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 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.

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

© 2017 The Authors. Published by Elsevier Inc. on behalf of Neoplasia Press, Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1476-5586

https://doi.org/10.1016/j.neo.2017.11.005
Approaches to measuring presenting symptoms in cancer patient populations

In this Review, we make frequent use of two terms: symptom signature and diagnostic difficulty. Symptom signature denotes the nature and relative frequency of symptoms (or symptom combinations) reported at presentation by patients later diagnosed with a particular cancer [13,14]. We describe symptom signatures as being ‘narrow’ when most patients present with a particular symptom (as is the case for breast lump in the context of breast cancer) or ‘broad’ when patients present with a larger range of symptoms (as is the case for colorectal cancer).

The term diagnostic difficulty (of a given cancer site) has previously been used to characterize cancer sites as “harder-to-suspect” (e.g. multiple myeloma, pancreatic cancer) or “easier-to-suspect” (e.g. breast cancer, melanoma) based on the profile of presenting symptoms [13]. It represents the perceived predictive value for cancer of the presenting symptoms of the ‘average’ patient.

We discuss methodological challenges in capturing data on symptoms at presentation and intervals to diagnosis and subsequently examine the symptom signatures of cancer sites and how this relates to diagnostic difficulty (Box 1). Diagnostic difficulty is related to the positive predictive value (PPV) of a symptom for a given disease, which is the proportion of all patients with the same symptom who will be found to have the disease. While PPV is a continuous measure, explicit threshold categories for all patients with the same symptom who will be found to have the disease.

Additionally, psychosocial barriers (such as embarrassment [47–49]) and perceived or actual time pressures during the consultation [50] may prevent complete disclosure of certain symptoms to the doctor. Coded information can also be less sensitive to qualitative distinctions in symptom experience such as temporal evolution, particularly if multiple symptoms are recorded.

Defining symptom signature and diagnostic difficulty

In this Review, we make frequent use of two terms: symptom signature and diagnostic difficulty. Symptom signature denotes the nature and relative frequency of symptoms (or symptom combinations) reported at presentation by patients later diagnosed with a particular cancer [13,14]. We describe symptom signatures as being ‘narrow’ when most patients present with a particular symptom (as is the case for breast lump in the context of breast cancer) or ‘broad’ when patients present with a larger range of symptoms (as is the case for colorectal cancer).

The term diagnostic difficulty (of a given cancer site) has previously been used to characterize cancer sites as “harder-to-suspect” (e.g. multiple myeloma, pancreatic cancer) or “easier-to-suspect” (e.g. breast cancer, melanoma) based on the profile of presenting symptoms [13]. It represents the perceived predictive value for cancer of the presenting symptoms of the ‘average’ patient.

How Can Presenting Symptoms and Intervals Before Diagnosis be Measured?

Capturing information on symptoms is challenging, as the majority cannot be objectively observed and their appraisal by individuals is influenced by sociocultural factors such as level of education and health literacy (including awareness of likely cancer symptoms), cancer fear, or fatalism [14,15]. When more than one symptom is experienced, the combination of symptoms could also influence appraisal and help-seeking. Additionally, several symptoms may have conflicting or overlapping meanings in lay and professional language, and this is reflected in heterogeneous terminology in published literature. For example, abdominal bloating (uncomfortable sensation of fullness) and distension (visible increase in abdominal girth) have been used interchangeably [16,17], while ‘change in bowel habit’ is often used by clinicians to denote a clinical suspicion of colorectal cancer beyond the presence of constipation or diarrhea alone [18]. Further, heterogeneity exists within certain nonspecific symptoms: ‘abdominal pain,’ for example, encompasses a range of presentations that vary greatly in nature, intensity, duration, and temporal evolution.

Self-reported symptom information.

