BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (http://bmjopen.bmj.com).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020980
Article Type:	Cohort profile
Date Submitted by the Author:	11-Dec-2017
Complete List of Authors:	Henson, Katherine; Public Health England, National Cancer Registration and Analysis Service Brock, Rachael; Public Health England, National Cancer Registration and Analysis Service Shand, Brian; Public Health England, National Cancer Registration and Analysis Service Coupland, Victoria; Public Health England, National Cancer Registration and Analysis Service Elliss-Brookes, Lucy; Public Health England, National Cancer Registration and Analysis Service Lyratzopoulos, Georgios ; University College London, Department of Epidemiology & Public Health, Health Behaviour Research Centre; Public Health England, National Cancer Registration and Analysis Service Godfrey, Philip; NHS Business Services Authority Haigh, Abigail; NHS Business Services Authority Hunter, Kelvin; Addenbrookes Hospital, Department of Medicine (Cambridge University) McCabe, Martin; University of Manchester, Division of Cancer Sciences; Public Health England, National Cancer Registration and Analysis Service Mitchell, Graham; NHS Business Services Authority Monckton, Nina; NHS Business Services Authority Robson, Robert; NHS Business Services Authority Round, Thomas; King's College London, Department of Primary Care and Public Health Sciences Wong, Kwok; Public Health England, National Cancer Registration and Analysis Service Rashbass, Jem; Public Health England, National Cancer Registration and Analysis Service
Keywords:	ONCOLOGY, EPIDEMIOLOGY, PRIMARY CARE, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS

SCHOLARONE™
Manuscripts

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 Coupland[1], Lucy Elliss-Brookes[1], Georgios Lyratzopoulos[1, 2], Phil 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 Rashbass[1]

[1] National Cancer Registration and Analysis Service, Public Health England, Skipton House, London, SE1 6LH, UK

[2] ECHO (Epidemiology of Cancer Healthcare and Outcomes) Group, Department of Behavioural Science and Health, University College London, Gower Street, Bloomsbury, London WC1E 6BT, UK

[3] NHS Business Services Authority, Stella House, Goldcrest Way, Newburn Riverside, Newcastle upon Tyne, NE15 8NY, UK

[4] Department of Medicine (Cambridge University), Addenbrookes Hospital, Box 157, Hills Road, Cambridge, CB2 0QQ

[5] Institute of Cancer Sciences, University of Manchester, Manchester Academic Health Science Centre, Core Technology Facility, The University of Manchester, 46 Grafton Street, Manchester, M13 9NT, UK

[6] Department of Primary Care and Public Health Sciences, King's College London, Addison House, Guys Campus, London SE1 1UL

Corresponding author: Dr Katherine Henson

National Cancer Registration and Analysis Service, Public Health England, Skipton House, London, SE1 6LH, UK

T: 020 765 48304; E: katherine.henson@phe.gov.uk

Disclaimers: All authors declare that they have no conflicts of interest in relation to this work.

Manuscript length: 2,833 words

For peer review only

ABSTRACT

Purpose

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

Participants

The linked prescriptions cancer registry data resource is available for April to July 2015 currently for all patients diagnosed with cancer in England with a prescription in that time frame.

The prescription data is collected by NHS Prescription Services, and the cancer registry data is processed by Public Health England. All data is routine healthcare data, used for secondary purposes.

Detailed demographic and clinical information on the type of cancer diagnosed and treatment is collected by the cancer registry. The prescription data contains basic demographic information, geography measures of the prescription, drug information (quantity, strength and presentation), cost of the drug, and the date that the dispensed prescription was submitted to NHSBSA.

Findings to date

Findings include a study of end of life prescribing in the community among cancer patients, an investigation of repeat prescriptions to derive measures of prior morbidity status in cancer patients, 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.

For peer review only

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 cancer patients 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 drug is not included in the data which may be problematic for drugs with multiple indications, for example anti-depressants can be indicated for depression or chronic pain.
- For pharmacovigilance studies, be cautious as dispensed drug information does not guarantee that the patient has taken the medication.

For peer review only

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 (NHS) in England, Public Health England holds cancer and other disease data; NHS Digital collects hospital activity [1] 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 1 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 multiply morbid 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 Public Health England (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 skilled registration officers into a clinically comprehensive record for each tumour. The majority of cancer care is delivered in secondary care; therefore the data collected reflects this.

A Data Sharing Agreement was established between PHE and NHSBSA for NHSBSA to supply pseudonymised prescriptions data to PHE, providing hitherto missing information on this aspect of community healthcare for cancer patients. We aim to describe the prescription data, the linkage,

For peer review only

and key data quality implications, with a focus on the linked data available within Public Health England.

COHORT DESCRIPTION

Data collected

The prescriptions data is 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 at 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 formats 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 is 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 other healthcare institutions, for example urgent care centres.

Demographic, prescriber and drug details are collected, including quantity, strength (dose) and presentation (tablet, injection etc.) The full drug item is recorded as per the British National Formulary (BNF) version 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.

