

Validation of self-reported help-seeking, and measurement of the patient interval, for cancer symptoms: an observational study to inform methodological challenges in symptomatic presentation research

Sarah M. Smith

Katriina L. Whitaker

Amanda H. Cardy

Alison M. Elliott

Philip C. Hannaford

Peter Murchie

This is a pre-copyedited, author-produced version of an article accepted for publication in Family Practice following peer review. The version of record: Smith, S.M., et al. (2019) Validation of self-reported help-seeking, and measurement of the patient interval, for cancer symptoms: an observational study to inform methodological challenges in symptomatic presentation research.

is available online at DOI:

<https://doi.org/10.1093/fampra/cmz047>

Title: Validation of self-reported help-seeking, and measurement of the patient interval, for cancer symptoms: an observational study to inform methodological challenges in symptomatic presentation research:

Short title: Validation of self-reported help-seeking, and measurement of the patient interval, for cancer symptoms

Article Category: Health Service Research

Sarah M. Smith¹, Katriina L. Whitaker², Amanda H. Cardy¹, Alison M. Elliott³, Phil C. Hannaford¹, Peter Murchie¹

¹Institute of Applied Health Sciences, University of Aberdeen, Foresterhill, Aberdeen, AB25 2ZD, UK.

²School of Health Sciences, University of Surrey, Guildford, Surrey, GU2 7XH, UK.

³Graduate School, Abertay University, Bell Street, Dundee, DD1 1HG, UK.

Corresponding author: *Professor P. Murchie, Institute of Applied Health Sciences, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, UK. (p.murchie@abdn.ac.uk)*

Disclosures: No conflicts of interest were reported by the authors of this paper.

Key messages:

- Careful survey design offers reasonably accurate data on medical help-seeking
- General practice records are unreliable for measuring the patient interval
- There is arguably a need to standardise the way the patient interval is recorded

ABSTRACT

Background

To improve earlier presentation with potential symptoms of cancer accurate data is needed on how people respond to these symptoms. It is currently unclear how self-reported medical help-seeking for symptoms associated with cancer by people from the community correspond to what is recorded in their general practice records, or how well the patient interval (time from symptom onset to first presentation to a health-professional) can be estimated from patient records.

Method

Data from two studies that reviewed general practice electronic records of residents in Scotland, 1) the 'Useful Study': respondents to a general population survey who reported experiencing symptoms potentially associated with one of four common cancers (breast, colorectal, lung, and upper gastro-intestinal) 2) the 'Detect Cancer Early' programme: cancer patients with one of the same four cancers. Survey respondents' self-reported help-seeking (yes/no) was corroborated; Cohen's Kappa assessed level of agreement. Combined data on the patient interval was evaluated using descriptive analysis.

Results

'Useful Study' respondents' self-report of help-seeking showed exact correspondence with general practice electronic records in 72% of cases (n=136, kappa 0.453, moderate agreement). Between both studies, 1269 patient records from 35 general practices were reviewed. The

patient interval could not be determined in 44% (N=809) of symptoms presented by these individuals.

Conclusions

Patient self-report of help-seeking for symptoms potentially associated with cancer offer a reasonably accurate method to research responses to these symptoms. Incomplete patient interval data suggests routine general practice records are unreliable for measuring this important part of the patient's symptom journey.

MeSH: Cancer symptoms; community surveys; electronic medical records; health seeking behaviour, primary care.

Introduction

Earlier diagnosis is strongly advocated as a means to address the relatively poor five-year cancer survival in the UK.^{1,2,3} As primary care is the first point of contact for most people with symptoms possibly indicative of cancer,⁴ general practitioners play an important role in the early identification of patients with cancer.⁵ Also, much of our knowledge surrounding symptoms of cancer is based on what is recorded in primary care records.^{6,7,8} However, while this method provides valuable information it has tended to be restricted to those individuals who already have cancer, and those who have chosen to consult their general practitioner prior to diagnosis.⁹ Correspondingly, community-based surveys have been effective in identifying what specific symptoms of cancer trigger reports of help-seeking behaviour in the general population.^{10,11} However, it is not clear 1) how reliable self-reporting of help-seeking in primary care for symptoms potentially associated with cancer compares with what is documented in general practice (GP) records, 2) how easily the time between symptom onset and first presentation to a medical practitioner, known as the patient interval,^{12,13} can be ascertained from GP records. This information is fundamental for researchers investigating responses to symptoms and the patient interval, and also for assessing the success of interventions to prompt earlier presentation by patients with potential cancer symptoms.