Information on presenting symptoms can be directly elicited from patients through semistructured interviews [26–31] or questionnaires [32,33]. Such methods can elicit valuable first-hand insights into the symptomatic and diagnostic experience. Patients may be prompted to identify their presenting symptoms from a predefined list (symptom recognition) or to describe them without any prompting (symptom recall), which can affect the degree of recall inaccuracies or bias. 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 [34]. Studies can also be distinguished by whether the information is collected before or after the diagnosis. Collecting data about presenting symptoms after diagnosis is more convenient due to easier identification of cases but it can lead to both recall and survivorship bias. The latter results in underrepresentation of cancer patients with poor prognosis, whose presenting symptoms could be different to those of the studied patients [35]. In comparison, collecting information prospectively (before a diagnosis of cancer is made) has the advantage of minimizing such potential biases [36–38].

Records-based symptom information.

Alternatively, information on presenting symptoms can be recorded during healthcare encounters (e.g., with a primary care physician) and captured as part of the patients’ health records [39–41]. Both coded and free-text information may be extracted [42–45].

In principle, studies collecting symptom information from patient records are less prone to the risk of selection and recall bias, as information on presenting symptoms is collected prospectively and prior to diagnosis for all patients. However, such methods critically rely on the symptoms both being elicited during the consultation and being accurately recorded; in many instances, these assumptions may not be met [46,47]. Additionally, psychosocial barriers (such as embarrassment [47–49]) and perceived or actual time pressures during the consultation [50] may prevent complete disclosure of certain symptoms to the doctor. Coded information can also be less sensitive to qualitative distinctions in symptom experience such as temporal evolution, particularly if multiple symptoms are recorded.

Approaches to measuring presenting symptoms in cancer patient populations

Patient may be prompted to identify their presenting symptoms from a predefined list (symptom recognition) or to describe them without any prompting (symptom recall), which can affect the degree of recall inaccuracies or bias. 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 [34]. Studies can also be distinguished by whether the information is collected before or after the diagnosis. Collecting data about presenting symptoms after diagnosis is more convenient due to easier identification of cases but it can lead to both recall and survivorship bias. The latter results in underrepresentation of cancer patients with poor prognosis, whose presenting symptoms could be different to those of the studied patients [35]. In comparison, collecting information prospectively (before a diagnosis of cancer is made) has the advantage of minimizing such potential biases [36–38].

Records-based symptom information.

Alternatively, information on presenting symptoms can be recorded during healthcare encounters (e.g., with a primary care physician) and captured as part of the patients’ health records [39–41]. Both coded and free-text information may be extracted [42–45].

In principle, studies collecting symptom information from patient records are less prone to the risk of selection and recall bias, as information on presenting symptoms is collected prospectively and prior to diagnosis for all patients. However, such methods critically rely on the symptoms both being elicited during the consultation and being accurately recorded; in many instances, these assumptions may not be met [46,47]. Additionally, psychosocial barriers (such as embarrassment [47–49]) and perceived or actual time pressures during the consultation [50] may prevent complete disclosure of certain symptoms to the doctor. Coded information can also be less sensitive to qualitative distinctions in symptom experience such as temporal evolution, particularly if multiple symptoms are recorded.
Similar challenges exist when measuring prediagnostic intervals experienced by cancer patients. Existing methods rely on the validity of the recall of particular dates of significance along the diagnostic pathway including the date of symptom onset, the date of first relevant symptomatic presentation (help-seeking), and the date of first referral to secondary care [11].

Two principal study designs have been described to examine diagnostic intervals: collecting self-reported information from patients and extracting information from patients’ health records [9,11]. We propose that these approaches are also relevant to the study of presenting symptoms (Box 2). Inconsistencies between self-reported and records-based information have been described, reflecting that both approaches have limitations [19–22]. Nonetheless, medical record studies offer the opportunity to examine prediagnostic symptoms (and intervals) in large and representative samples of patients, additionally facilitating the study of patients with rarer cancers [10,23–25].

