

BMJ Open Cohort profile: prescriptions dispensed in the community linked to the national cancer registry in England

Katherine E Henson,¹ Rachael Brock,¹ Brian Shand,¹ Victoria H Coupland,¹ Lucy Ellis-Brookes,¹ Georgios Lyratzopoulos,^{1,2} Philip Godfrey,³ Abigail Haigh,³ Kelvin Hunter,⁴ Martin G McCabe,^{1,5} Graham Mitchell,³ Nina Monckton,³ Robert Robson,³ Thomas Round,⁶ Kwok Wong,¹ Jem Rashbass¹

To cite: Henson KE, Brock R, Shand B, *et al*. Cohort profile: prescriptions dispensed in the community linked to the national cancer registry in England. *BMJ Open* 2018;**8**:e020980. doi:10.1136/bmjopen-2017-020980

► Prepublication history for this paper is available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-020980>).

Received 11 December 2017
Revised 9 March 2018
Accepted 1 June 2018



¹National Cancer Registration and Analysis Service, Public Health England, London, UK

²Department of Behavioural Science and Health, ECHO (Epidemiology of Cancer Healthcare and Outcomes) Group, University College London, London, UK

³NHS Prescription Services, NHS Business Services Authority, Newcastle upon Tyne, UK

⁴Department of Medicine (Cambridge University), Addenbrooke's Hospital, Cambridge, UK

⁵Institute of Cancer Sciences, Manchester Academic Health Science Centre, Core Technology Facility, University of Manchester, Manchester, UK

⁶Department of Primary Care and Public Health Sciences, King's College London, London, UK

Correspondence to

Dr Katherine E Henson;
katherine.henson@phe.gov.uk

ABSTRACT

Purpose The linked prescriptions cancer registry data resource was set up to extend our understanding of the pathway for patients with cancer past secondary care into the community, to ultimately improve patient outcomes.

Participants The linked prescriptions cancer registry data resource is currently available for April to July 2015, for all patients diagnosed with cancer in England with a dispensed prescription in that time frame. The dispensed prescriptions data are collected by National Health Service (NHS) Prescription Services, and the cancer registry data are processed by Public Health England. All data are routine healthcare data, used for secondary purposes, linked using a pseudonymised version of the patient's NHS number and date of birth. Detailed demographic and clinical information on the type of cancer diagnosed and treatment is collected by the cancer registry. The dispensed prescriptions data contain basic demographic information, geography measures of the dispensed prescription, drug information (quantity, strength and presentation), cost of the drug and the date that the dispensed prescription was submitted to NHS Business Services Authority.

Findings to date Findings include a study of end of life prescribing in the community among patients with cancer, an investigation of repeat prescriptions to derive measures of prior morbidity status in patients with cancer and studies of prescription activity surrounding the date of cancer diagnosis.

Future plans This English linked resource could be used for cancer epidemiological studies of diagnostic pathways, health outcomes and inequalities; to establish primary care comorbidity indices and for guideline concordance studies of treatment, particularly hormonal therapy, as a major treatment modality for breast and prostate cancer which has been largely delivered in the community setting for a number of years.

INTRODUCTION

National healthcare systems make it more feasible to routinely collect detailed healthcare data for the entire population, which are vital for population health research. For example, in the National Health Service

Strengths and limitations of this study

- The linkage of community-dispensed prescriptions data to cancer registration data can greatly enhance our understanding of the patient pathway, as it can provide novel insight into symptom profiles and potentially identify patterns which could indicate opportunities for the earlier diagnosis of cancer.
- The key strength of this data source is its population coverage, and the pseudonymisation process strengthens its value, as it has permitted the existence of an anonymised national control population for comparison purposes, to understand how patients with cancer differ from the general population.
- Sex is not available in the prescriptions data, therefore cannot be used to match controls of non-cancer patients.
- The indication for the dispensed drug is not included in the data which may be problematic for drugs with multiple indications, for example, antidepressants can be indicated for depression or chronic pain.
- For pharmacovigilance studies, one should be aware that dispensed drug information does not guarantee that the patient has taken the medication.

(NHS) in England, Public Health England (PHE) holds cancer and other disease data; NHS Digital collects hospital activity¹ and other process data, while the NHS Business Services Authority (NHSBSA) gathers administrative data for payments of services. NHSBSA gathers this through NHS Prescription Services, which calculates the remuneration and reimbursement owed to dispensing pharmacy contractors in the community across England, as per the Drug Tariff.²

The resulting data resource collected by NHSBSA is vast. Over one billion prescription items were dispensed in England in 2014, an increase of over 55% since 2004 at a cost of nearly £9 billion per year.³ As in many developed countries, prescribing rates have

continued to rise in England, largely due to an increasingly old and multimorbid population.⁴ Therefore, the importance of collecting and using such data is increasing.

Since April 2015, NHSBSA expanded the dataset to include NHS number, which is the primary patient identifier in England. This has transformed the data, allowing linkage to other health data, for example, national cancer registration data.

