	Ref.	
Non-white	27	9%	139	7%	1.3 (0.8–2.0)	0.336
No of symptoms						
1	290	92%	1820	91%	Ref.	
2 or more	24	8%	182	9%	0.5 (0.2–1.6)	0.252
Symptom group						
Breast lump only	236	75%	1534	77%	Ref.	
Non-lump only	39	12%	223	11%	1.1 (0.8–1.7)	
Both lump and non-lump	16	5%	123	6%	1.4 (0.4–5.1)	
Non-breast symptoms only	19	6%	106	5%	0.7 (0.4–1.3)	0.716
Lump and non-breast symptoms	2	0.6%	10	0.5%	2.1 (0.3–16.0)	
Non-lump and non-breast symptoms	2	0.6%	5	0.2%	1.8 (0.2–19.2)	
Breast lump, non-lump, non-breast symptoms	0	0.0%	1	0.0%	N/A	
Place of presentation						
General Practice	236	75%	1952	98%	Ref.	
Outpatients	20	6%	12	0.6%	14.0 (6.6–29.7)	
A&E	7	2%	5	0.2%	14.8 (4.4–50.3)	**<0.001**
Out of Hours	1	0.3%	1	0.0%	9.0 (0.5–151.8)	
Other	42	13%	29	1.4%	11.5 (7.0–19.0)	
Unknown	8	3%	3	0.1%	20.8 (5.4–80.5)	

A6.3 Full list of symptoms among women with breast cancer (n=2,316)

	Symptom signature and frequency		Pre-presentation			Post-presentation				
	N	% relative frequency (95% CI)	Patient Interval Median (IQR) 90th	% Patient Interval > 90 days (95th CI)	% missing	Primary Care Interval Median (IQR) 90th	% Primary Care Interval > 90 days (95th CI)	% missing	% 2+ pre-referral consultations	% missing
Breast lump	192	83.0% (81.4–84.5%)	7 (1–27) 75	8% (7–9%)	18%	0 (0–0) 3	1% (1–2%)	4%	6%	13%
Nipple abnormalities	158	6.8% (5.9–7.9%)	17 (2–71) 275	23% (17–31%)	21%	0 (0–1) 7	1% (0.4–5%)	3%	12%	15%
Breast pain	149	6.4% (5.5–7.5%)	10 (3–41) 96	12% (8–19%)	12%	0 (0–3) 34	3% (1–7%)	3%	20%	8%
Breast skin abnormalities	46	2.0% (1.5–2.6%)	13 (1–30) 129	10% (4–24%)	15%	0 (0–1) 3	2% (0.4–12%)	2%	8%	17%
Axillary lump	27	1.2% (0.8–1.7%)	2.5 (0–12) 15	0% (0–15%)	19%	0 (0–14) 34	4% (1–18%)	0%	36%	19%
Breast ulceration	25	1.1% (0.7–1.6%)	122 (0–276) 594	56% (27–81%)	64%	0 (0–1) 1	0% (0–15%)	16%	7%	40%
Back pain	24	1.0% (0.7–1.5%)	9.5 (1–51) 107.5	10% (3–30%)	17%	21 (0–105) 145	26% (13–46%)	4%	65%	4%
Breast contour abnormalities	17	0.7% (0.5–1.2%)	5 (4–18) 184	15% (4–42%)	24%	0 (0–1) 3	0% (0–20%)	12%	7%	18%
Breast infection or inflammation	15	0.6% (0.4–1.1%)	2.5 (0–30) 366	21% (8–48%)	7%	9 (0–23) 37	7% (1–31%)	7%	60%	0%
Breast swelling	14	0.6% (0.4–1.0%)	3.5 (0–14) †	10% (2–40%)	29%	0 (0–3.5) 8	0% (0–24%)	14%	15%	7%
Musculoskeletal pain	14	0.6% (0.4–1.0%)	0.5 (0–22) †	10% (2–40%)	29%	54 (0–187.5) 399	25% (9–53%)	14%	75%	14%
Breathlessness	11	0.5% (0.3–0.8%)	5 (0–35.5) †	0% (0–49%)	64%	1 (0–10.5) †	0% (0–32%)	27%	57%	36%
Breast rash	10	0.4% (0.2–0.8%)	0 (0–16) †	0% (0–39%)	40%	0 (0–7) †	0% (0–32%)	20%	20%	0%
Neck lump or lymph node abnormalities	9	0.4% (0.2–0.7%)	0 (0–10) †	0% (0–39%)	33%	4.5 (0–19.5) †	0% (0–32%)	11%	29%	22%
Abdominal pain	8	0.3% (0.2–0.7%)	39 (18–62) †	17% (3–56%)	25%	3 (2–6) †	0% (0–43%)	38%	71%	13%
Other breast abnormalities	8	0.3% (0.2–0.7%)	6 (0–8) †	0% (0–43%)	38%	0 (0–98) †	33% (10–70%)	25%	14%	13%
Chest pain	8	0.3% (0.2–0.7%)	18 (10–43) †	0% (0–32%)	0%	24 (9.5–83) †	25% (7–59%)	0%	75%	0%
Fatigue or weakness	7	0.3% (0.1–0.6%)	10.5 (1.5–33) †	0% (0–49%)	43%	2 (0–27) †	14% (3–51%)	0%	29%	0%
Weight Loss	6	0.3% (0.1–0.6%)	56 (51–61) †	0% (0–66%)	67%	18 (11–22) †	0% (0–43%)	17%	60%	17%
Cough	6	0.3% (0.1–0.6%)	5.5 (0–11) †	0% (0–66%)	67%	13.5 (6.5–38) †	0% (0–49%)	33%	60%	17%
Axillary pain	5	0.2% (0.1–0.5%)	15 (0–126) †	33% (6–79%)	40%	5 (1–8) †	0% (0–43%)	0%	40%	0%
Breast bruising	5	0.2% (0.1–0.5%)	7 (7–14) †	0% (0–43%)	0%	0 (0–8) †	0% (0–43%)	0%	40%	0%