**What are the Symptom Signatures of Different Cancer Sites?**

Understanding the nature and relative frequency of presenting symptoms associated with different cancer sites is necessary before investigating how symptoms may influence diagnostic timeliness. We therefore reviewed the literature to examine the symptom signature of common and rarer cancers (see Box 3) and present the findings here. We consider cancers in three groups based on symptom signatures, taking into account symptom heterogeneity (the ‘breadth’ of the symptom signature) and their predictive value (see Figure 1).

**Cancers With a Narrow Symptom Signature**

In this category, we consider several cancers where the majority of patients present with one symptom with adequately strong association with a given cancer (these are also known as ‘alarm’ symptoms). 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; see Table 1 [44,59,60]. Similarly, most bladder cancer patients present with macroscopic (visible, frank) hematuria (Table S3.1) [43,61,62].

Likewise, the symptomatic presentations of thyroid cancer, melanoma, testicular cancer, penile cancer, vaginal cancer, vulval cancer, and sarcoma are also narrow and are likely to have meaningfully high predictive values, although empirical documentation of the symptom signatures of these cancers is limited [63,64]. Importantly, a relatively narrow symptom signature does not necessarily guarantee swift or easy diagnostic resolution for all patients. Firstly, the nature of the symptoms associated with sarcomas (soft tissue lump or bone pain) suggests the level of diagnostic difficulty should be low, but the relative rarity of sarcomas among the general population means that alternative diagnoses are often provided [65]. Raising awareness of these symptoms could achieve earlier diagnosis for this cancer type given the reasonable predictive values for malignancy [66], but...

---

**Box 3**

Data sources and findings of literature review on the symptom signatures of cancer and associations between symptoms and diagnostic intervals

**Methodology**

We searched for studies describing the frequency of presenting symptoms of cancer patients based on either primary care records or prospectively collected self-reported information, supplemented by expert knowledge of relevant evidence. Studies describing self-reported symptoms captured retrospectively (after diagnosis) were excluded due to the high risk of bias. Studies on pediatric, teenager, and young adult cancer patient populations and studies based in low- and middle-income settings were excluded as they were deemed not comparable. 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.

There is no standard assessment tool for risk of bias in observational nonrandomized studies, and so we developed a risk of bias tool based on the REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) and the Quality Assessment of Diagnostic Accuracy Studies 2nd version (QUADAS-2) checklists [51,52]. The resulting quality appraisal tool assessed risk of bias across six dimensions: setting, study population, symptoms, external validity, data cleaning, and other sources of bias (Table S1). This tool was used to further exclude any studies that had three or more dimensions with “high” risk of bias.

**Summary of findings**

We identified a total of 41 studies including information on presenting symptoms for 16 cancer sites (see Supplementary Material for symptom frequencies by cancer site). All included studies were based in the UK and were mostly case-control or cohort studies 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, 18 (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) (Table S2). Nearly all studies focused on single cancer sites, with the exception of colorectal, esophagogastric, and renal tract cancers, which were treated as single entities, respectively [53–58]. Most evidence related to colorectal cancer (eight studies, Table S3.4), pancreatic cancer (six studies, Table S3.9), and lung cancer (five studies, Table S3.6); only a single publication was identified for five cancers (brain, cervical, endometrial, leukemia, and myeloma).

No evidence on the frequency of presenting symptoms before diagnosis could be identified for the following 12 cancers: laryngeal cancer; liver cancer; melanoma; mesothelioma; oral cancer; penile cancer; sarcoma; small intestinal cancer; testicular cancer; thyroid cancer; vaginal cancer; and vulval cancer. Of the 41 studies included in the Review, only four also contained evidence on associations between individual symptoms and intervals to diagnosis [36–38,44]. One study was not included in the consideration of symptom signatures as symptom information was collected through a combination of recall and recognition; however, evidence pertaining to symptom-specific diagnostic timeliness has been considered here [22].
Cancers with a narrow symptom signature of sufficiently high predictive value
Breast cancer
Bladder cancer
Melanoma
Thyroid cancer
Testicular cancer

Cancers with a broad symptom signature of varying predictive value
Colorectal cancer
Gastric cancer
Lung cancer
Esophageal cancer
Oropharyngeal cancer
Ovarian cancer
Renal cancer

Cancers with a broad symptom signature of low predictive value
Brain/CNS cancers
Hematological cancers

Figure 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. CNS: Central Nervous System.