This paper draws on data from two studies, the 'Useful Study', a community-based survey, and the 'Detect Cancer Early' programme, a cancer patient service evaluation. Both studies examined GP electronic records to explore information documented on the patient interval for symptoms

associated with four common cancers (breast, colorectal, lung and upper gastro-intestinal (GI)). Data from the 'Useful Study' GP record review has also been used to corroborate patient self-report of primary care help-seeking/no help-seeking in response to these symptoms.

Methods

The 'Useful Study' GP record review

'Useful Study' participants

This GP record review is the final phase of the 'Useful Study', a three-phase mixed-method prospective study conducted to better understand behavioural responses in the community to symptoms potentially associated with cancer. The study involved people aged 50 and over as they are at a higher risk of cancer than younger patients. Twenty-one symptoms associated with breast, colorectal, lung, and upper GI cancer were included in the Phase 1 questionnaire survey along with four symptoms intended to mask the purpose of the study (Box 1). Although the survey asked about symptom experiences throughout the questionnaire, reference to cancer was not made to avoid causing undue anxiety,¹⁴ or influencing responses.¹⁰

'Useful study' inclusion/exclusion criteria

Individuals eligible for the GP medical record review were sampled from respondents to the 'Useful Study' questionnaire survey and who agreed that their GP medical records could be examined, (70%, N=12744/16778). Respondents also needed to have reported experiencing at least one new potential cancer symptom in the previous year, with the same symptom present

in the previous month. Ten per cent of respondents met these additional criteria (1651/16778) for inclusion in the GP record review.

Next, eleven practices (four in England, seven in Scotland) containing 984 of the 1651 respondents were selected to incorporate a good spread across deprivation, rurality and geographical area. Further exclusion criteria were then applied for any of these 984 respondents who had masking symptoms unlikely to be related to cancer (Box 1), or who had substantial missing data on survey questions (e.g. symptom occurrence, help-seeking actions, and demographics) as this would limit our ability to corroborate self-reported primary care consultations with symptoms potentially associated with cancer. This resulted in 66% (N=651/974) of respondents from the 11 selected practices being eligible for the GP record review (60% in Scotland (N=390)). To reduce travel time and costs the review was restricted to the seven practices in Scotland.

BOX 1.

(a) Symptoms associated with breast, colorectal, lung and upper GI cancer included in the questionnaire:

Persistent indigestion/heartburn, difficulty swallowing, stomach or abdominal pain, chest pain, hoarseness, loss of appetite, unexplained weight loss, persistent cough, change in ongoing cough, persistent diarrhoea, coughing up phlegm, coughing up blood, shortness of breath, change in bowel habits, blood in stool or rectal bleeding, persistent vomiting, vomiting up blood, lump in breast, breast change other than lump, tired all the time.

(b) Masking symptoms:

Headaches, wheezy chest, change in bladder habits, back or joint pain.

'Useful Study' data sample

To reach a target of 250 GP records of survey respondents in Scotland, a sample of 300 was required (to allow for missing records) and selected to incorporate an even spread of rurality, deprivation.

'Useful Study' Data collection

For each eligible participant data on all consultations during the 12 months prior to, and the 12 months following, the questionnaire survey was collected from their GP electronic medical record (January 2014 to December 2016) by AC and two other NHS Research Scotland Primary Care network employees. Data collected included: date patient first presented with symptom (s); type of symptom; information on the patient interval. Sociodemographic characteristics were obtained from responses to the questionnaire survey and included age, sex, smoking status, rurality,¹⁵ and socio-economic status (defined by the Scottish Index of Multiple Deprivation (SIMD) comprising 5 indicators).¹⁶

The 'Detect Cancer Early' Programme GP record review

'DCE' participants

The GP electronic record review formed part of an evaluation in northeast Scotland of the impact of the Scottish Government's 'Detect Cancer Early' (DCE) campaign where information was retrospectively extracted on the diagnostic pathway of 2,102 adult cancer patients (≥ 18 years) from 31 GP practices in the NHS Grampian and NHS Shetland regions, who were diagnosed with

one of six common cancers (breast, colorectal, lung, melanoma, prostate and upper GI) between January 2007 and October 2013.