For peer review only

Table 1: Description of data items included, with data quality completeness of linked prescriptions cancer registry data, for patients diagnosed with malignant cancer (excluding non-melanoma skin cancer) after 1994 with a 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 number and date of birth	
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 Postcode	The postcode of the prescribing practice, as held by the NHSBSA master database and the HSCIC Organisation Data Service (ODS). It is in text and numerical format as per standard postcodes.	27%
Primary care Organisation Code	The code allocated by ODS to the Prescriber's Clinical Commissioning Group (CCG) or other primary care organisation (PCO). It is a 5 digit code containing numbers and letters.	99.7%
Primary care organisation 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. It is a 6 digit code containing numbers and letters.	>99.9%
Practice name		>99.9%
Net Ingredient Cost (NIC)	This is the cost (in pence) of the supplied 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%
Number of Items	This shows the number 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 i.e. tablet, ampoule, vial, 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 number of special containers dispensed.	100%
Drug Presentation paid	A description of the product reimbursed as held by the Dictionary of Medicines and Devices (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 prescribeable generic, the description of the generic product without supplier or trade name information is given. NB: On many occasions this will be the same as that prescribed.	100%

For peer review only

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 upon the British National Formulary, 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%

For peer review only

Prescription data is 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. During April to July, 332,655,118 prescription items among 29,481,344 individuals are identified. 10% of the prescription 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 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 – July 2015, and had a prescription dispensed during the same period. Amongst this cohort, 99% of their prescriptions were issued by a GP practice [9] (see Table 2).

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

Prescribing setting from NHS Digital's GP Practice list [9]	Prescription items	Percent of all dispensed prescription items	Percent of prescription items dispensed by a listed practice [9]
GP practice	1,820,523	98.82%	99.49%
Out of Hours Practice	3,806	0.21%	0.21%
Community health service	2,810	0.15%	0.15%
Other	1,106	0.06%	0.06%
Hospital service	456	0.02%	0.02%
Walk In Centre Practice	389	0.02%	0.02%
Urgent & 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	-	-

For peer review only

Pseudonymisation process

The partnership between NCRAS and NHSBSA has allowed NCRAS to obtain prescribing in the community information for both cancer patients, and non-cancer patients. The information on non-cancer patients has enabled the information on (anonymous) individuals without cancer to 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 prescriptions data, and the pseudonymised data was sent to NCRAS on an encrypted disk. The same pseudonymisation procedure was also run by NCRAS on the identifiers of cancer patients stored in the cancer registry. A script was then run which linked the pseudonymised prescriptions data with patient records in the cancer registry using the pseudonym identifiers. A subset of the pseudonymised row-level 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 approach [10]). 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 by both NHSBSA and NCRAS is identical. This allows us to match the pseudonyms in the 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 re-identify NHS numbers.

Data quality

NHS Prescription Services reprocess a random sample of 50,000 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-July 2015 data. In these items, the prescription 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 peer review only

for 0.3% of prescription items. Amongst the FP10 prescription items, age at prescription is missing for 19% (in the unlinked data). However, the age at prescription can be approximated for cancer patients 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 prescription items in the linked data. An inconsistency was defined as a difference in age of more than one year. Examples of this included a recorded age on the prescription record of 2 years, compared to 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 prescriptions recorded for the first day in either April or May due to system downtime. A spike in the number of 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 cancer patients who did and did not receive 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. Amongst these 100,424 cancer patients, 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 cancer patients with and without a prescription record in the same period (Figure 2). The confidence intervals did not overlap for all but two factors; however the absolute differences were small. The most marked difference was for age at cancer diagnosis, where, for example, 49% of cancer patients with a prescription were aged 70-89, but among cancer patients without a 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 cancer patients who don’t have any

For peer review only

prescriptions 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).

FINDINGS TO DATE

Timely, record-level community-based prescription data provides a rich data source for public health; especially when the coverage is national and near-complete. The linkage of prescription 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 cancer patients, an investigation of repeat prescriptions to derive measures of prior morbidity status in cancer patients, and studies of prescription activity surrounding the date of cancer diagnosis [13].

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 health care 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.

For peer review only

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 prescribed pre- and post-cancer diagnosis (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. The pseudonymisation process strengthens its value, as it has permitted the existence of an anonymised national control population for comparison purposes, to understand how cancer patients differ from the general population.

There are a number of limitations of this data resource, as with all routinely collected healthcare data [18-19]. Firstly, sex is not captured within the prescriptions dataset, neither is an alternative piece of information which could be used to derive sex (e.g. 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 (CPRD) and The Health Improvement Network (THIN) database, but these only cover 6.9% [20] and 6.2% [21] of the UK population, respectively.

For peer review only

The full date of prescription and the prescribing organisations' postcode is only available for prescriptions from the EPS. In the linked data (cancer patients diagnosed after 1994), the proportion of prescription items without the day of prescription or prescribers' postcode 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 6th 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 prescription 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 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 prescription data linked to cancer registration data.