	Symptom signature and frequency		Pre-presentation			Post-presentation				
	N	% relative frequency (95% CI)	Patient Interval Median (IQR) 90th	% Patient Interval > 90 days (95th CI)	% missing	Primary Care Interval Median (IQR) 90th	% Primary Care Interval > 90 days (95th CI)	% missing	% 2+ pre-referral consultations	% missing
Oedema of upper limb	5	0.2% (0.1–0.5%)	76 (19–133) †	50% (9–91%)	60%	0.5 (0–1) †	0% (0–49%)	20%	0%	0%
Anorexia or loss of appetite	3	0.1% (0.0–0.4%)	11 (11–11) †	0% (0–79%)	67%	39 (17–61) †	0% (0–66%)	33%	50%	33%
Mental conditions	3	0.1% (0.0–0.4%)	13 (1–25) †	0% (0–66%)	33%	7 (7–7) †	0% (0–79%)	67%	.	33%
Other	2	0.1% (0.0–0.3%)	44 (10–78) †	0% (0–66%)	0%	5.5 (4–7) †	0% (0–66%)	0%	50%	0%
Abdominal lump/ mass	2	0.1% (0.0–0.3%)	80.5 (18–143) †	50% (9–91%)	0%	1 (0–2) †	0% (0–66%)	0%	50%	0%
Confusion	2	0.1% (0.0–0.3%)	13 (1–25) †	0% (0–66%)	0%	7 (7–7) †	0% (0–79%)	50%	.	50%
Headache	2	0.1% (0.0–0.3%)	43 (43–43) †	0% (0–79%)	50%	64 (64–64) †	0% (0–79%)	50%	.	50%
Other lymph node abnormalities	2	0.1% (0.0–0.3%)	16 (16–16) †	0% (0–79%)	50%	10.5 (3–18) †	0% (0–66%)	0%	.	0%
Malaise	2	0.1% (0.0–0.3%)	0 (0–0) †	0% (0–79%)	50%	49.5 (1–98) †	50% (9–91%)	0%	.	50%
Vomiting	2	0.1% (0.0–0.3%)	8 (8–8) †	0% (0–79%)	50%	16 (16–16) †	0% (0–79%)	50%	.	50%
Abdominal distension	1	0.04% (0.0–0.2%)	0 (0–0) †	0% (0–79%)	0%	61 (61–61) †	0% (0–79%)	0%	.	0%
Ascites	1	0.04% (0.0–0.2%)	8 (8–8) †	0% (0–79%)	0%	16 (16–16) †	0% (0–79%)	0%	.	0%
Anaemia	1	0.04% (0.0–0.2%)	-	-	100%	-	-	100%	.	0%
Chest infection	1	0.04% (0.0–0.2%)	0 (0–0) †	0% (0–79%)	0%	1 (1–1) †	0% (0–79%)	0%	.	0%
Loss of consciousness	1	0.04% (0.0–0.2%)	0 (0–0) †	0% (0–79%)	0%	0 (0–0) †	0% (0–79%)	0%	.	0%
Constipation	1	0.04% (0.0–0.2%)	18 (18–18) †	0% (0–79%)	0%	2 (2–2) †	0% (0–79%)	0%	.	0%
Cardiovascular abnormalities	1	0.04% (0.0–0.2%)	-	-	100%	0 (0–0) †	0% (0–79%)	0%	.	0%
Diarrhoea	1	0.04% (0.0–0.2%)	-	-	100%	0 (0–0) †	0% (0–79%)	0%	.	0%
Dizziness	1	0.04% (0.0–0.2%)	-	-	100%	0 (0–0) †	0% (0–79%)	0%	.	0%
Epigastric pain	1	0.04% (0.0–0.2%)	3 (3–3) †	0% (0–79%)	0%	28 (28–28) †	0% (0–79%)	0%	.	0%
Falls	1	0.04% (0.0–0.2%)	-	-	100%	0 (0–0) †	0% (0–79%)	0%	.	0%
Fractures	1	0.04% (0.0–0.2%)	0 (0–0) †	0% (0–79%)	0%	0 (0–0) †	0% (0–79%)	0%	.	0%
Groin pain	1	0.04% (0.0–0.2%)	-	-	100%	-	-	100%	.	100%

	Symptom signature and frequency		Pre-presentation			Post-presentation				
	N	% relative frequency (95% CI)	Patient Interval Median (IQR) 90th	% Patient Interval > 90 days (95th CI)	% missing	Primary Care Interval Median (IQR) 90th	% Primary Care Interval > 90 days (95th CI)	% missing	% 2+ pre-referral consult ations	% missing
Haematemesis	1	0.04% (0.0–0.2%)	0 (0–0) †	0% (0–79%)	0%	44 (44–44) †	0% (0–79%)	0%	.	0%
Hoarseness/voice related	1	0.04% (0.0–0.2%)	-	-	100%	21 (21–21) †	0% (0–79%)	0%	.	0%
Vision related symptoms	1	0.04% (0.0–0.2%)	-	-	100%	0 (0–0) †	0% (0–79%)	0%	.	0%
Uncategorised lumps	1	0.04% (0.0–0.2%)	15 (15–15) †	0% (0–79%)	0%	57 (57–57) †	0% (0–79%)	0%	.	100%
Nausea	1	0.04% (0.0–0.2%)	1 (1–1) †	0% (0–79%)	0%	-	-	100%	.	0%
Nervous system symptoms	1	0.04% (0.0–0.2%)	0 (0–0) †	0% (0–79%)	0%	0 (0–0) †	0% (0–79%)	0%	.	0%
Post-coital bleeding	1	0.04% (0.0–0.2%)	0 (0–0) †	0% (0–79%)	0%	14 (14–14) †	0% (0–79%)	0%	.	100%
Speech & other cognitive abnormalities	1	0.04% (0.0–0.2%)	-	-	100%	0 (0–0) †	0% (0–79%)	0%	.	0%
Unsteadiness or impaired mobility	1	0.04% (0.0–0.2%)	-	-	100%	0 (0–0) †	0% (0–79%)	0%	.	0%
Unilateral weakness	1	0.04% (0.0–0.2%)	12 (12–12) †	0% (0–79%)	0%	45 (45–45) †	0% (0–79%)	0%	.	0%
Wheeze	1	0.04% (0.0–0.2%)	0 (0–0) †	0% (0–79%)	0%	0 (0–0) †	0% (0–79%)	0%	.	100%

NB Symptom frequencies do not add up to 100% or n=2,316 as some women had more than 1 symptom.
†90th centile PI and PCI values not shown for symptoms where there were <10 patients with non-missing values
PI: patient interval; PCI: primary care interval.