The National Cancer Registration and Analysis Service (NCRAS) in PHE is responsible for collecting data on all cases of cancer and certain benign tumours that occur in people living in England.⁵ Hospital trusts submit multiple feeds of electronic health data, which are processed and combined by trained registration officers into a clinically comprehensive record for each tumour. The majority of cancer care is delivered in secondary care; therefore, the data collected reflect this.

A Data Sharing Agreement was established between PHE and NHSBSA, for NHSBSA to supply pseudonymised dispensed prescriptions data to PHE, providing hitherto missing information on this aspect of community healthcare for patients with cancer. We aim to describe the dispensed prescriptions data, the linkage and key data quality implications, with a focus on the linked data available within PHE.

COHORT DESCRIPTION

Data collected

The dispensed prescriptions data are collected from the Electronic Prescribing Service (EPS) and captured from the paper FP10 prescription forms using high speed scanners and character recognition software, with human input when required. FP10 forms are the legal prescription forms issued to patients in England, which they take to their chosen pharmacy. As of summer 2017, over half of the prescriptions dispensed in the community in England were issued via the EPS service.⁶ This has increased since 2015. Both EPS and FP10 forms are legal prescriptions, which are usually issued by a doctor, but may also be issued by a nurse, pharmacist prescriber or supplementary prescriber. The data are collected from prescriptions dispensed at a community pharmacy. Prescriptions are not filled at community pharmacies in the following scenarios: drugs supplied during a hospital stay, including those filled at discharge; drugs provided by private healthcare institutions; drugs supplied during a stay at a hospice or drugs supplied by other healthcare institutions, for example, urgent care centres.

Demographic, prescriber and drug details are collected for the dispensed prescriptions, including quantity, strength (dose) and presentation (tablet, injection etc). The full drug item is recorded as per the British National Formulary (BNF) V.68.⁷ Exemptions to this classification are detailed in the BNF Classification booklet published by NHSBSA,⁸ including additional items not covered by the BNF, for example, dressings and appliances. A full

data dictionary is shown in [table 1](#), with the percentage completeness for each data item.

Dispensed prescriptions data are available from April 2015, with data for April to July 2015 currently linked to cancer registration data. Data from August 2015 onwards will be available in due course, with updated linked data available in 2018. The objective is to link updated prescriptions data to the cancer registry data on a quarterly basis, with an approximate lag to real time of 6 months. During April to July, 332 655 118 dispensed prescription items among 29 481 344 individuals are identified. Ten per cent of the dispensed items and 6% of the individuals were linked to the cancer registry data, and therefore had a history of cancer (restricting to malignant tumours excluding non-melanoma skin cancer [ICD-10 (International Statistical Classification of Diseases and Related Health Problems 10th Revision) C00-C97 excl. C44] diagnosed after 1994). Using this linked data, 90 840 patients were identified as being diagnosed with a malignant tumour during April to July 2015, and had a prescription dispensed during the same period. Among this cohort, 99% of their dispensed prescriptions were issued by a General Practitioner (GP) practice⁹ (see [table 2](#)).

Pseudonymisation process

The partnership between NCRAS and NHSBSA has allowed NCRAS to obtain prescribing dispensed in the community information for both patients with and without cancer. The information on (anonymous) individuals without cancer can be used for analytical purposes as 'controls'. It was required to be in an anonymised format because NCRAS does not have the legal permissions to hold identifiable data on patients without a cancer diagnosis. Therefore, a secure pseudonymisation process was developed within NCRAS to allow the data linkage to be performed securely, without requiring the individual identities of patients to be shared.

The stages performed by both NHSBSA and NCRAS of the data extraction, pseudonymisation and linkage are detailed in [figure 1](#). The pseudonymisation procedure was run at source by NHSBSA on the dispensed prescriptions data, and the pseudonymised data were sent to NCRAS on an encrypted disk. The same pseudonymisation procedure was also run by NCRAS on the identifiers of patients with cancer stored in the cancer registry. A script was then run which linked the pseudonymised dispensed prescriptions data with patient records in the cancer registry using the pseudonym identifiers. A subset of the pseudonymised row-level dispensed prescriptions data linked to the cancer registry data was therefore securely created.

The pseudonymisation procedure uses standard third-party encryption and hashing modules (a modified version of the OpenPseudonymiser approach¹⁰). Fields which identify a patient are encrypted (hashed), and these encrypted values replace the original. The fields used are a patient's NHS number and their date of birth. The pseudonymisation procedure is run at source by NHSBSA on the prescriptions data. The procedure used

Table 1 Description of data items included, with data quality completeness of linked dispensed prescriptions cancer registry data, for patients diagnosed with malignant cancer (excluding non-melanoma skin cancer) after 1994 with a dispensed prescription record during April to July 2015 (n=1 680 764 patients and 33 669 294 prescription items)