For peer review only

PHE has been granted specific legal permission to collect information about cancer patients 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 [22]. 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 [23], the Data Protection Act 1998 [24] and the 7 Caldicott Principles [25]. Applications to access this linked prescriptions data for cancer patients should be directed through the ODR [26], and application forms are available on their website [26].

For peer review only

CONCLUSIONS

The need to understand the healthcare of cancer patients in the community is paramount, both following their diagnosis and to improve earlier diagnosis. Re-purposing 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 prescription 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 prescriptions data, and with more prescriptions 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.

CONTRIBUTORSHIP STATEMENT

Conception and design of the work: KH, RB, JR, LEB, GL, MM and TR. Acquisition and pseudonymisation of the data: KH, BS, PG, AH, KH, GM, NM and RR. Analysis of the data: KH, VC 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.

For peer review only

COMPETING INTERESTS

All authors declare that they have no conflicts of interest in relation to this work.

FUNDING

There is no funding to declare in relation to this work.

ACKNOWLEDGEMENTS

The authors acknowledge and thank the entire team and NHS Business Services Authority for supporting and facilitating this partnership. Data for this study is based on patient-level information collected by the NHS, as part of the care and support of cancer patients. The data is collated, maintained and quality assured by the National Cancer Registration and Analysis Service, which is part of Public Health England (PHE). GL is supported by a Cancer Research UK Advanced Clinician Scientist Fellowship (C18081/A18180).

REFERENCES

[1] Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol* 2017; 46(4):1093-1093i. doi: 10.1093/ije/dyx015.

[2] NHS Prescription Services. Drug Tariff. Available from: https://www.nhsbsa.nhs.uk/pharmacies-gp-practices-and-appliance-contractors/drug-tariff [Accessed 21/07/2017]

[3] NHS Digital. Prescriptions Dispensed in the Community, England 2004-14. Available from: http://content.digital.nhs.uk/catalogue/PUB17644/pres-disp-com-eng-2004-14-rep.pdf [Accessed 21/07/2017]

[4] National Institute for Health and Care Excellence (NICE). Multimorbidity and polypharmacy, Key therapeutic topic. Available from: https://www.nice.org.uk/guidance/ktt18/resources/multimorbidity-and-polypharmacy-pdf-58757959453381 [Accessed 21/07/2017]

[5] National Cancer Registration and Analysis Service. Patient Information. Available from: http://www.ncras.nhs.uk/patientinfo/ [Accessed 21/07/2017]

[6] NHS Digital. Statistics and progress. Available from: https://digital.nhs.uk/Electronic-Prescription-Service/statistics-and-progress [Accessed 08/11/2017]

[7] Joint Formulary Committee. British National Formulary. 68. London: BMJ Group and Pharmaceutical Press; 2014.

[8] NHS Prescription Services. BNF Classification and Pseudo Classification Used by the NHS Prescription Services. Available from: https://www.nhsbsa.nhs.uk/sites/default/files/2017-04/BNF_Classification_Booklet-2017.pdf [Accessed 21/07/2017]

[9] NHS Digital. Data for GP Practices. Available from: https://digital.nhs.uk/organisation-data-service/data-downloads/gp-data [Accessed 21/07/2017]

[10] University of Nottingham. OpenPseudonymiser. Available from: https://www.openpseudonymiser.org/Default.aspx [Accessed 21/07/2017]

For peer review only

[11] National Institute of Standards and Technology (NIST). FIPS Publication 180-4: Secure Hash Standard (SHS); 2015. Available from: http://dx.doi.org/10.6028/NIST.FIPS.180-4

[12] NHS Business Services Authority. NHSBSA Prescription Processing Accuracy, December 2016 Processed Prescriptions. Available from: https://cms.nhsbsa.nhs.uk/sites/default/files/2017-05/Accuracy%20Measures%20for%20Publication%20-%20new%20website.pdf [Accessed 21/07/2017]

[13] Henson K, Shand B, Hunter K, Coupland V, Godfrey P, Elliss-Brookes E, Brock R. 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.

[14] Alvarez-Madrazo S, McTaggart S, Nangle C, Nicholson E, Bennie M. Data Resource Profile: The Scottish National Prescribing Information System (PIS). *Int J Epidemiol* 2016; 45(3):714-715f. doi: 10.1093/ije/dyw060.

[15] Guldbrandt LM, Møller H, Jakobsen E, Vedsted P. General practice consultations, diagnostic investigations, and prescriptions in the year preceding a lung cancer diagnosis. *Cancer Medicine* 2017; 6(1):79-88. doi:10.1002/cam4.965.

[16] Pottegård A, Hallas J. New use of prescription drugs prior to a cancer diagnosis. *Pharmacoepidemiol Drug Saf* 2017; 26(2):223-227. doi:10.1002/pds.4145.