Appendix 7. Appendices relating to Chapter 7

A7.1 Related publication in Journal of Public Health

Open Access text available at: https://doi.org/10.1093/pubmed/fdx188

Journal of Public Health | pp. 1–8 | doi:10.1093/pubmed/fdx188

The nature and frequency of abdominal symptoms in cancer patients and their associations with time to help-seeking: evidence from a national audit of cancer diagnosis

Minjoung Monica Koo[1], Christian von Wagner[1], Gary A. Abel[2], Sean McPhail[3], William Hamilton[2], Greg P. Rubin[4], Georgios Lyratzopoulos[1]

[1]University College London, 1-19 Torrington Place, London WC1E 6BT, UK
[2]University of Exeter Medical School, St Luke's Campus, Heavitree Road, Exeter EX1 2LU, UK
[3]National Cancer Registration and Analysis Service, Public Health England Zone A, 2nd Floor, Skipton House, 80 London Road, London SE1 6LH, UK
[4]Institute of Health and Society, Newcastle University, Sir James Spence Institute, Royal Victoria Infirmary, Newcastle upon Tyne NE1 4LP, UK
Address correspondence to Minjoung Monica Koo, E-mail: Monica.koo.14@ucl.ac.uk

ABSTRACT

Background Raising awareness of possible cancer symptoms is important for timely help-seeking; recent campaigns have focused on symptom groups (such as abdominal symptoms) rather than individual alarm symptoms associated with particular cancer sites. The evidence base supporting such initiatives is still emerging however; understanding the frequency and nature of presenting abdominal symptoms among cancer patients could inform the design and evaluation of public health awareness campaigns.

Methods We examined eight presenting abdominal symptoms (abdominal pain, change in bowel habit, bloating/distension, dyspepsia, rectal bleeding, dysphagia, reflux and nausea/vomiting) among 15 956 patients subsequently diagnosed with cancer in England. We investigated the cancer site case-mix and variation in the patient interval (symptom-onset-to-presentation) by abdominal symptom.

Results Almost a quarter (23%) of cancer patients presented with abdominal symptoms before being diagnosed with one of 27 common and rarer cancers. The patient interval varied substantially by abdominal symptom: median (IQR) intervals ranged from 7 (0–28) days for abdominal pain to 30 (4–73) days for dysphagia. This variation persisted after adjusting for age, sex and ethnicity ($P < 0.001$).

Conclusions Abdominal symptoms are common at presentation among cancer patients, while time to presentation varies by symptom. The need for awareness campaigns may be greater for symptoms associated with longer intervals to help-seeking.

Keywords cancer, health promotion, public health

Introduction

Diagnosing cancer early in symptomatic patients is a prominent feature of contemporary cancer control strategies.[1,2] A range of pioneering studies during the last decade have established associations between the knowledge ('awareness') of likely symptoms of cancer among the general public and timely presentation, diagnosis, and outcomes.[3–6] Public health agencies have consequently implemented educational interventions aimed at raising awareness of cancer symptoms in order to promote timely presentation.[7–9] However, the evidence base supporting the design of such interventions is still emerging.

Previous symptom awareness campaigns have tended to take a cancer-based approach, by targeting 'red-flag' or 'alarm' symptoms explicitly associated with specific cancers, such as 'blood in poo' and colorectal cancer.[10–12] There is however growing interest in targeting symptoms relating to a body area or system, partly as this provides an opportunity

Minjoung Monica Koo, PhD Student
Christian von Wagner, Reader in Behavioural Science and Health
Gary A. Abel, Senior Lecturer
Sean McPhail, Senior Analyst
William Hamilton, Professor of Primary Care Diagnostics
Greg P. Rubin, Professor of General Practice and Primary Care
Georgios Lyratzopoulos, Professor of Cancer Epidemiology

A7.2 Patients with missing patient interval

Among 2,253 cancer patients who had a single abdominal symptom at presentation, 21% (n=470) were missing information on the patient interval. These cancer patients were more likely to first present in places other than general practice and there was some variation by abdominal symptom, without evidence for variation in missing patient interval by age group and sex.

	Missing		Non-missing		Adjusted Odds Ratio†	Joint Wald test p-value
	N	%	N	%		
Total	470	21%	1783	79%		
Age group						
15–49 years	35	7%	143	8%	0.9 (0.6–1.3)	
50–69 years	195	41%	754	42%	Ref.	0.680
70+ years	240	51%	886	50%	1.0 (0.8–1.3)	
Sex						
Male	231	49%	969	54%	Ref.	
Female	239	51%	814	46%	1.1 (0.9–1.4)	0.220
Symptom group						
Abdominal pain	168	36%	502	28%	Ref.	
Change in bowel habit	91	19%	434	24%	0.8 (0.6–1.1)	
Rectal bleeding	81	17%	414	23%	0.6 (0.5–0.9)	
Dysphagia	43	9%	224	13%	0.7 (0.5–1.1)	**<0.001**
Dyspepsia	30	6%	88	5%	1.3 (0.8–2.0)	
Bloating or distension	26	6%	70	4%	1.2 (0.7–2.0)	
Nausea or vomiting	24	5%	29	2%	2.8 (1.5–5.1)	
Reflux	7	1%	22	1%	1.1 (0.5–2.9)	
Place of presentation						
General Practice	346	74%	1675	94%	Ref.	
Outpatients	24	5%	18	1.0%	7.0 (3.7–13.0)	
A&E	69	15%	41	2.3%	7.6 (5.0–11.4)	**<0.001**
Out of Hours	0	0.0%	6	0.3%	†	
Other	22	5%	42	2.4%	2.5 (1.4–4.2)	
Unknown	9	2%	1	0.1%	40.5 (5.1–323.7)	

†All six patients with a single abdominal symptom who had presented 'out of hours' had complete information on the patient interval and were therefore dropped from the model resulting in a sample size of 2,247

A7.3 Supplementary analyses: abdominal symptom constructs

In order to assess the robustness and validity of how I had defined the abdominal symptoms, I examined the 18 symptom constructs within each of the eight abdominal symptoms. Five of the eight examined abdominal symptoms represented aggregates of multiple symptom constructs. Three symptoms (dyspepsia, reflux and rectal bleeding) consisted of only a single symptom construct respectively, but were included for consistency and to aid comparison. Variation in the patient interval was examined using descriptive statistics; Kruskal-Wallis tests were used to statistically verify variation in median interval length by symptom construct within each symptom.

The symptom 'abdominal pain' represented an aggregation of five symptom constructs, of which abdominal pain NOS (not otherwise specified) was the most common followed by epigastric pain. On average, patients with epigastric pain waited longest before seeking help compared to other types of abdominal pain, although this was still relatively timely compared to abdominal pain NOS (median patient interval=12 days versus 7 days).

Three symptom constructs were aggregated to create the symptom 'change in bowel habit': change in bowel habit (verbatim), diarrhoea, and constipation. Constipation was least common but associated with shortest time to help-seeking compared to diarrhoea or change in bowel habit (median patient interval: 7 days versus 27 and 41 respectively).