Data item	Description	Completeness* (%)
Pseudonymised patient identifier	Code derived from patient NHS no and DOB	
Month	The month for which the claim for dispensed items has been submitted to NHSBSA.	100
Prescribed date	The date on which the prescription item was prescribed and submitted to EPS by the prescriber organisation. Note: only available for EPS messages	27
Prescribers postal code	The postal code of the prescribing practice, as held by the NHSBSA master database and the NHS Digital Organisation Data Service (ODS). It is in text and numerical format as per standard postal codes.	27
Primary care organisation (PCO) code	The code allocated by ODS to the prescriber's Clinical Commissioning Group or other PCO. It is a five-digit code containing numbers and letters.	99.7
PCO name		100
Practice code	The code of the general practice that carried out the prescribing, or other cost centre which is coded as a general practice in the NHSBSA master database. The code is allocated by NHSBSA.	>99.9
Practice name	It is a six-digit code containing numbers and letters.	>99.9
Net ingredient cost	This is the cost (in pence) of the dispensed item calculated from the reimbursement price held on the NHSBSA master database. The reimbursement price is the basic price for the quantity supplied of those drugs, appliances and chemical reagents as listed in parts VIII or IX of the Drug Tariff. For a drug not listed in part VIII, the price is normally determined by the list price published by the manufacturer, wholesaler or supplier of the pack size used for a prescription of that quantity.	100
Prescribed quantity	A specific quantity prescribed of a drug or appliance. Note: only available for EPS messages	100
No of items	This shows the no of times a product appears on a prescription form or EPS message.	100
Unit of measure	A description of the physical entity that can be handled, that is, tablet, ampoule, phial, mL. The unit of measure is held by the Dictionary of Medicines and Devices (dm+d) where this is appropriate to the product.	100
Pay quantity	The quantity of a drug or appliance which the NHSBSA has reimbursed. Where the product is packed in a special container, the quantity may be expressed as the no of special containers dispensed.	100
Drug presentation paid	A description of the product reimbursed as held by the dm+d. Where this is an actual product, the description includes the name of the proprietary or generic item, along with the supplier/manufacturer name. Where the product is a prescribable generic, the description of the generic product without supplier or trade name information is given. Note: on many occasions this will be the same as that prescribed.	100
BNF code	The NHSBSA database allocates BNF codes to drugs and medical devices: these codes are structured according to a series of hierarchical levels. Each level provides information about either the therapeutic use or pharmacological group of the drug or appliance. The structured coding is based on the BNF, with drug and appliance entries being arranged according to the BNF chapters, plus additional pseudo-chapters created by the NHSBSA.	>99.9
Patient age	The age of the patient, in years, on the date the item was prescribed.	82
Exemption category	The exemption category declared by the patient on the prescription form or included in the EPS message. Where the age/DOB is computer generated on the prescription, this is used for age exemption purposes.	
	The exemption category declared by the patient on the prescription form (note that this can be completed in prescription items from EPS messages).	99
	The exemption category included in the EPS message.	27
	Indicator whether exemption category was declared by the patient on the prescription form or included in the EPS message.	100

*The number of dispensed prescription items with a known value as a percentage of all dispensed prescription items (both FP10 and EPS). BNF, British National Formulary; DOB, date of birth; EPS, Electronic Prescribing Service; NHSBSA, NHS Business Services Authority; NHS, National Health Service.

by both NHSBSA and NCRAS is identical. This allows us to match the pseudonyms in the dispensed prescriptions data, where possible, with those in the cancer registry without the need to reveal the original patient details. Current security technology shows that a 'brute force' approach could theoretically be used to decrypt the data, but in practice the probability is minuscule,¹¹ therefore, in practice it is not possible to reidentify NHS numbers.

Data quality

NHS Prescription Services reprocess a random sample of 50 000 dispensed prescription items a month to assess the accuracy of these data. As of December 2016, the processing accuracy was 99.5%.¹²

Preliminary analyses of the linked data highlighted a number of data quality issues. A number of these issues relate to the FP10 forms, which account for 73% of the items in the April to July 2015 data. In these items, the date information is the month that the claim was submitted to NHSBSA, rather than the date that the drug was dispensed. There are therefore cases of inconsistent date orders, for example, the date of patient death is before the date of prescription for 0.3% of dispensed prescription items. Among the FP10-dispensed prescription items, age at prescription is missing for 19% (in the unlinked data). However, the age at prescription can be approximated for patients with cancer using the prescription date and date of birth of the patient (from cancer registration data). For patients without a missing age at prescription, this calculation highlighted inconsistencies for 0.2% of dispensed prescription items in the linked data.

An inconsistency was defined as a difference in age of more than 1 year. Examples of this included a recorded age on the prescription record of 2 years, compared with a calculated age of 102 years.

NHSBSA processing of Electronic Transfer of Prescriptions (ETP) messages has been interrupted on a small number of days and there are no dispensed prescriptions recorded for the first day in either April or May due to system downtime. A spike in the number of dispensed prescriptions exists at the end of each of the affected months, which could indicate a recording bias due to bulk processing of missed records. Another data quality issue is that the net ingredient cost for a particular drug can change each month, and assessments of cost must adjust for this; this field is present in both FP10 and EPS items.