[17] Hansen PL, Hjertholm P, Vedsted P. Increased diagnostic activity in general practice during the year preceding colorectal cancer diagnosis. *Int J Cancer* 2015; 137: 615–624. doi:10.1002/ijc.29418

[18] Pottegård A, Schmidt SA, Wallach-Kildemoes H, Sørensen HT, Hallas J, Schmidt M. Data Resource Profile: The Danish National Prescription Registry. *Int J Epidemiol* 2016; 46(3): 798-798f. doi: 10.1093/ije/dyw213

For peer review only

[19] Johannesdottir SA, Horváth-Puhó E, Ehrenstein V, Schmidt M, Pedersen L, Sørensen HT. Existing data sources for clinical epidemiology: The Danish National Database of Reimbursed Prescriptions. *Clin Epidemiol* 2012; 4:303-313. doi:10.2147/CLEP.S37587.

[20]Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, Smeeth L. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; 44(3):827-36. doi: 10.1093/ije/dyv098.

[21] The Health Improvement Network (THIN) Research Team. THIN Database. Available from: https://www.ucl.ac.uk/pcph/research-groups-themes/thin-pub/database [Accessed 21/07/2017]

[22]NHS Health Research Authority. Section 251 and the Confidentiality Advisory Group (CAG). Available from: http://www.hra.nhs.uk/about-the-hra/our-committees/section-251/ [Accessed 21/07/2017]

[23] The rules are extrapolated from the decisions of the courts

[24] Great Britain. Data Protection Act. London: Stationery Office; 1998.

[25] Department of Health. Information: To share or not to share? The Information Governance Review. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/192572/2900774_InfoGovernance_accv2.pdf [Accessed 21/07/2017]

[26] Public Health England. The Office of Data Release. Available from: https://www.gov.uk/government/publications/accessing-public-health-england-data/about-the-phe-odr-and-accessing-data [Accessed 21/07/2017]