The symptom 'dysphagia' mostly consisted of patients who had experienced dysphagia alone but 9 patients who had experienced odynophagia. The relatively small sample sizes should be taken into consideration but it is unsurprising that odynophagia (which is distinguished from dysphagia by the presence of pain) was associated with shorter time to help-seeking than dysphagia (median patient interval: 13 versus 30 days).

Summary statistics for the patient interval (measured in days) and the proportion of patients that experienced intervals exceeding 60 days, by abdominal symptom construct

				Patient interval[4]					
Symptom	Symptom construct	N	% (95% CI)	Mean	25th	50th	75th	90th	% 60+ days
Abdominal Pain	Abdominal pain[1]	515	23% (21–25%)	27	0	7	28	62	12%
	Epigastric pain	86	4% (3–5%)	39	3	12	31	105	17%
	Loin pain & renal colic	47	2% (2–3%)	22	0	2	31	61	13%
	Right iliac fossa pain	18	0.8% (0.5–1.3%)	11	0	1	5	31	7%
	Suprapubic pain	4	0.2% (0.1–0.5%)	0	0	0	0	0	0%
Change in Bowel Habit	Change in bowel habit	262	12% (10–13%)	78	13	41	94	201	42%
	Diarrhoea	176	8% (7–9%)	57	3	27	61	151	26%
	Constipation	87	4% (3–5%)	29	1	7	27	83	15%
Dyspepsia	Dyspepsia and related epigastric symptoms[2]	118	5% (4–6%)	31	0	14	30	87	15%
Dysphagia	Dysphagia	258	11% (10–13%)	49	10	30	61	118	26%
	Odynophagia	9	0.4% (0.2–0.8%)	17	0	13	31	60	0%
Reflux	Reflux	29	1.3% (0.9–1.8%)	41	0	16	61	128	27%
Bloating or Distension	Abdominal bloating	60	3% (2–3%)	46	3	26	62	92	27%
	Abdominal distension	27	1.2% (0.8–1.7%)	50	0	28	92	123	36%
	Ascites	9	0.4% (0.2–0.8%)	61	0	16	122	213	25%
Nausea or Vomiting	Vomiting	35	2% (1–2%)	24	1	4	17	105	12%
	Nausea	18	0.8% (0.5–1.3%)	85	0	15	98	183	33%
Rectal bleeding	Rectal bleeding[3]	495	22% (20–24%)	55	1	16	59	136	25%
All abdominal symptoms	-	2253	100%	47	1	16	54	122	23%

1 Abdominal pain that was not otherwise specified, excluding acute abdominal pain
2 includes dyspepsia, indigestion, waterbrash, gastritis, burping, belching, "GI upset", and "upper GI symptoms".
3 includes blood in stool or rectal bleeding, excludes acute rectal bleeding
4 n=1,783 as 21% had missing patient interval values
5 Kruskal-Wallis test of variation in median patient interval values across patients with one of 18 symptom constructs; $p<0.001$

Appendix 8. Appendices relating to Chapter 8

A8.1 Missing outcome data

A8.1.1 Primary care interval

Among 2,253 cancer patients who had a single abdominal symptom at presentation, 10% (n=236) were missing information on the primary care interval. As observed for the patient interval (see Appendix 7.2), cancer patients with missing information were more likely to first present in places other than general practice and there was some variation by abdominal symptom, without evidence for variation in missing primary care interval by age group and sex.

	Missing		Non-missing		Adjusted Odds Ratio†	Joint Wald test p-value
	N	%	N	%		
Total	236	10%	2017	90%		
Age group						
15–49 years	18	8%	160	8%	0.7 (0.4–1.3)	
50–69 years	104	44%	845	42%	Ref.	0.419
70+ years	114	48%	1012	50%	0.9 (0.6–1.2)	
Sex						
Male	107	45%	1093	54%	Ref.	
Female	129	55%	924	46%	1.4 (1.0–1.9)	0.062
Symptom group						
Abdominal pain	115	49%	555	28%	Ref.	
Change in bowel habit	36	15%	489	24%	0.6 (0.3–0.9)	
Rectal bleeding	41	17%	454	23%	0.5 (0.3–0.7)	
Dysphagia	11	5%	256	13%	0.3 (0.2–0.6)	**<0.001**
Dyspepsia	5	2%	113	6%	0.2 (0.1–0.7)	
Bloating or distension	12	5%	84	4%	0.8 (0.4–1.6)	
Nausea or vomiting	13	6%	40	2%	1.9 (0.9–4.3)	
Reflux	3	1%	26	1%	0.8 (0.2–3.3)	
Place of presentation						
General Practice	106	45%	1915	95%	Ref.	
Outpatients	27	11%	15	0.7%	36.5 (18.4–72.5)	
A&E	70	30%	40	2.0%	27.5 (17.5–43.2)	**<0.001**
Out of Hours	0	0.0%	6	0.3%	†	
Other	26	11%	38	1.9%	11.9 (6.8–20.7)	
Unknown	7	3%	3	0.1%	40.2 (9.8–164.6)	

†All six patients with a single abdominal symptom who had presented 'out of hours' had complete information on the primary care interval and were therefore dropped from the model resulting in a sample size of 2,247

A8.1.2 Number of pre-referral consultations

Variation in the odds of missing outcome data was examined using multivariate logistic regression models.

Among 2,253 cancer patients who had a single abdominal symptom at presentation, 12% (n=263) were missing information on the number of pre-referral consultations. Cancer patients with missing information were more likely to first present in places other than general practice and there was some variation by abdominal symptom, without evidence for variation by age group and sex.