Comparing the linked data to cancer registration data, there were systematic differences between the patients with cancer who were and were not dispensed a prescription. To demonstrate this, a cohort of patients diagnosed with malignant cancer (ICD-10: C00–C97, excluding C44), during April to July 2015 was identified. Among these 100 424 patients with cancer, 90 840 patients received a prescription in the same time period (90%). The distribution of demographic and tumour factors, for example, age and cancer type, varied between patients with cancer with and without a dispensed prescription record in the same period (figure 2). The confidence intervals did not overlap for all but two factors; however, the absolute differences

Table 2 Number of dispensed prescription items by the prescribing institution for a cohort of patients with cancer diagnosed during April to July 2015, with a prescription dispensed in the same time period

Prescribing setting from NHS Digital's GP practice list ⁹	Dispensed prescription items	Per cent of all dispensed prescription items (%)	Per cent of prescription items dispensed by a listed practice ⁹ (%)
GP practice	1 820 523	98.82	99.49
Out of hours practice	3806	0.21	0.21
Community health service	2810	0.15	0.15
Other	1106	0.06	0.06
Hospital service	456	0.02	0.02
Walk in centre practice	389	0.02	0.02
Urgent and emergency care	215	0.01	0.01
Hospice	168	0.01	0.01
Public health service	119	0.01	0.01
Walk in centre+out of hours practice	98	0.01	0.01
Care home/nursing home	95	0.01	0.01
Optometry service	8	0.00	0.00
Prison	6	0.00	0.00
Organisation not listed as a GP practice by NHS Digital, ⁹ assumed secondary care	12 396	0.67	–
Total	1 842 195	–	–

GP, general practitioner; NHS, National Health Service.

were small. The most marked difference was for age at cancer diagnosis, where, for example, 49% of patients with cancer with a dispensed prescription were aged 70–89, but among patients with cancer without a dispensed prescription only 18% were aged 70–89. Variation by age is expected clinically, as individuals often need treatment for an increasing number of conditions as they age (therefore the background rate changes), and patients with cancer who do not have any prescriptions dispensed in the community may be particularly

unwell and admitted to hospital. This highlights the need to control for age during analyses, particularly when matching to the general population. For sex, cancer site, stage at diagnosis and ethnicity, the absolute differences were small (≤ 1.0).

Patient and public involvement

Patients and the public were not involved in the development of this study. Data for this study are based on information collected by the NHS.