'DCE' Data collection

The GP review was carried out by SS (October 13 to January 2015). Comprehensive data were collected on consultations with relevant symptoms¹⁷ for two years prior to referral or admission. This included the date of first presentation with symptoms and, where available, the date of symptom onset, or any information on when the patient first became aware of symptoms. This allowed information on the patient interval to be gathered for each patient.

'DCE' Inclusion/exclusion criteria

For assessing the patient interval, patients were excluded if they were diagnosed through screening, if their first presentation with symptoms was not to a general practitioner, or if their cancer was an incidental finding.

'Useful Study' and 'DCE' data management

Both GP electronic record reviews were conducted by skilled researchers who held honorary NHS contracts and used password protected laptop computers. Proformas were developed in Microsoft Excel ('Useful Study') and Microsoft Access ('DCE') prior to the reviews to ensure standardised data collection. No personal identifiable information was collected during either review. Using a unique ID number, data from the 'Useful Study' review was subsequently merged with related data from the questionnaire survey and transferred into SPSS version 24 for analysis.

Similarly, the 'DCE' database and relevant data from the NHS Grampian secondary care Cancer Care Pathways database, available in Microsoft Excel, were exported into SPSS version 23.0 and linked by a unique patient ID number to create a unified dataset. To align with the 'Useful Study', anonymised data on the patient interval for breast, colorectal, lung, and upper GI cancer was then extracted from the unified 'DCE' database and analysed using SPSS version 24.

A more comprehensive methodology for the 'DCE' study has been described elsewhere.¹⁸

The 'Useful Study' and 'DCE' Analysis

Data on the patient interval for symptoms potentially associated with breast, colorectal, lung and upper GI cancer from the 'DCE' study and the 'Useful Study' were combined and descriptive analysis carried out using SPSS version 24. The ease with which it was possible to determine the patient interval from GP records was placed in one of four categories: "straightforward", "straightforward with assumptions", "impossible to determine", and "no information available", examples are given in box 2. Data from the 'Useful Study' questionnaire survey responses (yes/no) on whether a general practitioner had been consulted with symptoms potentially associated with cancer in the previous year were assessed for corroborative evidence within the GP records and any differences documented. Cohen's Kappa statistic was calculated to quantify level of agreement between these two sources of data. A summary box of both studies is provided (box 3).

Box 2.**Examples of how the ease of determining the patient interval was categorised**

(a) Straightforward	Exact date given 00/00/00 'three months' 'one week' 'forty-eight hours' 'past two weeks' 'since Christmas day' 'since the Queens Jubilee'
(b) Straightforward with assumptions	'a few weeks' = at least three weeks 'for over a month now' = at least five weeks 'last few months' = at least three months 'several months' = at least four months
(c) Impossible to determine	'a history of' 'lately' 'again' 'since discharge' 'worse the last few days' 'ongoing' 'current' 'on and off for years' 'periodic' 'a while' 'for some time' 'chronic' 'increased over time'
(d) No information available	No mention of the patient interval in GP records

Box 3.**The 'Useful Study' Summary**

Background: To improve earlier presentation with potential symptoms of cancer, accurate data is needed on how people from the community respond to these symptoms.

Purpose: To explore available information on the patient interval and corroborate self-reported help-seeking behaviour for potential symptoms of four common cancers (breast, colorectal, lung, upper gastro-intestinal).

Structure: A GP review of respondents to the 'Useful Study' questionnaire survey who agreed to their records being reviewed. Validation of participant self-report of having experienced symptoms possibly indicative of cancer, and their response to these symptoms (help-seeking/no help-seeking). Information on the patient interval was documented.

THE 'DCE' Study Summary

Background: Late stage diagnosis is a contributory factor to Scotland's poor five-year cancer survival rates. To enable earlier stage diagnosis and treatment, Scottish Government health policy has emphasised the importance of early presentation of symptoms and early referral to secondary care.

Purpose: To assess how well the patient interval is documented in medical records as a prolonged patient interval can be linked to later stage diagnosis for cancer patients.