The cost-effectiveness of a population-wide intervention for such rare cancers would be a concern. Secondly, a minority of patients with ‘narrow symptom signature’ cancers will have atypical presentations, which tend to be associated with a longer time to presentation and referral [44,67].

Cancers With a Broad Symptom Signature
In this category, we consider cancer sites characterized by a broad symptom signature. For some cancers, this includes certain alarm symptoms (e.g., colorectal, lung, pancreatic, esophagogastric, and ovarian cancers), while for other cancers, presenting symptoms are chiefly nonspecific (e.g., hematological malignancies, and brain and CNS cancers).

Broad Symptom Signature, Varying Predictive Value. Many common cancers have broad symptom signatures consisting of multiple symptoms, of which only relatively few (e.g. one or two) are alarm symptoms that are strongly predictive of cancer. For example, eight studies report rectal bleeding, which has only relatively high predictive value, as a common presenting symptom of colorectal cancer, although estimates vary substantially (16%-60%) [18,38,68–73]. Other common presenting symptoms among colorectal cancer patients include abdominal pain, diarrhea, and constipation, which are associated with much lower predictive values and greater diagnostic difficulty (Table S3.4). Lung cancer also has a broad symptom signature with symptoms of varying predictive value: while it includes hemoptysis, a highly predictive symptom of malignancy [74], evidence from six studies suggests that this is a relatively rare presenting symptom, reported in less than a quarter (20%-23%) of patients subsequently diagnosed with lung cancer (Table 2) [36,74–76].

We identified six studies describing the frequencies of the presenting symptoms of pancreatic cancer (Table S3.9) [37,43,77–80]. Jaundice has a high predictive value for pancreatic cancer, but reported frequencies range from 12% to 43% among patients, and it is often a sign of advanced disease [37,43,77–80]. The most common presenting symptom among pancreatic cancer patients is abdominal pain (reported range: 40%-57% of cases), while other upper gastrointestinal symptoms such as indigestion and nausea and vomiting are also common— and given their frequency among primary care consultees, these symptoms have naturally low predictive values. Studies also reported frequencies of back pain and nonlocalizing symptoms such as weight loss, 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 (Table S3.9).

Current data on the symptom signatures of esophageal and gastric cancers are limited to studies that describe them in combination (Table S3.7) [53–56]. While dysphagia is the most common presenting symptom in this cancer patient population (an alarm symptom shown to be highly predictive of malignancy), 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 [53–56].

Likewise, ovarian cancer has a symptom signature encompassing a broad spectrum of abdominal symptoms, although existing evidence tends to be based on smaller study populations due to the low incidence of