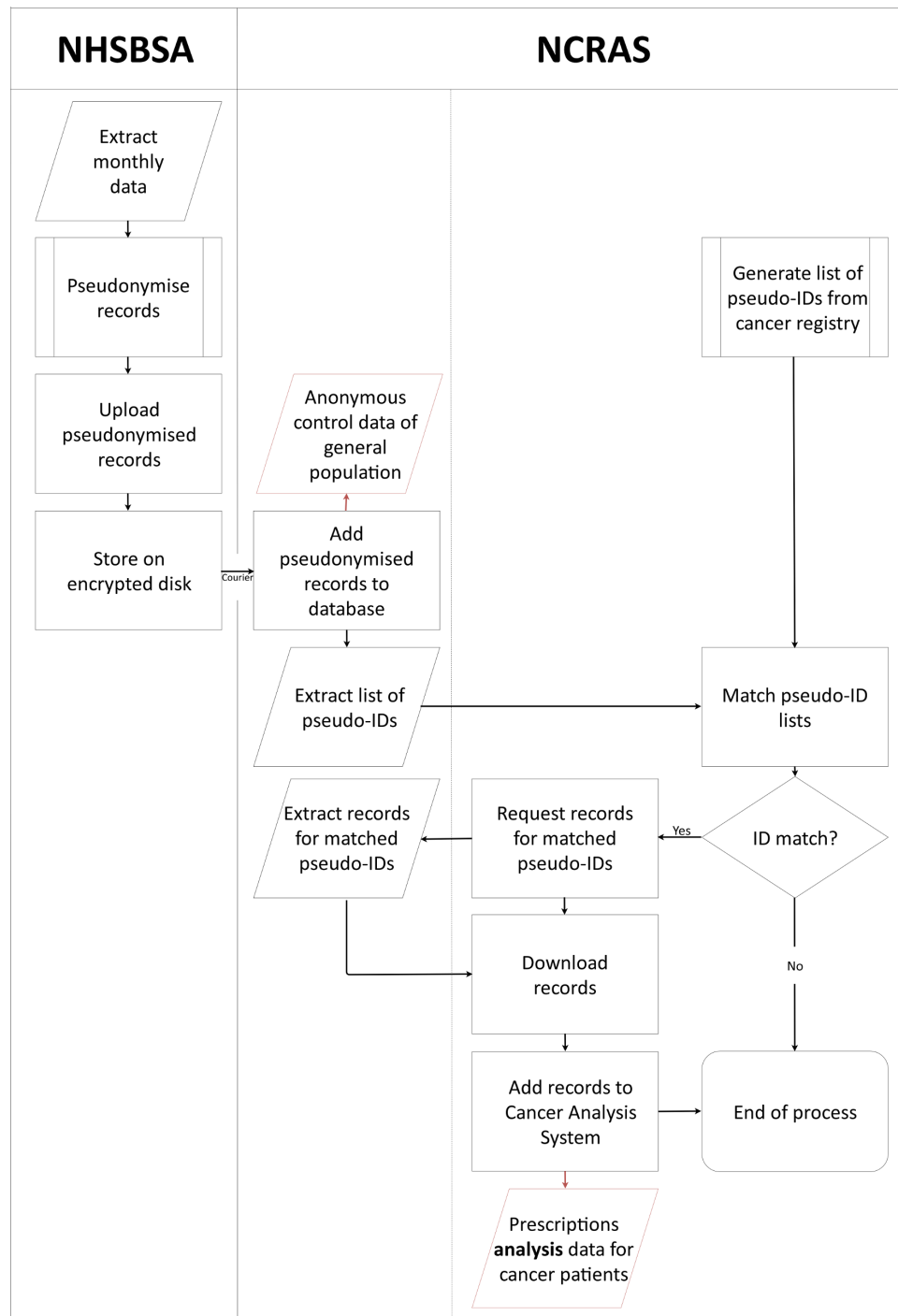


Figure 1 Pseudonymisation and linkage process for the dispensed prescriptions data and cancer registry data. NCRAS, National Cancer Registration and Analysis Service; NHSBSA, NHS Business Services Authority.

FINDINGS TO DATE

Timely, record-level community-dispensed prescriptions data provide a rich data source for public health; especially when the coverage is national and near complete. The linkage of dispensed prescriptions data to national cancer registration data creates a powerful resource that facilitates a better understanding of healthcare use over the duration of illness. This may, to our knowledge, become the largest resource of its kind globally, due to the underlying population size of England and longitudinal coverage of the national cancer registry. Projects of interest to NCRAS include a study of end of life prescribing in the community among patients with cancer, an investigation of repeat prescriptions to derive measures of prior morbidity status in patients with cancer and studies of prescription activity surrounding the date of cancer diagnosis.¹³ This work has not yet (as of March 2018) been published in peer-reviewed journals.

A very similar resource is available in Scotland (the Prescribing Information System), and the existence of their unique numeric patient identifier allows linkage to other Scottish healthcare data, including the Scottish National Cancer Registry. The population size of Scotland is smaller than England, but the benefit of such a linked resource has been demonstrated.¹⁴ The national structure of healthcare in Denmark has also afforded a similar resource to be created, and epidemiological studies investigating the impact on early diagnosis have been recently

emerging.^{15–17} This English linked resource could be used for cancer epidemiological studies of diagnostic pathways, health outcomes and inequalities; to establish primary care comorbidity indices and for guideline concordance studies of treatment. Hormonal therapy is one example of this, as it is a major treatment modality for breast and prostate cancer which has been largely delivered in the community setting for a number of years.

STRENGTHS AND LIMITATIONS

The linkage of community-dispensed prescriptions data to cancer registration data can greatly enhance our understanding of the patient pathway. It can provide novel insight into symptom profiles and improve our understanding of long-term trends in the patterns of drugs dispensed before and after a diagnosis of cancer (the latter of which may have utility as a proxy for quality of life). It could also potentially identify patterns which could indicate opportunities for the earlier diagnosis of cancer. The key strength of this data source is its population coverage. Linkage to the cancer registration data also provides information on the date and cause of death of the patient, allowing effective censoring. This linked dispensed prescriptions and cancer registration data resource can also be linked to other datasets held by NCRAS, including Hospital Episodes Statistics,¹⁸ the Radiotherapy Dataset¹⁹ and the Systemic Anti-Cancer