	Missing		Non-missing		Adjusted Odds Ratio†	Joint Wald test p-value
	N	%	N	%		
Total	263	12%	1990	88%		
Age group						
15–49 years	21	8%	157	8%	0.9 (0.5–1.6)	
50–69 years	106	40%	843	42%	Ref.	0.672
70+ years	136	52%	990	50%	1.1 (0.8–1.5)	
Sex						
Male	140	53%	1060	53%	Ref.	
Female	123	47%	930	47%	0.9 (0.6–1.2)	0.379
Symptom group						
Abdominal pain	99	38%	571	29%	Ref.	
Change in bowel habit	37	14%	488	25%	0.7 (0.5–1.2)	
Rectal bleeding	71	27%	424	21%	1.4 (1.0–2.2)	
Dysphagia	25	10%	242	12%	1.1 (0.6–1.9)	0.021
Dyspepsia	4	2%	114	6%	0.2 (0.1–0.7)	
Bloating or distension	14	5%	82	4%	1.5 (0.7–3.1)	
Nausea or vomiting	10	4%	43	2%	1.5 (0.6–3.6)	
Reflux	3	1%	26	1%	1.2 (0.3–4.6)	
Place of presentation						
General Practice	123	47%	1898	95%	Ref.	
Outpatients	28	11%	14	0.7%	31.2 (15.8–61.4)	
A&E	76	29%	34	1.7%	35.7 (22.4–56.9)	
Out of Hours	1	0.4%	5	0.3%	2.6 (0.3–22.3)	<0.001
Other	29	11%	35	1.8%	12.3 (7.2–21.0)	
Unknown	6	2%	4	0.2%	25.3 (6.7–96.3)	

†Five of six patients with a single abdominal symptom who had presented 'out of hours' had complete information on the number of pre-referral consultations and so model sample size was 2,253 patients

A8.2 Frequencies of cancer signature by abdominal symptom

A8.2.1 Cancer signature of change in bowel habit

Cancer	N	% (95% CI)
Colorectal	791	78% (76–81%)
Ovarian	50	5% (4–6%)
Prostate	32	3% (2–4%)
Pancreatic	31	3% (2–4%)
Other	19	2% (1–3%)
Lymphoma	12	1% (1–2%)
Lung	11	1% (1–2%)
Small Intestine	11	1% (1–2%)
Renal	10	1% (1–2%)
Stomach	8	0.8% (0.4–1.6%)
Bladder	7	0.7% (0.3–1.4%)
Leukaemia	7	0.7% (0.3–1.4%)
Oesophageal	5	0.5% (0.2–1.2%)
Liver	4	0.4% (0.2–1.0%)
Sarcoma	4	0.4% (0.2–1.0%)
Breast	2	0.2% (0.1–0.7%)
Myeloma	2	0.2% (0.1–0.7%)
Cervical	1	0.1% (0.02–0.6%)
Endometrial	1	0.1% (0.02–0.6%)
Gallbladder	1	0.1% (0.02–0.6%)
Thyroid	1	0.1% (0.02–0.6%)

A8.2.2 Cancer signature of rectal bleeding

Cancer	N	% (95% CI)
Colorectal	705	92% (90–94%)
Other	24	3% (2–5%)
Prostate	8	1.0% (0.5–2.0%)
Lymphoma	7	0.9% (0.4–1.9%)
Bladder	4	0.5% (0.2–1.3%)
Pancreatic	4	0.5% (0.2–1.3%)
Leukaemia	3	0.4% (0.1–1.1%)
Renal	3	0.4% (0.1–1.1%)
Small Intestine	3	0.4% (0.1–1.1%)
Stomach	3	0.4% (0.1–1.1%)
Endometrial	2	0.3% (0.1–0.9%)
Lung	1	0.1% (0.02–0.7%)
Ovarian	1	0.1% (0.02–0.7%)

A8.2.3 Cancer signature of dysphagia

Cancer	N	% (95% CI)
Oesophageal	311	74% (70–78%)
Stomach	33	8% (6–11%)
Oropharyngeal	19	5% (3–7%)
Lung	15	4% (2–6%)
Laryngeal	11	3% (1–5%)
Lymphoma	9	2% (1–4%)
Pancreatic	5	1% (1–3%)
Other	5	1% (1–3%)
Colorectal	4	1.0% (0.4–2.4%)
Thyroid	2	0.5% (0.1–1.7%)
Brain	1	0.2% (0.04–1.3%)
Liver	1	0.2% (0.04–1.3%)
Melanoma	1	0.2% (0.04–1.3%)
Myeloma	1	0.2% (0.04–1.3%)

A8.2.4 Cancer signature of abdominal pain

Cancer	N	% (95% CI)
Colorectal	418	33% (30–36%)
Ovarian	148	12% (10–14%)
Pancreatic	133	10% (9–12%)
Stomach	75	6% (5–7%)
Renal	68	5% (4–7%)
Prostate	66	5% (4–7%)
Lymphoma	59	5% (4–6%)
Oesophageal	55	4% (3–6%)
Lung	41	3% (2–4%)
Other	41	3% (2–4%)
Bladder	28	2% (2–3%)
Gallbladder	25	2% (1–3%)
Liver	19	1% (1–2%)
Small Intestine	18	1% (1–2%)
Endometrial	16	1% (1–2%)
Leukaemia	12	0.9% (0.5–1.6%)
Myeloma	11	0.9% (0.5–1.5%)
Breast	9	0.7% (0.4–1.3%)
Cervical	9	0.7% (0.4–1.3%)
Sarcoma	7	0.6% (0.3–1.1%)
Testicular	5	0.4% (0.2–.9%)
Melanoma	2	0.2% (0.04–0.6%)
Mesothelioma	2	0.2% (0.04–0.6%)
Laryngeal	1	0.1% (0.01–0.4%)

A8.2.5 Cancer signature of nausea or vomiting

Cancer	N	% (95% CI)
Colorectal	62	24% (19–29%)
Pancreatic	34	13% (9–18%)
Oesophageal	30	11% (8–16%)
Stomach	29	11% (8–16%)
Lung	21	8% (5–12%)
Ovarian	13	5% (3–8%)
Other	11	4% (2–7%)
Lymphoma	10	4% (2–7%)
Brain	9	3% (2–6%)
Small Intestine	9	3% (2–6%)
Liver	7	3% (1–5%)
Renal	6	2% (1–5%)
Prostate	5	2% (1–4%)
Breast	3	1.1% (0.4–3.3%)
Myeloma	3	1.1% (0.4–3.3%)
Leukaemia	2	0.8% (0.2–2.8%)
Bladder	1	0.4% (0.1–2.1%)
Endometrial	1	0.4% (0.1–2.1%)
Gallbladder	1	0.4% (0.1–2.1%)
Oropharyngeal	1	0.4% (0.1–2.1%)
Sarcoma	1	0.4% (0.1–2.1%)
Testicular	1	0.4% (0.1–2.1%)
Thyroid	1	0.4% (0.1–2.1%)

A8.2.6 Cancer signature of dyspepsia

Cancer	N	% (95% CI)
Oesophageal	95	37% (31–43%)
Stomach	61	24% (19–29%)
Colorectal	30	12% (8–16%)
Pancreatic	26	10% (7–14%)
Lymphoma	10	4% (2–7%)
Lung	7	3% (1–6%)
Gallbladder	6	2% (1–5%)
Ovarian	5	2% (1–4%)
Other	5	2% (1–4%)
Liver	3	1.2% (0.4–3.4%)
Prostate	2	0.8% (0.2–2.8%)
Renal	2	0.8% (0.2–2.8%)
Leukaemia	1	0.4% (0.1–2.2%)
Myeloma	1	0.4% (0.1–2.2%)
Oropharyngeal	1	0.4% (0.1–2.2%)
Small Intestine	1	0.4% (0.1–2.2%)