<table>
<thead>
<tr>
<th>Study</th>
<th>Setting/Source of Data</th>
<th>Study Period</th>
<th>Sample Size</th>
<th>Study Population</th>
<th>Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walker et al., 2014</td>
<td>Primary care, CPRD data (Read coded)</td>
<td>2000-09</td>
<td>3166</td>
<td>40+ years</td>
<td>Breast lump 44.1% Breast pain 2.4% Nipple discharge 1.0% Nipple retraction 1.0% Breast lump 3.5% Breast pain 4.6% Nipple distortion 1.5% Nipple eczema 0.2% Breast skin changes 0.2% Bloody nipple 0.01%</td>
</tr>
<tr>
<td>Redaniel et al., 2015</td>
<td>Primary care, CPRD data (Read coded)</td>
<td>1998-09</td>
<td>8544</td>
<td>15+ years</td>
<td>Breast lump 83% Nipple abnormalities 6.8% Breast pain 6.4% Breast skin abnormalities 2.0% Axillary lump 1.2% Breast ulceration 1.1% Back pain 1.0% Breast contour abnormalities 0.7% Breast infection or inflammation 0.6% Breast swelling 0.6% Musculoskeletal pain 0.6% Breathlessness 0.5% Breast rash 0.4%</td>
</tr>
<tr>
<td>Koo et al., 2017</td>
<td>Primary care, audit data (free text)</td>
<td>2009-10</td>
<td>2316</td>
<td>20+ years</td>
<td>Breast lump 44.1% Breast pain 2.4% Nipple discharge 1.0% Nipple retraction 1.0% Breast lump 3.5% Breast pain 4.6% Nipple distortion 1.5% Nipple eczema 0.2% Breast skin changes 0.2% Bloody nipple 0.01%</td>
</tr>
</tbody>
</table>

* All symptom frequencies calculated 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).
† Symptoms in 10 or more women listed only; further symptoms listed in Supplementary files of original paper.

Table 1. Population-Based Estimates of the Frequencies of Presenting Symptoms Among Breast Cancer Patients [44,59,60]
Abdominal distension has been shown to have a reasonable predictive value for cancer, but it is only reported by 8% to 38% of patients before diagnosis; similarly common presenting symptoms including abdominal bloating and abdominal pain have much lower predictive values [17]. Additionally, a wide variety of other symptoms with low predictive values have been identified as presenting symptoms of ovarian cancer, such as vaginal bleeding, upper and lower gastrointestinal symptoms, and systemic symptoms [47,81–83].

**Broad Symptom Signature, Low Predictive Value.** Some cancers have broad symptom signatures comprising almost entirely symptoms of low predictive value. Results from four studies indicate that hematological cancers (leukemia, lymphoma, and multiple myeloma) have such symptom signatures, comprised of vague or nonlocalizing symptoms such as fatigue and weight loss or common complaints such as back pain (Table S13-15) [33,84–86]. Consequently, some hematological cancers, and multiple myeloma in particular, are associated with a high level of diagnostic difficulty, as also evidenced by high frequency of multiple consultations in primary care before specialist referral [10].

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 nonspecific symptoms, with very low predictive value (Table 3) [87]. Achieving earlier diagnosis of brain and other neurological cancers is therefore associated with substantial diagnostic difficulties.

### Table 2. Population-Based Estimates of the Frequencies of Presenting Symptoms Among Lung Cancer Patients [36,60,74–76]

<table>
<thead>
<tr>
<th>Study</th>
<th>Setting/Source of Data</th>
<th>Study Period</th>
<th>Sample Size</th>
<th>Study Population Age/Range</th>
<th>Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hamilton et al., 2005a</td>
<td>Primary care, data from 21 general practices in Exeter</td>
<td>1998-02</td>
<td>247</td>
<td>40+ years</td>
<td>Hemoptysis 20%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Weight loss 27%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Loss of appetite 19%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dyspnea 56%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chest or rib pain 42%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fatigue 35%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fingerclubbing 4.5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Thrombocytosis 14%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Abnormal ophthalmoscopy 9.7%</td>
</tr>
<tr>
<td>Hippisley-Cox &amp; Coupland, 2011a</td>
<td>Primary care, QResearch data (Read coded)</td>
<td>2000-10</td>
<td>2196</td>
<td>30-84 years</td>
<td>Hemoptysis 23.0% *</td>
</tr>
<tr>
<td>Ades et al., 2014</td>
<td>Primary care, data from 21 general practices in Exeter</td>
<td>1998-02</td>
<td>247</td>
<td>40+ years</td>
<td>Cough 64.8%</td>
</tr>
<tr>
<td>Redaniel et al., 2015</td>
<td>Primary care, CPRD data (read coded)</td>
<td>1998-09</td>
<td>5737</td>
<td>15+ years</td>
<td>Hemoptysis 8.8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SVC obstruction 0.4%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stridor 0.1%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Asthma 1.7%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cervical lymphadenopathy 0.5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chest signs 2.8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chest/rib pain 14.9%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cough 40.9%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dyspnea 18.5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fatigue 4.1%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fingerclubbing 0.5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hoarseness 1.9%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Shoulder pain 5.0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Coughing up blood 21.6%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cough or worsening cough 56.2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Breathlessness or worsening breathlessness 41.2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chest/shoulder pain 55.3%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hoarseness 12.4%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Decreased appetite 22.2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Unexplained weight loss 15%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fatigue or tiredness 45.1%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Feeling different “in yourself” 34.6%</td>
</tr>
</tbody>
</table>