Structure: A GP review as part of a service evaluation of the diagnostic pathway of patients with four common cancers (breast, colorectal, lung, upper gastro-intestinal) that included the patient interval.

RESULTS

The Useful Study

After four general practices in three Scottish NHS health boards (Highland, Grampian, and Lothian) had been visited and 188 participant records had been reviewed, a decision was taken to halt data collection. This was because it had become clear at that point that relatively few participants had sufficient information documented which enabled the patient interval to be estimated.

Socio-demographic characteristics

Comparison of socio-demographic characteristics of respondents to the questionnaire survey and the GP medical records review sample showed a similar balance of gender, age and smokers, but with more remote rural participants in the GP medical records review and more affluent participants in the main survey respondents (table 1).

Corroboration of survey data with GP medical records data

Corroboration of survey data with the GP records for participants in the 'Useful Study' showed that participant recall of consulting with symptoms potentially associated with cancer in the year prior to completing the survey agreed in 72% of cases (N=136; kappa statistic 0.453, moderate agreement), i.e. there was corroborative evidence that they had or had not consulted their general practitioner with such symptoms (table 2). Reported consulting behaviour was not corroborated in 17% of cases (N=32), i.e. 13% (N=25) reported they had consulted with symptoms for which there was no record in the notes, and 4% (N=7) had a note of having consulted when

the respondent reported they had not. For the remaining 10% of patients (N=19) with two or more symptoms, there was corroborative evidence for some but not all symptoms.

Ease of determining the patient interval - the 'Useful Study' and the 'DCE' study

Over the two-year study period, 99 respondents in the 'Useful Study' had a first presentation with 101 symptoms potentially associated with cancer, and in the 'DCE' study 1170 cancer patients had a first presentation with 1754 related symptoms. Both studies showed similar findings in each assessment category for the patient interval. In total there were 1269 cases of first presentations with 1855 symptoms associated with breast, colorectal, lung and upper GI cancer. The ease with which the patient interval could be ascertained was "straightforward" in 48% (N=889) of symptoms, "straightforward with assumptions" in 9% (N=157) of symptoms and "impossible to determine" in 9% (N=166), with "no information available" for 35% (N=643) of symptoms (table 3).

Discussion

Summary of main findings

To our knowledge, this study is the first to validate survey respondents' self-report of help-seeking/no help-seeking, for symptoms potentially indicative of cancer, with GP record data in a community-based sample. Our data suggest that most people, when asked about help-seeking in response to these symptoms, respond with reasonable accuracy. They also show that GP records are often a poor source of information for estimating the patient interval. This may be because patients have difficulty in pinpointing when symptoms first began or the information

was not ascertained or documented. The ease of determining the patient interval varied according to the nature of the symptom, this may be due to a greater awareness of a possible link with cancer for some symptoms than others.¹⁹ Given the increasing reliance on GP medical records for research and policy purposes,⁹ there is arguably a need for a standardised method of documenting the patient interval for symptoms potentially associated with cancer.

Strengths and limitations

A key strength of this paper was the ability to combine two data sets, each with similar findings, to produce a larger sample size for exploring the documentation of information related to the calculation of the patient interval. We collected information from 35 practices over a broad span of the population in terms of socio-economic status and geography, thereby increasing the generalisability of the findings. The prospective nature of the 'Useful Study' allowed corroboration of the self-reporting of help-seeking actions for symptoms potentially associated with cancer by the general population, this is the first study to do so. However, due to the small sample size in this part of the study these findings should be interpreted with caution as they may not reflect the help-seeking actions of the entire population.

The standardized data collection forms used by study researchers in both studies helped to optimize the quality of data available for analysis.

Comparison with existing literature

Studies exploring concordance of patient records with patient self-report showed variable results. Mant et al,²⁰ investigating the accuracy of data in GP records compared with data from patient questionnaires on smoking use and alcoholic intake, found agreement between the two sources was moderate for both smoking (kappa = 0.50) and alcohol use (kappa = 0.52). On the other hand, Barbara et al.²¹, exploring concordance between self-reported medical visits for respiratory symptoms and GP medical records among 176 participants in an influenza prevention study, reported a range of kappa statistics for different symptoms (from 0.05 for chills, to 0.41 for cough and 0.51 for earache). Only fair agreement was shown (kappa=0.25) between diabetic patients' reports of attending for annual eye examinations and GP medical records, with 64.6% discordance because of lack of evidence in the medical record to support self-report.²² Variation in agreement may be due to differences in the design of questionnaires in each study and/or insufficient information documented in medical records.