A8.2.7 Cancer signature of bloating or distension

Cancer	N	% (95% CI)
Ovarian	112	45% (39–51%)
Colorectal	50	20% (16–25%)
Pancreatic	15	6% (4–10%)
Stomach	15	6% (4–10%)
Other	12	5% (3–8%)
Liver	8	3% (2–6%)
Lymphoma	7	3% (1–6%)
Endometrial	6	2% (1–5%)
Prostate	5	2% (1–5%)
Renal	4	2% (1–4%)
Small Intestine	3	1% (0.4–3%)
Breast	2	0.8% (0.2–2.9%)
Gallbladder	2	0.8% (0.2–2.9%)
Leukaemia	2	0.8% (0.2–2.9%)
Mesothelioma	2	0.8% (0.2–2.9%)
Bladder	1	0.4% (0.1–2.2%)
Cervical	1	0.4% (0.1–2.2%)
Lung	1	0.4% (0.1–2.2%)
Melanoma	1	0.4% (0.1–2.2%)
Oesophageal	1	0.4% (0.1–2.2%)

In men who presented with bloating or distension:

Cancer	N	% (95% CI)
Colorectal	24	37% (26–49%)
Liver	7	11% (5–21%)
Stomach	7	11% (5–21%)
Pancreatic	5	8% (3–17%)
Prostate	5	8% (3–17%)
Lymphoma	4	6% (2–15%)
Gallbladder	2	3% (1–11%)
Mesothelioma	2	3% (1–11%)
Renal	2	3% (1–11%)
Other	2	3% (1–11%)
Bladder	1	1.5% (0.3–8.2%)
Leukaemia	1	1.5% (0.3–8.2%)
Melanoma	1	1.5% (0.3–8.2%)
Oesophageal	1	1.5% (0.3–8.2%)
Small Intestine	1	1.5% (0.3–8.2%)

Appendix

In women who presented with bloating or distension:

Cancer	N	% (95% CI)
Ovarian	112	61% (53–67%)
Colorectal	26	14% (10–20%)
Pancreatic	10	5% (3–10%)
Other	10	5% (3–10%)
Stomach	8	4% (2–8%)
Endometrial	6	3% (1–7%)
Lymphoma	3	1.6% (0.6–4.7%)
Breast	2	1.1% (0.3–3.9%)
Renal	2	1.1% (0.3–3.9%)
Small Intestine	2	1.1% (0.3–3.9%)
Cervical	1	0.5% (0.1–3.0%)
Leukaemia	1	0.5% (0.1–3.0%)
Liver	1	0.5% (0.1–3.0%)
Lung	1	0.5% (0.1–3.0%)

A8.2.8 Cancer signature of reflux

Cancer	N	% (95% CI)
Oesophageal	35	49% (38–61%)
Stomach	15	21% (13–32%)
Pancreatic	8	11% (6–21%)
Colorectal	4	6% (2–14%)
Small Intestine	2	3% (1–10%)
Endometrial	1	1.4% (0.2–7.6%)
Leukaemia	1	1.4% (0.2–7.6%)
Liver	1	1.4% (0.2–7.6%)
Lung	1	1.4% (0.2–7.6%)
Ovarian	1	1.4% (0.2–7.6%)
Renal	1	1.4% (0.2–7.6%)
Other	1	1.4% (0.2–7.6%)

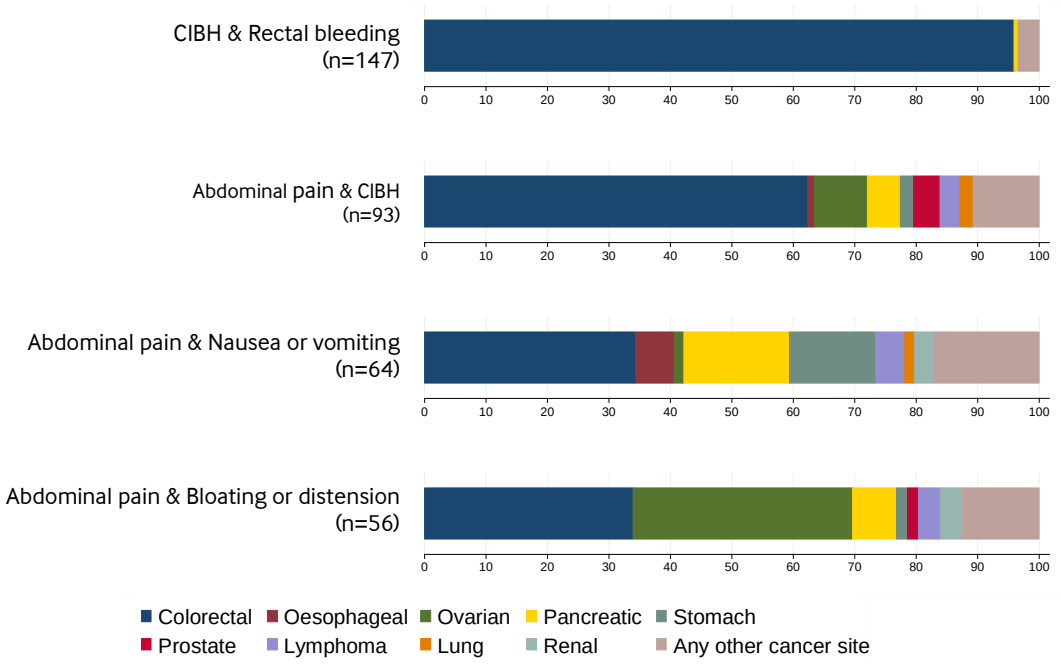
A8.3 Supplementary analyses: patients with multiple abdominal symptoms – post presentation

As described in Section 7.4.4, patients with the four most common abdominal symptom combinations were further analysed alongside eight abdominal symptoms as single symptoms (n=3,438, 94% of all patients with an abdominal symptom in the sample). The four symptom pairs were:

- Change in bowel habit (CIBH) and rectal bleeding;
- Abdominal pain and CIBH;
- Abdominal pain and nausea/vomiting; and
- Abdominal pain and bloating/distension.