SVC, superior vena cava.

* Frequencies of other symptoms included in study were not reported.

† Same study population as Hamilton et al., 2005a; frequencies of additional/different symptoms displayed only.

### Table 3. Population-Based Estimates of the Frequencies of Presenting Symptoms Among Brain or CNS Cancer Patients [87]

<table>
<thead>
<tr>
<th>Study</th>
<th>Setting/Source of Data</th>
<th>Study Period</th>
<th>Sample Size</th>
<th>Study Population Age/Range</th>
<th>Symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hamilton et al., 2007</td>
<td>Primary care, CPRD data</td>
<td>1988-06</td>
<td>3505</td>
<td>18+ years</td>
<td>Headache 10.2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Motor loss 8.8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>New onset seizure 4.4%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Confusion 3.1%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Weakness 2.7%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Memory loss 1.1%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Visual disorder 1.0%</td>
</tr>
</tbody>
</table>
groupings of presenting symptoms have been used among specific cancer patient populations, such as lump versus no lump among either breast or sarcoma patients [91,92].

Available evidence on individual symptoms and diagnostic timeliness is currently limited to four cancers (breast, colorectal, lung, pancreatic) and is derived from study designs that combine prospectively collected patient information with primary and secondary care records, or examine data collected as part of clinical audit initiatives (see Box 4). Expanding this line of enquiry to other cancer patient populations, such as lump versus no lump among breast or sarcoma patients [91,92], may not be fully representative of the symptomatic patient population. Examining all presenting symptoms of a cohort might not have been captured and reported symptom frequencies in the literature or clinical guidelines. Rarer or less-specific symptoms with prespecified symptoms (identified a priori) either from relevant literature or clinical guidelines. Rarer or less-specific symptoms might not have been captured and reported symptom frequencies may not be fully representative of the symptomatic patient population. Examining all presenting symptoms of a cohort of cancer patients without prior restrictions can bring valuable insights [36–38,44,101].

Implications for Early Diagnosis Initiatives
Most patients with cancers characterized by a narrow symptom signature (such as breast and bladder cancer) experience relatively difficulty. Given the likely variation across countries in diagnostic activity and access to investigations, international comparisons through collaborative efforts such as the International Cancer Benchmarking Partnership (ICBP) could be valuable [96].

Further, some presenting symptoms such as jaundice are likely to represent advanced disease. In these patients, diagnostic difficulty could be minimal, but expediting their diagnosis may not necessarily lead to favorable clinical outcomes or alter prognosis. Understanding associations between symptoms and stage at diagnosis is an important area for future research [97].

There was substantial variation in reported symptom frequencies between studies, reflecting the heterogeneity in how symptom information was reported, extracted, and collated. Optimizing data capture by improving the application of existing clinical coding systems (and physician compliance) is important, particularly as novel technologies such as machine learning and natural language processing are used to extract information from electronic health records [98–100]. Until we are able to capture population-based information on symptoms systematically and reliably before diagnosis, existing methodologies such as clinical audits and prospective cohort studies offer opportunities to examine the presenting symptoms of cancer and associated diagnostic intervals [36–38,44].