We found few studies reporting on the patient interval and even fewer exploring how easily the patient interval could be determined from GP records. This is in line with Koshiaris et al.²³ who describe the lack of studies on the patient interval as a limitation to their systematic review and meta-analysis to quantify time intervals to diagnosis of myeloma. One study in England analysed National Audit data of more than 10,000 cancer patients with 28 different cancers²⁴ and reported a variation in missing interval data (patient and primary care) that ranged from 28% to 49% depending on the type of cancer. The authors suggest that commitment to regular collection of pre-referral interval data is critical for evaluating and monitoring interventions aimed at increasing the earlier diagnosis of cancer.

Related to this are the concerns about the general accuracy and completeness of data in electronic primary care records.^{9, 25,26,27} For example, imprecise coding may have been why self-report overestimated the number of patients visiting an eye care provider in the previous year and underestimated the prevalence of eye disease.²⁸ Also, it is suggested that the increasing reliance on clinical coding within electronic GP medical records may affect the collection and recording of detailed information provided by patients^{29,30,31} which may include the patient interval.

Our observations support the need for efforts to improve the routine and accurate recording of the patient interval in GP medical records. However, there are practical difficulties to consider before this could feasibly occur since eliciting the patient interval for each symptom potentially associated with cancer from patients with multiple relevant symptoms could prove too time-consuming. Even so, it is clear that poor or variable quality of data in medical records needs to be better understood and effective interventions for improvement developed.³² One solution suggested to increase data quality in medical records has been to make data more accessible to patients which might act as an impetus to increase data quality.³³ The dynamic nature of care-recording³³ together with the adoption of new approaches will hopefully lead to the patient interval being recorded in a more standardised way.

Conclusion

The moderate agreement found between self-completed questionnaire responses and primary-care held records suggests that carefully designed questionnaires can be used to obtain reasonable quality data about medical help-seeking in primary care. Caution should be exercised, however, when using primary care records to estimate the patient interval, whether for audit or research purposes. Given the increasing importance of patient intervals in health-care research we suggest that if initiatives to standardise data recording were to include the patient interval it would strengthen the possibility of the GP electronic data systems being configured to deliberately collect this information.

Declarations: Ethical approval for the 'Useful Study' study was obtained by the National Research Ethics Committee (NRES) East Midlands Derby on 14/08/2014 (14/EM/1124). NHS Research and Development approval was also obtained prior to conducting the study.

Anonymised data were collected for the 'DCE' study as part of the 'Detect Cancer Early' project, an NHS Grampian service improvement exercise. The study was reviewed by the Caldicott Guardian for NHS Grampian and approved by the North of Scotland Research Ethics Committee, who confirmed that the project was a service evaluation and that there was no requirement to contact individual patients prior to review of their medical records.

Funding: The 'Useful Study' was funded by a Cancer Research UK National Awareness and Early Diagnosis Initiative project award (C45810/A17927). The 'DCE' Study was funded by the Chief Scientist Office of the Scottish Government.

Conflict of interest: The authors declare no conflict of interest

Contributorship: AME, PM, KLW and PCH conceived and conducted the study. AC co-ordinated the collection of the case-note data. SM and PM analysed the data with support from AE and PCH. All authors contributed to writing the manuscript.

Acknowledgments: The authors wish to thank the patients who participated in the study and their respective primary-care practices. They also wish to acknowledge the data-co-ordinators from NHS Research Scotland Primary Care who collected the case note review data. They would also like to acknowledge contributions to this aspect of the 'Useful Study' by Alison Thornton and Susan Hall from the Academic Primary Care Research Group, University of Aberdeen.

Data Statement: It is recognized that data from the 'Useful Study' could support other future researchers in understanding symptom prevalence and people's reactions to them in the

community. The 'Useful Study' data are available they will be stored in an accessible data repository at the University of Aberdeen in full accordance with Cancer Research UK's guidance on data-sharing. Subsequent sharing requests will be made to the PI and granted only upon careful consideration of the specific case made for why the data is to be shared, how it is to be used and the likely outcome of its use. Decisions to share data will only be taken in the light of advice from the regulatory bodies approving the study and from Research and Innovation, University of Aberdeen. Where a decision is taken to share data, we will ensure secure transfer protocols are used and that optimal security will be employed by the data recipient.