For peer review only

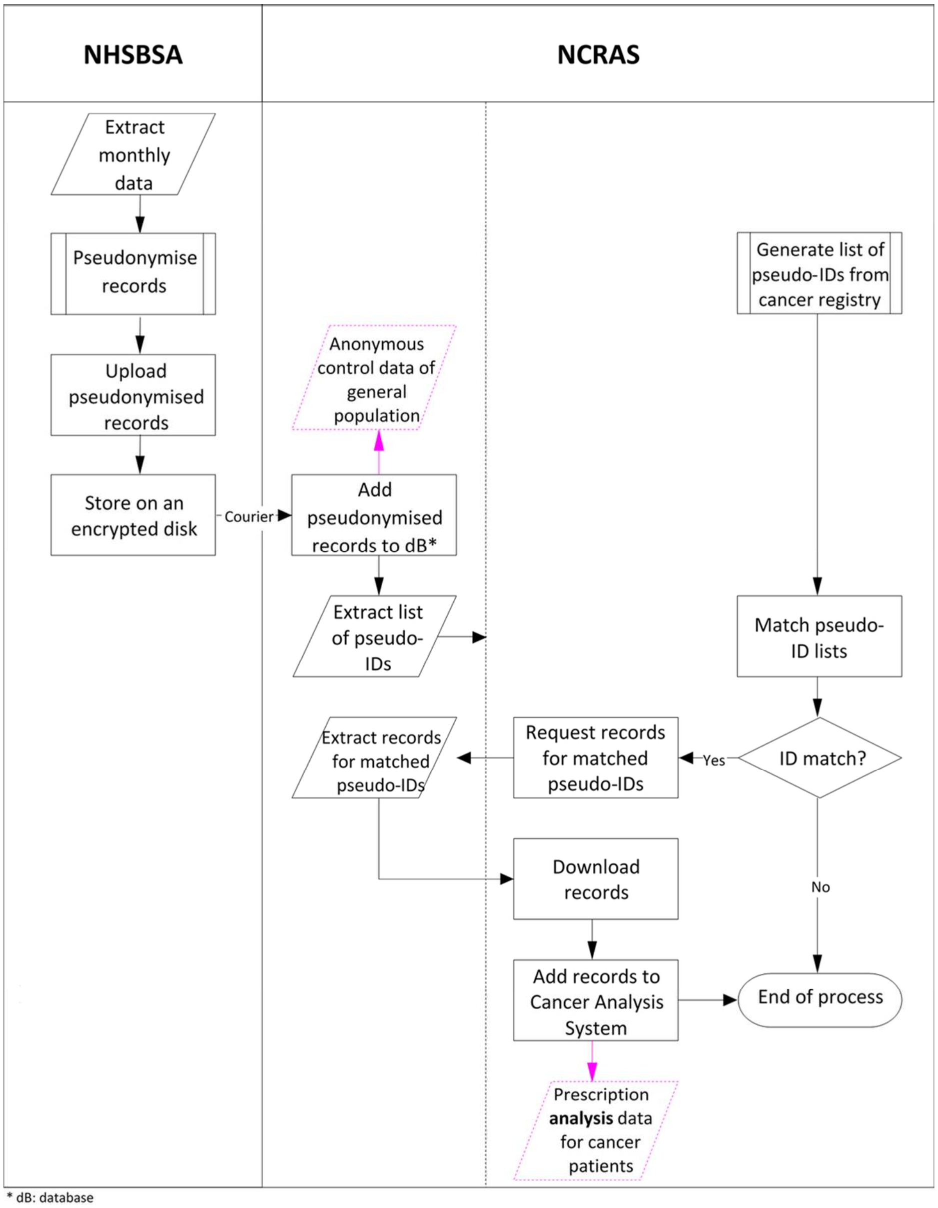

Figure 1: Pseudonymisation and linkage process for the prescription and cancer registry data

99x128mm (300 x 300 DPI)

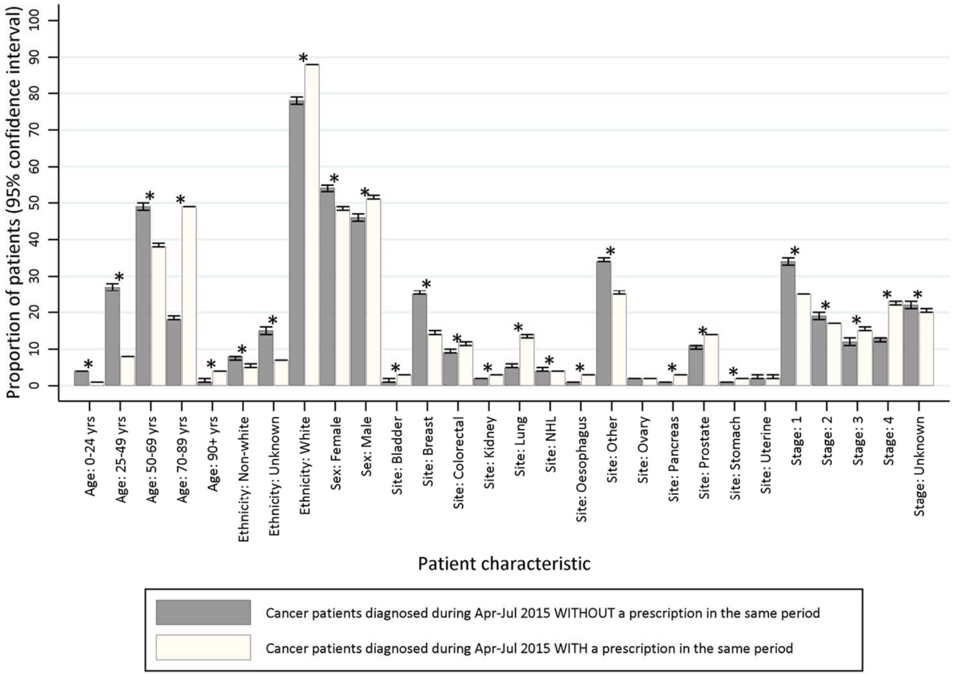

Figure 2: Representativeness of linked cancer registry and prescription data, as compared to a cohort of cancer patients using cancer registry data alone, by key patient and tumour characteristics (age at diagnosis, ethnicity, sex, tumour site and stage at cancer diagnosis)

121x88mm (300 x 300 DPI)

w only

BMJ Open

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

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-020980.R1
Article Type:	Cohort profile
Date Submitted by the Author:	09-Mar-2018
Complete List of Authors:	Henson, Katherine; Public Health England, National Cancer Registration and Analysis Service Brock, Rachael; Public Health England, National Cancer Registration and Analysis Service Shand, Brian; Public Health England, National Cancer Registration and Analysis Service Coupland, Victoria; Public Health England, National Cancer Registration and Analysis Service Elliss-Brookes, Lucy; Public Health England, National Cancer Registration and Analysis Service Lyratzopoulos, Georgios ; University College London, Department of Epidemiology & Public Health, Health Behaviour Research Centre; Public Health England, National Cancer Registration and Analysis Service Godfrey, Philip; NHS Business Services Authority Haigh, Abigail; NHS Business Services Authority Hunter, Kelvin; Addenbrookes Hospital, Department of Medicine (Cambridge University) McCabe, Martin; University of Manchester, Division of Cancer Sciences; Public Health England, National Cancer Registration and Analysis Service Mitchell, Graham; NHS Business Services Authority Monckton, Nina; NHS Business Services Authority Robson, Robert; NHS Business Services Authority Round, Thomas; King's College London, Department of Primary Care and Public Health Sciences Wong, Kwok; Public Health England, National Cancer Registration and Analysis Service Rashbass, Jem; Public Health England, National Cancer Registration and Analysis Service
Primary Subject Heading:	Oncology
Secondary Subject Heading:	Epidemiology, Public health, Health informatics, Pharmacology and therapeutics
Keywords:	ONCOLOGY, EPIDEMIOLOGY, PRIMARY CARE, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS

SCHOLARONE™
Manuscripts

For peer review only

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 Elliss-Brookes[1],
Georgios Lyratzopoulos[1, 2], Phil 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 Rashbass[1]

[1] National Cancer Registration and Analysis Service, Public Health England, Skipton House, London, SE1 6LH, UK

[2] ECHO (Epidemiology of Cancer Healthcare and Outcomes) Group, Department of Behavioural Science and Health, University College London, Gower Street, Bloomsbury, London WC1E 6BT, UK

[3] NHS Business Services Authority, Stella House, Goldcrest Way, Newburn Riverside, Newcastle upon Tyne, NE15 8NY, UK

[4] Department of Medicine (Cambridge University), Addenbrookes Hospital, Box 157, Hills Road, Cambridge, CB2 0QQ

[5] Institute of Cancer Sciences, University of Manchester, Manchester Academic Health Science Centre, Core Technology Facility, The University of Manchester, 46 Grafton Street, Manchester, M13 9NT, UK

[6] Department of Primary Care and Public Health Sciences, King's College London, Addison House, Guys Campus, London SE1 1UL

Corresponding author: Dr Katherine Henson

National Cancer Registration and Analysis Service, Public Health England, Skipton House, London, SE1 6LH, UK

T: 020 765 48304; E: katherine.henson@phe.gov.uk

Disclaimers: All authors declare that they have no conflicts of interest in relation to this work.

Manuscript length: 3,159 words

For peer review only

ABSTRACT

Purpose

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

Participants

The linked prescriptions cancer registry data resource is available for April to July 2015 currently for all patients diagnosed with cancer in England with a dispensed prescription in that time frame.

The dispensed prescriptions data is collected by NHS Prescription Services, and the cancer registry data is processed by Public Health England. All data is 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 contains 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 NHSBSA.

Findings to date

Findings include a study of end of life prescribing in the community among cancer patients, an investigation of repeat prescriptions to derive measures of prior morbidity status in cancer patients, 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

For peer review only

modality for breast and prostate cancer which has been largely delivered in the community setting for a number of years.

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 cancer patients 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 anti-depressants 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.

For peer review only

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 (NHS) in England, Public Health England holds cancer and other disease data; NHS Digital collects hospital activity [1] 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 1 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 multiply morbid 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 Public Health England (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 reflects 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 cancer patients. We aim to describe the dispensed prescriptions

For peer review only

data, the linkage, and key data quality implications, with a focus on the linked data available within Public Health England.

COHORT DESCRIPTION

Data collected

The dispensed prescriptions data is 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 at 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 formats 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 is 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 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) version 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.

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

Table 1: Description of data items included, with data quality completeness of linked 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 [a]
Pseudonymised Patient Identifier	Code derived from Patient NHS number and date of birth	
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 Postcode	The postcode of the prescribing practice, as held by the NHSBSA master database and the HSCIC Organisation Data Service (ODS). It is in text and numerical format as per standard postcodes.	27%
Primary care Organisation Code	The code allocated by ODS to the Prescriber's Clinical Commissioning Group (CCG) or other primary care organisation (PCO). It is a 5 digit code containing numbers and letters.	99.7%
Primary care organisation 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. It is a 6 digit code containing numbers and letters.	>99.9%
Practice name		>99.9%
Net Ingredient Cost (NIC)	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%
Number of Items	This shows the number 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 i.e. tablet, ampoule, vial, 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 number of special containers dispensed.	100%
Drug Presentation paid	A description of the product reimbursed as held by the Dictionary of Medicines and Devices (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 prescribeable generic, the description of the generic product without supplier or trade name information is given. NB: On many occasions this will be the same as that prescribed.	100%

For peer review only

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 upon the British National Formulary, 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%

[a] the number of dispensed prescription items with a known value as a percentage of all dispensed prescription items (both FP10 and EPS)

Dispensed prescriptions data is 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 six months. During April to July, 332,655,118 dispensed prescription items among 29,481,344 individuals are identified. 10% 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 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 – July 2015, and had a prescription dispensed during the same period. Amongst this cohort, 99% of their dispensed prescriptions were issued by a GP practice [9] (see Table 2).

For peer review only

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

Prescribing setting from NHS Digital's GP Practice list [9]	Dispensed prescription items	Percent of all dispensed prescription items	Percent of prescription items dispensed by a listed practice [9]
GP practice	1,820,523	98.82%	99.49%
Out of Hours Practice	3,806	0.21%	0.21%
Community health service	2,810	0.15%	0.15%
Other	1,106	0.06%	0.06%
Hospital service	456	0.02%	0.02%
Walk In Centre Practice	389	0.02%	0.02%
Urgent & 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	-	-

Pseudonymisation process

The partnership between NCRAS and NHSBSA has allowed NCRAS to obtain prescribing dispensed in the community information for both cancer patients, and non-cancer patients. The information on non-cancer patients has enabled the information on (anonymous) individuals without cancer to 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.