Many of the studies included in this Review investigated patients with prespecified symptoms (identified a priori) either from relevant literature or clinical guidelines. Rarer or less-specific symptoms might not have been captured and reported symptom frequencies may not be fully representative of the symptomatic patient population. Examining all presenting symptoms of a cohort of cancer patients without prior restrictions can bring valuable insights [36–38,44,101].

**Box 4**

**Symptoms and time to diagnosis: emerging evidence**

Data from three SYMPTOM studies in England on lung, colorectal, and pancreatic cancers provide some early insights into variation in intervals to diagnosis by individual symptoms [36–38]. Symptom information was collected 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) [36–38]. The quantitative examination of presenting symptoms and intervals to diagnosis has also been enhanced (triangulated) with qualitative analysis of in-depth patient and healthcare professional interviews [93,94].

The multicenter DECCIRE study used a comparable design to collect information on the diagnostic process for 795 colorectal cancer patients in Spain [95]. Symptom information was elicited from patients shortly after diagnosis by a combination of recall and recognition, and corroborated with medical records from which prediagnostic intervals were estimated [22]. Of the examined symptoms, bowel obstruction was the only independent predictor of a shorter diagnostic interval (time from symptom onset to diagnosis), although investigators noted significant differences in interval length depending on the method of data collection (patient interview, hospital records, primary care records) [22].

The presenting symptoms of breast cancer and associated patient and primary care intervals have been examined using primary care data in England [44]. The study documented that women with nonlump breast symptoms, and women with both breast lump and nonlump breast symptoms sought help later than those who presented with breast lump alone. Risks of recall or selection bias were minimal, as for the SYMPTOM/DECCIRE studies, and the study had a large representative sample (n = 2316 women) [44]. Further utilization of health records data could enable the investigation of symptom-specific timeliness of presentation and referral in greater detail.

**Discussion**

Measuring presenting symptoms and intervals before the diagnosis of cancer in patient populations is challenging. Currently, there are two main approaches: self-report versus records-based information. We identified 41 population-based studies describing information on the symptom signatures of 16 common and rarer cancers. Based on our findings, we described these symptom signatures as narrow (e.g., breast, bladder cancers), broad comprising cancers with some highly predictive symptoms (e.g., colorectal or pancreatic cancer), or broad comprising cancers characterised by mostly nonspecific symptoms. Evidence on how presenting symptoms relate to prediagnostic intervals was limited, but emerging findings indicate notable variation that could be used to guide interventions.

We have reviewed the interrelated concepts of symptom signature and diagnostic difficulty, the latter being an expression of symptom-specific predictive values. It should be acknowledged however that organizational or system factors can also influence the difficulty of diagnosis of a particular cancer site. For example, both the availability and the accessibility of different clinical investigations may influence the diagnostic difficulty of a cancer. Full blood counts can be more readily ordered than paraprotein studies in primary care; therefore, leukemia may be investigated with a lower threshold of cancer suspicion than myeloma and has lower overall diagnostic difficulty. Given the likely variation across countries in diagnostic activity and access to investigations, international comparisons through collaborative efforts such as the International Cancer Benchmarking Partnership (ICBP) could be valuable [96].

Further, some presenting symptoms such as jaundice are likely to represent advanced disease. In these patients, diagnostic difficulty could be minimal, but expediting their diagnosis may not necessarily lead to favorable clinical outcomes or alter prognosis. Understanding associations between symptoms and stage at diagnosis is an important area for future research [97].