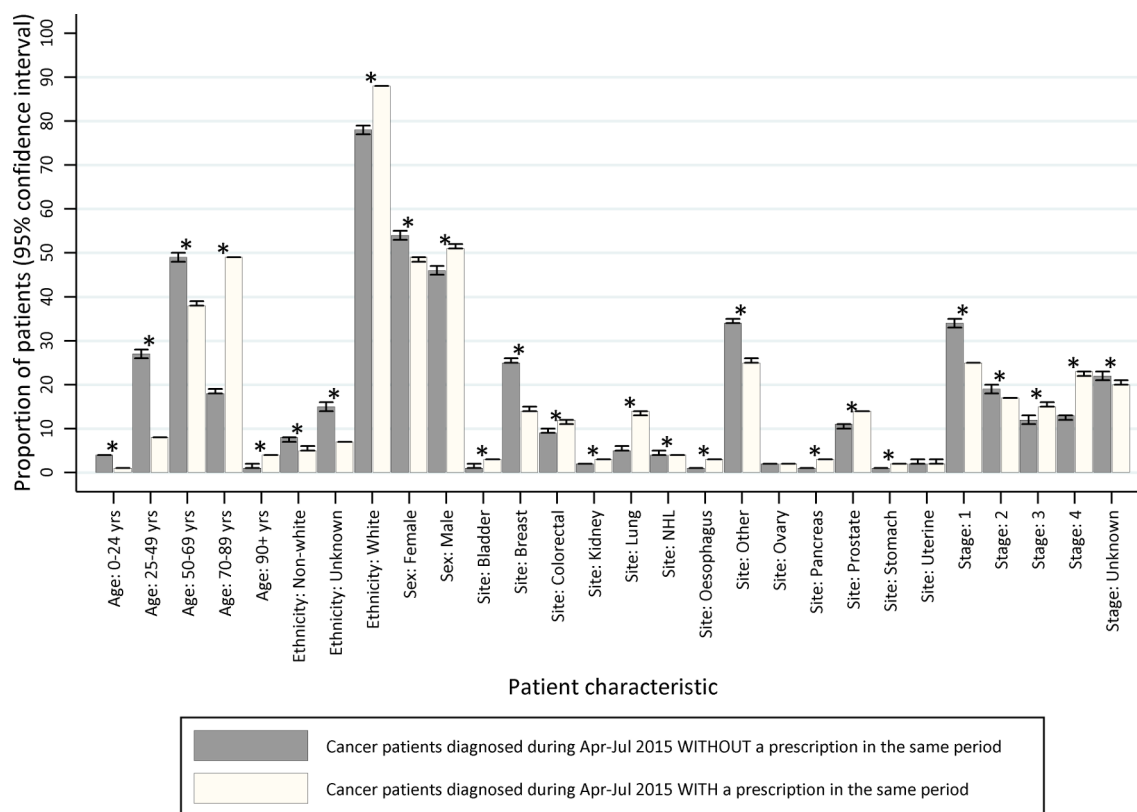


Figure 2 Representativeness of linked cancer registry and dispensed prescription data as compared with cancer registry data alone, by key patient and tumour characteristics (age at diagnosis, ethnicity, sex, tumour site and stage at cancer diagnosis). *Indicates non-overlapping confidence intervals.

Therapy dataset.²⁰ The pseudonymisation process strengthens its value, as it has permitted the existence of an anonymised national control population for comparison purposes, to understand how patients with cancer differ from the general population.

There are a number of limitations of this data resource, as with all routinely collected healthcare data.^{21 22} First, prescriptions dispensed in a private setting, prison setting or without an NHS number recorded are not captured by the data. However, the impact of this is estimated to be less than 3%,²³ less than 1%^{24 25} and 10% of all prescriptions dispensed. In addition, prescriptions that were written but not dispensed, or not submitted by the pharmacy to NHS Prescription Services are not captured, though this is thought to be minimal as the dispenser would not be reimbursed. Sex is not captured within the dispensed prescriptions dataset, neither is an alternative piece of information which could be used to derive sex (eg, title). This is only problematic for the control population, and limits the type of comparison that can be made.

The drug indication is not included in the prescription, therefore cannot be captured. For drugs with multiple indications, this may be problematic as it may be unclear which underlying condition required the prescription. To overcome this, the patients' prescribing history could be used to approximate the indication. Future work could link drugs prescribed to likely indications as recommended by the BNF. Ideally, the patient's medical history, as recorded in primary care, would be available. However, a national data source of this has not been identified. Many primary care databases are available, for example, the Clinical Practice Research Datalink and The Health Improvement Network database, but these only cover 6.9%²⁶ and 6.2%²⁷ of the UK population, respectively.

The full date of the dispensed prescription and the prescribing organisations' postal code is only available for dispensed prescriptions from the EPS. In the linked data (patients with cancer diagnosed after 1994), the proportion of dispensed prescription items from FP10 forms, therefore without the day of prescription or prescribers' postal code were 76% in April, 74% in May, 71% in June and 70% in July 2015, which is continually improving. The EPS is currently on phase 4 of deployment and release 2 of the system, with each phase improving the functionality for both GP practices and pharmacies, thereby improving ease of use. The latest statistics (as of 6 November 2017) showed that 91.6% of GP practices and 99.2% of pharmacies were using the EPS.⁶

Two further limitations exist, which are primarily relevant for pharmacovigilance studies, particularly long-term studies investigating whether a certain drug leads to cancer. Namely, the information captured is for drugs dispensed, irrespective of whether the patient actually took the drug. However, it can still be used as a proxy for an underlying disease or the presence of disease risk factor. Finally, for studies where detailed information on the strength of the drug is required, work must be undertaken to parse text fields and accurately calculate the

amount of active ingredient, using combinations of the data items available.