A8.3.1 Cancer signatures of abdominal symptom pairs

The cancer signatures of the four most common pairs of abdominal symptoms were comparable to that of each individual symptom (see figure below and Figure 8.3).



A8.3.2 Distribution of the primary care interval

Summary statistics of the primary care interval (in days) and proportion of patients that experienced primary care intervals exceeding 60 days by symptom combination (see table below and Table 8.3).

Symptom combination	N[1]	Mean	25th	50th[2]	75th	90th	% 60+ days
Abdominal pain alone	806	36	2	15	42	88	35%
CIBH alone	640	32	0	8	32	91	24%
Rectal bleeding alone	549	23	0	1	12	63	8%
Dysphagia alone	336	15	0	0	12.5	43	11%
Dyspepsia alone	161	51	6	21	57	124	35%
CIBH & Rectal bleeding	141	37	0	2	29	106	15%
Nausea or vomiting alone	122	25	0	8	29	61	39%
Bloating or distension alone	124	23	1	7	22	58	25%
Abdominal pain & CIBH	84	58	1.5	11	60	182	46%
Abdominal pain & Nausea or vomiting	55	45	2	14	38	147	35%
Abdominal pain & Bloating or distension	48	27	2	7	39.5	70	34%
Reflux alone	33	47	0	17	49	142	30%

CIBH: change in bowel habit; Symptom pairs are in bold print.
1 Number of patients (percentages) sum to 3,099 as 10% (n=339/3,438) of observations had missing information on the patient interval.
2 Kruskal-Wallis test of variation in median patient values across patients with one of 12 symptom combinations; p<0.001

A8.4 Supplementary analyses: variation in the referral interval

I compared the median length of the referral interval (from first presentation in primary care to first consultation in secondary care) between cancer patients who presented with each of the three abdominal alarm symptoms and were diagnosed with the 'typical' cancer(s), compared with those diagnosed with a 'non-typical' cancer. There was no evidence for a difference in median referral interval values, and the IQRs were also comparable (see table below).

Symptom	N	Median (IQR) referral interval (days)	Wilcoxon rank-sum test for median values
Cancer patients with CIBH	506	13 (8–24)	-
CRC patients with CIBH	422	13 (8–23)	0.067
Other cancer patients with CIBH	84	10 (5–28)	
Cancer patients with rectal bleeding	456	14 (8–29)	-
CRC patients with rectal bleeding	425	13 (8–30)	0.471
Other cancer patients with rectal bleeding	31	14 (4–28)	
Cancer patients with dysphagia	254	12 (8–17)	-
OG cancer patients with dysphagia	225	11 (8–16)	0.508
Other cancer patients with dysphagia	29	13 (7–26)	

CIBH: change in bowel habit; CRC: colorectal cancer; OG: oesophageal

Appendix 9. Appendices relating to Chapter 9

[Please see overleaf for landscape table]

A9.1 Symptoms excluded from alarm symptom classification

Symptom	Included in BCOC campaign[1]	Positive predictive value (PPV) reported in the literature[2]	Mentioned in NICE 2005 and associated with mandated action[3]	Mentioned in NICE 2015 and associated with mandated action[3]	Patients in NACDPC cohort reporting the symptom (n)[4]
Axillary lump	Secondary symptom (breast cancer campaign)[5]	Not available	Not included	Refer urgently for suspected breast cancer	86
Abdominal pain	Secondary symptom (CRC, ovarian, OG, and bladder & kidney cancer campaigns)	0.3% for ovarian cancer (Hamilton et al, 2009) 0.3% for OG cancer (Stapley et al, 2013) 0.3% for pancreatic cancer (Stapley et al, 2013) 1.15% for CRC (Hamilton et al, 2005) 0.2% for bladder cancer (Shephard et al, 2012)	Refer urgently for suspected pancreatic or OG cancer in combination with weight loss	Refer urgently for suspected colorectal or OG cancer (if in combination with weight loss) Refer urgently for suspected CRC if in combination with weight loss	1054
Abdominal bloating	Primary symptom (ovarian cancer campaign) Secondary symptom (OG cancer campaign)	0.3% for ovarian cancer (Hamilton et al, 2009)	Not included	Not included	183
Breast symptoms other than breast lump	Secondary symptom (breast cancer campaign)[5]	Not available	Refer urgently for suspected breast cancer	Refer urgently for suspected breast cancer	307
Chest pain	Primary symptom (lung cancer campaign)	0.82% for lung cancer (Hamilton et al, 2005) 0.1% for myeloma (Shephard et al, 2015)	Refer urgently for suspected lung/mesothelioma cancer (alone or in combination with systemic symptoms)	Refer urgently for suspected lung/mesothelioma cancer (alone or in combination with systemic symptoms)	

Symptom	Included in BCOC campaign[1]	Positive predictive value (PPV) reported in the literature[2]	Mentioned in NICE 2005 and associated with mandated action[3]	Mentioned in NICE 2015 and associated with mandated action[3]	Patients in NACDPC cohort reporting the symptom (n)[4]
Cough	Primary symptom (lung cancer campaign)	0.4% for lung cancer (Hamilton et al, 2005)	Refer urgently for suspected lung/mesothelioma cancer	Refer urgently for suspected lung/mesothelioma cancer (alone or in combination with systemic symptoms)	730
Diarrhoea	Primary symptom (CRC campaign)	0.94% for CRC (Hamilton et al, 2009)	Not included	Not included	383
Dyspepsia	Secondary symptom (OG cancer campaign)	0.7% for OG cancer (Stapley et al, 2013)	Refer urgently for suspected pancreatic or OG cancer if recent-onset or in combination with chronic gastrointestinal bleeding, dysphagia, weight loss, vomiting, anaemia, epigastric mass, or suspicious barium meal result	Refer urgently for suspected OG cancer if in combination with weight loss	269
Fatigue	Secondary symptom (CRC, lung, ovarian cancer campaigns)	0.43% for lung cancer (Hamilton et al, 2005)	Not included	Refer urgently for suspected lung or mesothelioma if in combination with chest pain, cough, shortness of breath, loss of appetite	482
Flatulence	Secondary symptom (OG cancer campaign)	Not available	Not included	Not included	39
Heartburn	Primary symptom (OG cancer campaign)	0.6% for OG cancer (Stapley et al, 2013)	Not included	Refer urgently for suspected OG cancer if in combination with weight loss	75

Symptom	Included in BCOC campaign[1]	Positive predictive value (PPV) reported in the literature[2]	Mentioned in NICE 2005 and associated with mandated action[3]	Mentioned in NICE 2015 and associated with mandated action[3]	Patients in NACDPC cohort reporting the symptom (n)[4]
Dyspnoea (shortness of breath)	Secondary symptom (lung cancer campaign)	0.66% for lung cancer (Hamilton et al, 2005) 0.06% for myeloma (Shephard et al, 2015)	Refer urgently for suspected lung cancer	Refer urgently for suspected lung/mesothelioma cancer or haematological malignancy if alone or in combination with other symptoms	724

BCOC: Be Clear on Cancer; CRC: colorectal cancer; OG: oesophago-gastric cancer

1 The 'primary' or 'secondary' status of symptoms was inferred based on the design and phrasing of campaign materials: symptoms that were used to headline individual campaigns or that were described as 'key' were considered to be primary symptoms, while other symptoms mentioned in supporting material were considered to be secondary symptoms.