There was substantial variation in reported symptom frequencies between studies, reflecting the heterogeneity in how symptom information was reported, extracted, and collated. Optimizing data capture by improving the application of existing clinical coding systems (and physician compliance) is important, particularly as novel technologies such as machine learning and natural language processing are used to extract information from electronic health records [98–100]. Until we are able to capture population-based information on symptoms systematically and reliably before diagnosis, existing methodologies such as clinical audits and prospective cohort studies offer opportunities to examine the presenting symptoms of cancer and associated diagnostic intervals [36–38,44].

Many of the studies included in this Review investigated patients with prespecified symptoms (identified a priori) either from relevant literature or clinical guidelines. Rarer or less-specific symptoms might not have been captured and reported symptom frequencies may not be fully representative of the symptomatic patient population. Examining all presenting symptoms of a cohort of cancer patients without prior restrictions can bring valuable insights [36–38,44,101].

**Implications for Early Diagnosis Initiatives**
Most patients with cancers characterized by a narrow symptom signature (such as breast and bladder cancer) experience relatively
short intervals to help-seeking [7-9,32], although a minority of patients have atypical presentations and experience prolonged intervals to presentation. Research efforts are needed to improve timely presentation in the latter group. Further, public health education campaigns about alarm symptoms remain important for improving awareness among minority groups and in the context of low- and middle-income countries [44,102–104]. For cancers with a broad symptom signature, promoting timely help-seeking is more challenging. While many patients diagnosed with such cancers will present with alarm symptoms, many others will present with symptoms of low predictive value. Thus, while raising awareness of alarm symptoms associated with those cancers remains important, complementary strategies need to be developed. Public health education campaigns could provide information on symptom combinations or symptom duration, and also address attitudinal and psychosocial barriers to help-seeking for new symptoms, such as cancer fear and fatalism [37,105,106].

Postpresentation, patients with alarm symptoms are likely to benefit from fast-track diagnostic assessment pathways [24,88], but patients with significant but nonlocalizing symptoms present a greater challenge. Innovation in diagnostic strategies, either through the development of new diagnostic tests or novel uses of existing technologies, is needed. Rapid access to specialist investigative expertise and testing strategies in the form of multidisciplinary diagnostic centers (MDCs) have recently been implemented in Denmark and are currently being developed in the United Kingdom [107,108]. In addition to improving the diagnosis of cancer, such services can additionally improve the diagnosis of a range of other serious (nonneoplastic) diseases [107,109,110]. Further, for patients with nonresolving vague symptoms of low specificity, planned reevaluation through safety-netting approaches can minimize prolonged time to diagnosis [111].

In conclusion, the diagnostic difficulty of a cancer is closely tied to its symptom signature. Expanding current scientific knowledge about the nature of presenting symptoms and how they are associated with diagnostic intervals will further our understanding of mechanisms that influence the diagnostic pathway at patient, healthcare professional, and system levels. Doing so will strengthen the evidence base to support the development and implementation of public health and healthcare interventions promoting early diagnosis, thereby resulting in improved outcomes for cancer patients.

Funding
This work was supported by a grant from the UK Department of Health [grant number no. 106/0001]. This work was part of the program of the Policy Research Unit in Cancer Awareness, Screening and Early Diagnosis. The Policy Research Unit in Cancer Awareness, Screening, and Early Diagnosis receives funding for a research program from the Department of Health Policy Research Programme. It 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 Peninsula Medical School/University of Exeter). G. L. is supported by a Cancer Research UK Advanced Clinician Scientist Fellowship [grant number: C18081/A18180]. F. M. W. is supported by a National Institute for Health Research (NIHR) Clinician Scientist award. The views expressed are those of the authors and not necessarily those of the Department of Health, Cancer Research UK, or NIHR.

Appendix A. Supplementary data
Supplementary data to this article can be found online at https://doi.org/10.1016/j.jneo.2017.11.005.

References
Symptoms and Timeliness in Cancer Patients

Koo et al.

Neoplasia Vol. 20, No. xx, 2018


[38] Collins GS and Altman DG (2013). Identifying patients with undetected renal tract cancer in primary care: an independent and external