For peer review only

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 was sent to NCRAS on an encrypted disk. The same pseudonymisation procedure was also run by NCRAS on the identifiers of cancer patients 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 approach [10]). 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 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 re-identify 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-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

For peer review only

inconsistent date orders: for example, the date of patient death is before the date of prescription for 0.3% of dispensed prescription items. Amongst 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 cancer patients 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 one year. Examples of this included a recorded age on the prescription record of 2 years, compared to 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 cancer patients 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. Amongst these 100,424 cancer patients, 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 cancer patients 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 were small. The most marked difference was for age at cancer diagnosis, where, for example, 49% of cancer patients with a dispensed prescription were aged 70-89, but among cancer patients 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

For peer review only

of conditions as they age (therefore the background rate changes), and cancer patients who don't 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 is based on information collected by the NHS.

FINDINGS TO DATE

Timely, record-level community-dispensed prescriptions data provides 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 cancer patients, an investigation of repeat prescriptions to derive measures of prior morbidity status in cancer patients, 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 health care 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-

For peer review only

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 pre- and post-cancer diagnosis (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 (HES) [18], the RadioTherapy DataSet (RTDS) [19] and the Systemic Anti-Cancer Therapy (SACT) 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 cancer patients differ from the general population.

There are a number of limitations of this data resource, as with all routinely collected healthcare data [21-22]. Firstly, 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

For peer review only

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 (e.g. 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 (CPRD) and The Health Improvement Network (THIN) 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' postcode is only available for dispensed prescriptions from the EPS. In the linked data (cancer patients diagnosed after 1994), the proportion of dispensed prescription items from FP10 forms, therefore without the day of prescription or prescribers' postcode 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 6th 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,

For peer review only

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 cancer patients 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 1998 [30] (to be superseded by the General Data Protection Regulation (EU) 2016/679) which will take effect on 25th May 2018) [31] and the 7 Caldicott Principles [32]. Applications to access this linked prescriptions data for cancer patients should be directed through the ODR [33], and application forms are available on their website [33].

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

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 protections 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 2002 [34]; and that appropriate ethical assurances are met.

CONCLUSIONS

The need to understand the healthcare of cancer patients in the community is paramount, both following their diagnosis and to improve earlier diagnosis. Re-purposing 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.

For peer review only

FIGURE LEGENDS

Figure 1: Pseudonymisation and linkage process for the dispensed prescriptions data and cancer registry data

Figure 2: Representativeness of linked cancer registry and dispensed prescription data as compared to cancer registry data alone, by key patient and tumour characteristics (age at diagnosis, ethnicity, sex, tumour site and stage at cancer diagnosis)

CONTRIBUTORSHIP STATEMENT

Conception and design of the work: KH, RB, JR, LEB, GL, MM and TR. Acquisition and pseudonymisation of the data: KH, BS, PG, AH, KH, GM, NM and RR. Analysis of the data: KH, VC 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.

COMPETING INTERESTS

All authors declare that they have no conflicts of interest in relation to this work.

FUNDING

There is no funding to declare in relation to this work.

ACKNOWLEDGEMENTS

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

For peer review only

part of Public Health England (PHE). GL is supported by a Cancer Research UK Advanced Clinician Scientist Fellowship (C18081/A18180).

REFERENCES

[1] Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol* 2017; 46(4):1093-1093i. doi: 10.1093/ije/dyx015.

[2] NHS Prescription Services. Drug Tariff. Available from: https://www.nhsbsa.nhs.uk/pharmacies-gp-practices-and-appliance-contractors/drug-tariff [Accessed 21/07/2017]

[3] NHS Digital. Prescriptions Dispensed in the Community, England 2004-14. Available from: http://content.digital.nhs.uk/catalogue/PUB17644/pres-disp-com-eng-2004-14-rep.pdf [Accessed 21/07/2017]

[4] National Institute for Health and Care Excellence (NICE). Multimorbidity and polypharmacy, Key therapeutic topic. Available from: https://www.nice.org.uk/guidance/ktt18/resources/multimorbidity-and-polypharmacy-pdf-58757959453381 [Accessed 21/07/2017]

[5] National Cancer Registration and Analysis Service. Patient Information. Available from: http://www.ncras.nhs.uk/patientinfo/ [Accessed 21/07/2017]

[6] NHS Digital. Statistics and progress. Available from: https://digital.nhs.uk/Electronic-Prescription-Service/statistics-and-progress [Accessed 08/11/2017]

[7] Joint Formulary Committee. British National Formulary. 68. London: BMJ Group and Pharmaceutical Press; 2014.