COLLABORATION

The Data Sharing Agreement outlined the terms of the partnership, data transfer and subsequent use. These state that NHSBSA will remain the primary data owner for the prescriptions dataset and will continue to be responsible for the continuity, quality, timeliness and availability of the data it contains. PHE will assume the responsibility of in-common data controller as appropriate for all data transferred to it from the NHSBSA under the terms of the agreement. PHE will be responsible for determining and approving the specific purposes that fit within the objectives of the agreed data exchange, for which the linked dispensed prescriptions and cancer registration dataset are used to support its statutory functions and core remit. PHE will be responsible for reviewing and approving the release of extracts of the linked data.

PHE has been granted specific legal permission to collect information about patients with cancer for specific purposes, without the need to seek consent. These purposes include health improvement and service provision. This permission was granted to PHE through Section 251 of the NHS Act 2006. This support is reviewed annually by the Confidentiality Advisory Group of the Health Research Authority.²⁸ PHE manages the release of potentially identifiable data through the Office for Data Release (ODR). The ODR provides a common governance framework for responding to requests to access PHE data, and is subject to strict confidentiality provisions in line with the requirements of the Common Law Duty of Confidentiality,²⁹ the Data Protection Act 1998³⁰ (to be superseded by the General Data Protection Regulation (EU) 2016/679 which will take effect on 25 May 2018)³¹ and the 7 Caldicott principles.³² Applications to access this linked prescriptions data for patients with cancer should be directed through the ODR,³³ and application forms are available on their website.³³

The ODR accepts applications from UK, EEA and international organisations; however, approvals to process any data controlled by PHE will be subject to adequate safeguards being established with the data recipient to ensure that: the level of protection afforded to individuals by UK data protection laws is not undermined; the purpose of any request complements the permissions to process the data without consent granted to PHE by the Secretary of State under the Health Service (Control of Patient Information) Regulations 2002³⁴ and that appropriate ethical assurances are met.

CONCLUSIONS

The need to understand the healthcare of patients with cancer in the community is paramount, both following their diagnosis and to improve earlier diagnosis. Repurposing administrative datasets is an efficient method of doing this, providing the data quality and content is

sufficient. National sources of reimbursed prescriptions, such as the resource described in this profile, have been valuable to epidemiological research in Scandinavian countries, particularly Denmark.^{15–17} We have demonstrated the value of dispensed prescriptions data for research and health monitoring in England, and the added value when linked to cancer registration information. The long-term coverage of the national cancer registration data has created a large linked resource even with 4 months of dispensed prescriptions data, and with more months of data the scale of the resource will expand creating a resource of unparalleled scale. The implications of data quality must be considered when designing studies, but the value to cancer research is clear.

Acknowledgements The authors acknowledge and thank the entire team at NHS Business Services Authority for supporting and facilitating this partnership. Data for this study are based on patient-level information collected by the NHS, as part of the care and support of cancer patients. The data are collated, maintained and quality assured by the National Cancer Registration and Analysis Service, which is part of Public Health England (PHE).

Contributors Conception and design of the work: KEH, RB, JR, LE-B, GL, MGM and TR. Acquisition and pseudonymisation of the data: KEH, BS, PG, AH, KH, GM, NM and RR. Analysis of the data: KEH, VHC and KW. All authors made substantial contributions to the interpretation of the findings. All authors contributed to drafting the manuscript or revising it critically for important intellectual content and approved the final version submitted. All authors have agreed to be accountable for all aspects of the work.

Funding GL is supported by a Cancer Research UK Advanced Clinician Scientist Fellowship (C18081/A18180).

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Collaborations can be proposed to the National Cancer Registration and Analysis Service via NCRASenquiries@phe.gov.uk. Enquiries for data access can be made to Public Health England's Office for Data Release (odr@phe.gov.uk).