References

- 1) Neal RD, Tharmanathan P, France B, et al. Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *Br J Cancer*. 2015;112 Suppl 1(Suppl 1):S92–S107. Published 2015 Mar 3. doi:10.1038/bjc.2015.48.
- 2) Richards MA. The National Awareness and Early Diagnosis Initiative in England: assembling the evidence. *Br J Cancer*. 2009;101 Suppl 2(Suppl 2):S1–S4. doi:10.1038/sj.bjc.6605382.
- 3) Scottish Government. Detect Cancer Early Programme, 2015. Available at: <https://www2.gov.scot/Topics/Health/Services/Cancer/Detect-Cancer-Early> Accessed October 2017
- 4) Green T, Atkin K, Macleod U. Cancer detection in primary care: insights from general practitioners. *Br J Cancer*. 2015;112 Suppl 1(Suppl 1):S41–S49. doi:10.1038/bjc.2015.41
- 5) Rubin G, Berendsen A, Crawford SM, et al. The expanding role of primary care in cancer control. *Lancet Oncol* 2015; 16(12):1231–1272.
- 6) Baughan P, O'Neill B, Fletcher E. Auditing the diagnosis of cancer in primary care: the experience in Scotland. *Br J Cancer* 2009; 101: S87-S91.
- 7) Stapley S, Peters TJ, Neal RD, Rose PW, Walter FM, Hamilton W. The risk of oesophago-gastric cancer in symptomatic patients in primary care: a large case-control study using electronic records. *Br J Cancer*. 2012;108(1):25–31. doi:10.1038/bjc.2012.551.
- 8) Hortedahl K, Vedsted P, Borgquist L, et al. 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*. 2017;3:e00328.
- 9) Shephard E, Stapley S, Hamilton W. The use of electronic databases in primary care research, *Fam Pract*. 2011;28(4):352–354, <https://doi.org/10.1093/fampra/cmr039>.
- 10) Whitaker KL, Scott SE, Winstanley K, Macleod U, Wardle J. Attributions of cancer 'alarm' symptoms in a community sample [published correction appears in *PLoS One*. 2015;10(3):e0118418]. *PLoS One*. 2014;9(12):e114028. Published 2014 Dec 2. Doi:10.1372/journal.pone.0114028.
- 11) Whitaker KL, Smith CF, Winstanley K, Wardle J. What prompts help-seeking for cancer 'alarm' symptoms? A primary care based survey. *Br J Cancer*. 2016;114(3):334–339. doi:10.1038/bjc.2015.445.
- 12) Weller D, Vedsted P, Rubin G, et al. The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. *Br J Cancer*. 2012;106(7):1262–1267. doi:10.1038/bjc.2012.68.
- 13) Scott SE, Walter FM, Webster A, Sutton S, Emery J. The Model of Pathways to Treatment: Conceptualization and integration with existing theory. *Br J Health Psychol*. 2013 02/01; 2019/04;18(1):45-65.
- 14) Brodersen J, McKenna SP, Doward LC, Thorsen H. Measuring the psychosocial consequences of screening. *Health Qual Life Outcomes* 2007; 5: 3.
- 15) Scottish Government. Scottish Government Urban Rural Classification 2016. <https://www.gov.scot/publications/scottish-government-urban-rural-classification-2016/pages/5/>. Accessed June 2017.