2 where individual PPVs were presented by age group or sex, the lowest value has been reported here

3 where phraseology indicated mandated action, usually two-week-wait referral ("refer") or urgent investigation ("offer"). Excludes symptoms for which guidance begins "consider".

4 among 15,956 cancer patients in NACDPC.

5 the primary focus of this campaign was to raise awareness of breast cancer among 70+ year old women rather than on raising awareness of particular presenting symptoms.

A9.2 Supplementary analyses: patients who presented and were referred within 30 days

I conducted crude and adjusted logistic regression analyses to examine the association between alarm symptoms and stage of diagnosis among patients who had a total pre-referral interval (sum of the patient interval and primary care interval) of 0–30 days. Given the more restrictive definition, this analysis was run on a smaller sample of patients compared to the main analysis (n=4,909, 70% of all patients who presented within 30 days). The proportion of late (distant) stage, and crude and adjusted odds ratios of late stage at diagnosis are presented below.

Variable	N	% distant stage	Crude OR[1] (95% CI)	P-value[2]	Adjusted OR[1] (95% CI)	P-value[2]
Alarm symptom[1]						
No alarm symptoms	2087	24% (22–26%)	Ref.		Ref.	
Abnormal mole	137	3% (1–7%)	0.1 (0.04–0.3)		**0.1 (0.04–0.3)**	
PMB	150	4% (2–8%)	0.1 (0.05–0.3)	**<0.001**	**0.1 (0.05–0.3)**	**<0.001**
Breast lump	1209	5% (4–7%)	0.2 (0.1–0.2)		**0.2 (0.1–0.2)**	
Haematuria	512	8% (6–10%)	0.3 (0.2–0.4)		**0.3 (0.2–0.4)**	
Rectal bleeding	200	14% (10–19%)	0.5 (0.3–0.8)		**0.5 (0.3–0.7)**	
Dysphagia	99	23% (16–32%)	1.0 (0.6–1.6)		1.0 (0.6–1.5)	
Haemoptysis	91	30% (21–40%)	1.3 (0.8–2.1)		1.3 (0.8–2.1)	
CIBH	166	27% (20–34%)	1.2 (0.8–1.6)		1.1 (0.8–1.6)	
Jaundice	79	35% (26–46%)	1.8 (1.1–2.8)		**1.7 (1.0–2.7)**	
Weight loss	97	46% (37–56%)	2.8 (1.8–4.2)		**2.7 (1.8–4.1)**	
Age group						
15–49 years	734	9% (7–12%)	0.5 (0.4–0.6)		0.7 (0.5–0.9)	
50–69 years	1838	18% (16–20%)	Ref.	**<0.001**	Ref.	**<0.001**
70+ years	2337	19% (17–20%)	1.1 (0.9–1.3)		1.0 (0.8–1.2)	
Sex						
Male	2146	20% (18–21%)	Ref.		Ref.	
Female	2763	15% (13–16%)	0.7 (0.6–0.8)	**0.003**	1.3 (1.1–1.5)	**<0.001**

CIBH: change in bowel habit; PMB: post-menopausal bleeding
1excludes 82 patients with multiple alarm symptoms
2Joint Wald test p-value

A9.3 Supplementary analyses: assuming patients with missing stage had late stage

I conducted further crude and adjusted logistic regression analyses examining the association between alarm symptoms and stage of diagnosis assuming that patients with missing information on stage at diagnosis were assigned to late stage (model n=7,467). The proportion of late (distant) stage, and crude and adjusted odds ratios of late stage at diagnosis are presented below.

Variable	N	% distant stage	Crude OR[1] (95% CI)	P-value[2]	Adjusted OR[1] (95% CI)	P-value[2]
Alarm symptom						
No alarm symptoms	3812	33% (31–34%)	Ref.		Ref.	
Abnormal mole	182	12% (8–18%)	**0.3 (0.2–0.4)**		**0.3 (0.2–0.5)**	
PMB	182	12% (8–18%)	**0.3 (0.2–0.4)**	**<0.001**	**0.2 (0.2–0.4)**	**<0.001**
Breast lump	1397	13% (12–15%)	**0.3 (0.3–0.4)**		**0.3 (0.3–0.4)**	
Haematuria	656	14% (11–17%)	**0.3 (0.3–0.4)**		**0.3 (0.3–0.4)**	
Rectal bleeding	289	21% (17–27%)	**0.6 (0.4–0.7)**		**0.5 (0.4–0.7)**	
Dysphagia	158	30% (23–37%)	0.9 (0.6–1.2)		0.8 (0.6–1.2)	
CIBH	348	32% (27–37%)	1.0 (0.8–1.2)		0.9 (0.7–1.1)	
Haemoptysis	136	35% (27–43%)	1.1 (0.7–1.5)		1.1 (0.7–1.5)	
Jaundice	106	42% (33–51%)	1.4 (1.0–2.1)		1.4 (0.9–2.0)	
Weight loss	201	53% (46–60%)	**2.3 (1.7–3.0)**		**2.2 (1.6–2.9)**	
Age group						
15–49 years	1034	16% (14–18%)	0.5 (0.4–0.6)		0.6 (0.5–0.7)	
50–69 years	2896	27% (26–29%)	Ref.	**<0.001**	Ref.	**<0.001**
70+ years	3683	30% (28–31%)	1.1 (1.0–1.3)		1.1 (1.0–1.2)	
Sex						
Male	3575	29% (27–30%)	Ref.		Ref.	
Female	4038	25% (24–27%)	0.9 (0.8–0.9)	**0.003**	1.2 (1.1–1.4)	**<0.001**

CIBH: change in bowel habit; PMB: post-menopausal bleeding
1excludes 146 patients with multiple alarm symptoms
2Joint Wald test p-value