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- Herbert A, Wijlaars L, Zylbersztejn A, *et al*. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol* 2017;46:1093–1093i.
- NHS Business Services, NHS Prescription Services. Drug tariff. <https://www.nhsbsa.nhs.uk/pharmacies-gp-practices-and-appliance-contractors/drug-tariff> (accessed 21 Jul 2017).
- NHS Digital. Prescriptions dispensed in the community, England 2004–14. <http://content.digital.nhs.uk/catalogue/PUB17644/presdisp-com-eng-2004-14-rep.pdf> (accessed 21 Jul 2017).
- National Institute for Health and Care Excellence (NICE). Multimorbidity and polypharmacy. Key therapeutic topic. <https://www.nice.org.uk/guidance/kt18/resources/multimorbidity-and-polypharmacy-pdf-58757959453381> (accessed 21 Jul 2017).
- National Cancer Registration and Analysis Service. Patient information. <http://www.ncras.nhs.uk/patientinfo/> (accessed 21 Jul 2017).
- NHS Digital. Statistics and progress. <https://digital.nhs.uk/Electronic-Prescription-Service/statistics-and-progress> (accessed 08 Nov 2017).
- Joint Formulary Committee. *British National Formulary*. 68. London: BMJ Group and Pharmaceutical Press, 2014.
- NHS Prescription Services. BNF Classification and Pseudo Classification Used by the NHS Prescription Services. https://www.nhsbsa.nhs.uk/sites/default/files/2017-04/BNF_Classification_Booklet-2017.pdf (accessed 21 Jul 2017).
- NHS Digital. Data for GP practices. <https://digital.nhs.uk/organisation-data-service/data-downloads/gp-data> (accessed 21 Jul 2017).
- University of Nottingham. OpenPseudonymiser. <https://www.openpseudonymiser.org/Default.aspx> (accessed 21 Jul 2017).
- National Institute of Standards and Technology (NIST). FIPS Publication 180-4: Secure Hash Standard (SHS). 2015 <http://dx.doi.org/10.6028/NIST.FIPS.180-4>.
- NHS Business Services Authority. NHSBSA prescription processing accuracy processed prescriptions. 2016 <https://cms.nhsbsa.nhs.uk/sites/default/files/2017-05/Accuracy%20Measures%20for%20Publication%20-%20new%20website.pdf> (accessed 21 Jul 2017).
- Henson K, Shand B, Hunter K, *et al*. Exploring a new dataset: Primary care prescription activity among cancer patients in England. *Poster Presented at: 2017 Annual Conference of the North American Association of Central Cancer Registries (NAACCR) 16-23rd June 2017*, Albuquerque, New Mexico.
- Alvarez-Madrado S, McTaggart S, Nangle C, *et al*. Data Resource Profile: The Scottish National Prescribing Information System (PIS). *Int J Epidemiol* 2016;45:714–5.
- Guldbrandt LM, Møller H, Jakobsen E, *et al*. General practice consultations, diagnostic investigations, and prescriptions in the year preceding a lung cancer diagnosis. *Cancer Med* 2017;6:79–88.
- Pottegård A, Hallas J. New use of prescription drugs prior to a cancer diagnosis. *Pharmacoepidemiol Drug Saf* 2017;26:223–7.
- Hansen PL, Hjertholm P, Vedsted P. Increased diagnostic activity in general practice during the year preceding colorectal cancer diagnosis. *Int J Cancer* 2015;137:615–24.
- NHS Digital. Hospital Episode Statistics. <http://content.digital.nhs.uk/hes> (accessed 22 Feb 2018).
- NHS Digital. Data set: Radio Therapy Dataset (RTDS). http://www.datadictionary.nhs.uk/data_dictionary/messages/clinical_data_sets/data_sets/radiotherapy_data_set_fr.asp?shownav=1 (accessed 22 Feb 2018).
- NHS Digital. Data set: Systemic Anti-Cancer Therapy (SACT) dataset. http://www.datadictionary.nhs.uk/data_dictionary/messages/clinical_data_sets/data_sets/systemic_anti-cancer_therapy_data_set_fr.asp?shownav=1 (accessed 22 Feb 2018).
- Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, *et al*. Data resource profile: the Danish National Prescription Registry. *Int J Epidemiol* 2017;46:dyw213–798.
- Johannesdottir SA, Horváth-Puhó E, Ehrenstein V, *et al*. Existing data sources for clinical epidemiology: The Danish National Database of Reimbursed Prescriptions. *Clin Epidemiol* 2012;4:303–13.
- The King's Fund. The UK private health market. 2014 <https://www.kingsfund.org.uk/sites/default/files/media/commission-appendix-uk-private-health-market.pdf> (accessed 22 Feb 2018).
- Ministry of Justice. Official statistics, prison population figures. 2018 <https://www.gov.uk/government/statistics/prison-population-figures-2018> (accessed 22 Feb 2018).
- NHS Commissioning, Direct Commissioning Change Projects Team. Strategic direction for health services in the justice system 2016–2020. 2016 <https://www.england.nhs.uk/wp-content/uploads/2016/10/hlth-justice-directions-v11.pdf>.
- Herrett E, Gallagher AM, Bhaskaran K, *et al*. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015;44:827–36.
- The Health Improvement Network (THIN) Research Team. THIN database. <http://www.ucl.ac.uk/pcph/research/thin-database/database> (accessed 21 Jul 2018).
- NHS Health Research Authority. Section 251 and the Confidentiality Advisory Group (CAG). <http://www.hra.nhs.uk/about-the-hra/our-committees/section-251/> (accessed 21 Jul 2017).
- The rules are extrapolated from the decisions of the courts*.
- Great Britain. *Data Protection Act*. London: Stationery Office, 1998.
- European Union. General Data Protection Regulation. 2018 <https://www.eugdpr.org/> (accessed 22 Feb 2018).
- Department of Health. Information: to share or not to share? The Information Governance Review. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/192572/2900774_InfoGovernance_accv2.pdf (accessed 21 Jul 2017).



33. Public Health England. The Office for Data Release. <https://www.gov.uk/government/publications/accessing-public-health-england-data/about-the-phe-odr-and-accessing-data> (accessed 21 Jul 2017).
34. Great Britain. The Health Service (control of patient information) Regulations. 2002 <http://www.legislation.gov.uk/uksi/2002/1438/made?view=plain> (accessed 22 Feb 2018).