- 16) Scottish Government. The Scottish Index of Multiple Deprivation 2016. Available at: <http://simd.scotland.gov.uk/publication-2012/>. Accessed June 2017
- 17) Scottish Executive Health Department. *Scottish Referral Guidelines for Suspected Cancer*. SEHD, Edinburgh, 2007.
- 18) Murchie P, Smith SM, Yule MS, et al. Does emergency presentation of cancer represent poor performance in primary care? Insights from a novel analysis of linked primary and secondary care data. *Br J Cancer*. 2017;116(9):1148–1158. doi:10.1038/bjc.2017.71.
- 19) Davis LL. Determining Time of Symptom Onset in Patients With Acute Coronary Syndromes: Agreement Between Medical Record and Interview Data. *Dimens Crit Care Nurs*. 2015;34(4):222–231. doi:10.1097/DCC.000000000000117.
- 20) Mant J, Murphy M, Rose P, et al. The accuracy of general practitioner records of smoking and alcohol use: comparison with patient questionnaires. *J Public Health Med* 2000;22:198–201.
- 21) Barbara AM, Loeb M, Dolovich L, Brazil K, Russell M. Agreement between self-report and medical records on signs and symptoms of respiratory illness. *Prim Care Respir J*. 2012;21:145-152.
- 22) Beckles GLA, Williamson DF, Brown AF, Gregg EW, Karter AJ, Kim C, Dudley RA, Safford MM, Stevens MR, Thompson TJ. Agreement between self-reports and medical records was only fair in a cross-sectional study of performance of annual eye examinations among adults with diabetes in managed care. *Med Care* 2007;45:876-883. doi: 10.1097/MLR.0b013e3180ca95fa.
- 23) Koshiaris C, Oke J, Abel L, et al. Quantifying intervals to diagnosis in myeloma: a systematic review and meta-analysis. *BMJ Open* 2018;8:e019758. doi:10.1136/bmjopen-2017-019758.
- 24) Lyratzopoulos G, Saunders CL, Abel GA, et al. The relative length of the patient and the primary care interval in patients with 28 common and rarer cancers. *Br J Cancer*. 2015;112 Suppl 1(Suppl 1):S35–S40. doi:10.1038/bjc.2015.40.
- 25) Dungey S., Beloff N, Williams R, Williams T, Puri S, Tate AR. (2015) Characterisation of Data Quality in Electronic Healthcare Records. In: Briassouli A., Benois-Pineau J., Hauptmann A. (eds) *Health Monitoring and Personalized Feedback using Multimedia Data*. Springer, Cham. https://doi.org/10.1007/978-3-319-17963-6_7.
- 26) Nightingale AL, Davidson JE, Molta CT, Kan HJ, McHugh NJ. Presentation of SLE in UK primary care using the Clinical Practice Research Datalink. *Lupus Sci Med*. 2017;4(1):e000172. Published 2017 Feb 10. doi:10.1136/lupus-2016-000172.
- 27) Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. *BMJ*. 2003;326(7398):1070. doi:10.1136/bmj.326.7398.1070.
- 28) MacLennan PA, McGwin G, Searcey K, Owsley C. Medical record validation of self-reported eye diseases and eye care utilization among older adults. *Curr Eye Res*. 2012;38(1):1–8. doi:10.3109/02713683.2012.733054.
- 29) Ford E, Nicholson A, Koeling R, et al. Optimising the use of electronic health records to estimate the incidence of rheumatoid arthritis in primary care: what information is

hidden in free text?. *BMC Med Res Methodol*. 2013;13:105. doi:10.1186/1471-2288-13-105

- 30) Price SJ, Stapley SA, Shephard E, Barraclough K, Hamilton WT. 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*. 2016;6(5):e011664. doi:10.1136/bmjopen-2016-011664
- 31) Tate AR, Martin AGR, Ali A, *et al* Using free text information to explore how and when GPs code a diagnosis of ovarian cancer: an observational study using primary care records of patients with ovarian cancer *BMJ Open* 2011;1:e000025. doi: 10.1136/bmjopen-2010-000025
- 32) Chan, Kitty S., Jinnat B. Fowles, and Jonathan P. Weiner. "Review: Electronic Health Records and the Reliability and Validity of Quality Measures: A Review of the Literature." *Medical Care Research and Review* 67, no. 5 (October 2010): 503–27. doi:10.1177/1077558709359007.
- 33) Dungey S, Glew S, Heyes B, Macleod J, Tate AR. Exploring practical approaches to maximising data quality in electronic healthcare records in the primary care setting and associated benefits. Report of panel-led discussion held at SAPC in July 2014. *Primary Health Care Research & Development*. 2016;17(5):448-452. doi:10.1017/S1463423615000596

Table 1. The 'Useful Study' socio-demographic characteristics of the general practice medical record review subset and all questionnaire survey respondents (2014-2016)

Characteristics	General practice medical records review		Questionnaire survey
		Questionnaire survey subset	All respondents
		N (%)	N (%)
Gender	Male	93 (49)	7745 (46)
	Female	95 (50)	9033 (54)
	Total	188 (100)	16778 (100)
Age band	50-59	72 (38)	5564 (33)
	60-69	62 (33)	6350 (38)
	70-79	40 (21)	3712 (22)
	80+	14 (7)	1152 (7)
	Total	188 (100)	16778 (100)
Smoking status	Never smoker	87 (46)	8872 (53)
	Ex-smoker	82 (44)	6095 (36)
	Current smoker	19 (10)	1471 (9)
	Total	188 (100)	16438 (100)
Rurality	Large Urban Areas	62 (33)	3234 (37)
	Other urban areas	0 (0)	1434 (16)
	Accessible Small Towns	20 (11)	640 (7)
	Remote Small Towns	35 (17)	1569 (18)
	Accessible Rural	20 (11)	959 (11)
	Remote Rural	51 (27)	1013 (11)
	Total	188 (100)	8849 (100)
	SIMD*	1	11 (6)
2		23 (12)	733 (8)
3		58 (31)	1856 (21)
4		76 (40)	2363 (27)
5		20 (11)	3373 (38)
Total		188 (100)	8849 (100)

Table 2. The ‘Useful Study’ corroboration of survey data with general practice medical records on self-reported consultation history with symptoms associated with cancer (2014-2016)

Agreement/disagreement	N (%)
Agreement of general practice medical records with survey data <i>(i.e. ‘yes’ I did consult, ‘no’ I did not consult)</i>	136 (72)
Disagreement of general practice medical records with survey data	32 (17)
<i>General practice medical records showed no consultation, survey data ‘yes’ I did consult</i>	25 (13)
<i>General practice medical records showed patient had consulted, survey data ‘no’ I did not consult</i>	7 (4)
Agreement with general practice medical records for some symptoms, disagreement for others	19 (10)
Missing survey data	1 (<1)
Total	188 (100)
Kappa measurement of agreement	0.453
Standard error	0.065

Table 3. Ease of determining the patient interval with symptoms associated with cancer based on information held in general practice medical records from the 'Useful Study' (2014-2016) and the 'Detect Cancer Early' study (2013-2015)

Symptom type	Straightforward N (%)	Straightforward with Assumptions N (%)	Impossible to determine N (%)	No information N (%)	Total
Blood in stool or rectal bleeding	53 (43)	24 (19)	8 (6)	39 (31)	124
Breast change other than a lump	47 (38)	6 (5)	13 (10)	58 (47)	124
Change in bowel habit	84 (47)	11 (6)	23 (13)	60 (34)	178
Change in ongoing cough	1 (100)	0 (0)	0 (0)	0 (0)	1
Chest pain	24 (41)	10 (17)	1 (2)	23 (40)	58
Coughing up blood	16 (62)	1 (4)	2 (8)	7 (27)	26
Coughing up phlegm	31 (40)	2 (3)	11 (14)	32 (42)	76
Difficulty swallowing	43 (61)	10 (14)	6 (9)	11 (16)	70
Hoarseness	8 (53)	1 (7)	0 (0)	6 (40)	15
Loss of appetite	24 (43)	6 (11)	2 (4)	24 (42)	56
Lump in breast	212 (66)	21 (7)	24 (8)	63 (20)	320
Persistent cough/cough	93 (51)	11 (6)	14 (8)	66 (36)	184
Persistent diarrhoea	19 (66)	3 (10)	4 (14)	3 (10)	29
Persistent indigestion/ heartburn	19 (31)	7 (11)	8 (13)	28 (45)	62
Persistent vomiting	13 (72)	2 (11)	0 (0)	3 (17)	18
Shortness of breath	42 (30)	8 (6)	27 (19)	64 (45)	141
Stomach or abdominal pain	110 (55)	20 (10)	11 (6)	58 (29)	199
Tired all the time	15 (22)	5 (7)	6 (9)	43 (62)	69
Unexplained weight loss	33 (32)	9 (9)	6 (6)	55 (53)	103
Vomiting up blood	2 (100)	0 (0)	0 (0)	0 (0)	2
Total	889 (48)	157 (9)	166 (9)	643 (35)	1855 (100)