[8] NHS Prescription Services. BNF Classification and Pseudo Classification Used by the NHS Prescription Services. Available from: https://www.nhsbsa.nhs.uk/sites/default/files/2017-04/BNF_Classification_Booklet-2017.pdf [Accessed 21/07/2017]

For peer review only

[9] NHS Digital. Data for GP Practices. Available from: https://digital.nhs.uk/organisation-data-service/data-downloads/gp-data [Accessed 21/07/2017]

[10] University of Nottingham. OpenPseudonymiser. Available from: https://www.openpseudonymiser.org/Default.aspx [Accessed 21/07/2017]

[11] National Institute of Standards and Technology (NIST). FIPS Publication 180-4: Secure Hash Standard (SHS); 2015. Available from: http://dx.doi.org/10.6028/NIST.FIPS.180-4

[12] NHS Business Services Authority. NHSBSA Prescription Processing Accuracy, December 2016 Processed Prescriptions. Available from: https://cms.nhsbsa.nhs.uk/sites/default/files/2017-05/Accuracy%20Measures%20for%20Publication%20-%20new%20website.pdf [Accessed 21/07/2017]

[13] Henson K, Shand B, Hunter K, Coupland V, Godfrey P, Elliss-Brookes E, Brock R. 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.

[14] Alvarez-Madrazo S, McTaggart S, Nangle C, Nicholson E, Bennie M. Data Resource Profile: The Scottish National Prescribing Information System (PIS). *Int J Epidemiol* 2016; 45(3):714-715f. doi: 10.1093/ije/dyw060.

[15] Guldbrandt LM, Møller H, Jakobsen E, Vedsted P. General practice consultations, diagnostic investigations, and prescriptions in the year preceding a lung cancer diagnosis. *Cancer Medicine* 2017; 6(1):79-88. doi:10.1002/cam4.965.

[16] Pottegård A, Hallas J. New use of prescription drugs prior to a cancer diagnosis. *Pharmacoepidemiol Drug Saf* 2017; 26(2):223-227. doi:10.1002/pds.4145.

[17] Hansen PL, Hjertholm P, Vedsted P. Increased diagnostic activity in general practice during the year preceding colorectal cancer diagnosis. *Int J Cancer* 2015; 137: 615–624. doi:10.1002/ijc.29418

For peer review only

For peer review only - http://bmjopen.bmj.com/site/about/guidelines.xhtml

[18] NHS Digital, Hospital Episode Statistics. http://content.digital.nhs.uk/hes [Accessed 22/02/2018]

[19] NHS Digital Data Set: Radio Therapy dataset. http://www.datadictionary.nhs.uk/data_dictionary/messages/clinical_data_sets/data_sets/radiotherapy_data_set_fr.asp?shownav=1 [Accessed 22/02/2018]

[20] NHS Digital Data Set: Systemic Anti-Cancer Therapy 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 22/02/2018]

[21] Pottegård A, Schmidt SA, Wallach-Kildemoes H, Sørensen HT, Hallas J, Schmidt M. Data Resource Profile: The Danish National Prescription Registry. *Int J Epidemiol* 2016; 46(3): 798-798f. doi: 10.1093/ije/dyw213

[22] Johannesdottir SA, Horváth-Puhó E, Ehrenstein V, Schmidt M, Pedersen L, Sørensen HT. Existing data sources for clinical epidemiology: The Danish National Database of Reimbursed Prescriptions. *Clin Epidemiol* 2012; 4:303-313. doi:10.2147/CLEP.S37587.

[23] The King's Fund, The UK private health market; 2014. Available from https://www.kingsfund.org.uk/sites/default/files/media/commission-appendix-uk-private-health-market.pdf [Accessed 22/02/2018]

[24] Ministry of Justice. Official Statistics, Prison population figures: 2018. Available from https://www.gov.uk/government/statistics/prison-population-figures-2018 [Accessed 22/02/2018]

[25] NHS Commissioning, Direct Commissioning Change Projects Team. Strategic direction for health services in the justice system: 2016-2020. 2016 Available from: https://www.england.nhs.uk/wp-content/uploads/2016/10/hlth-justice-directions-v11.pdf [Accessed 22/02/2018]

For peer review only

[26]Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, Smeeth L. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* 2015; 44(3):827-36. doi: 10.1093/ije/dyv098.

[27] The Health Improvement Network (THIN) Research Team. THIN Database. Available from: https://www.ucl.ac.uk/pcph/research-groups-themes/thin-pub/database [Accessed 21/07/2017]

[28]NHS Health Research Authority. Section 251 and the Confidentiality Advisory Group (CAG). Available from: http://www.hra.nhs.uk/about-the-hra/our-committees/section-251/ [Accessed 21/07/2017]

[29] The rules are extrapolated from the decisions of the courts

[30] Great Britain. Data Protection Act. London: Stationery Office; 1998.

[31] European Union. General Data Protection Regulation; 2018. Available from https://www.eugdpr.org/ [Accessed 22/02/2018]

[32] Department of Health. Information: To share or not to share? The Information Governance Review. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/192572/2900774_InfoGovernance_accv2.pdf [Accessed 21/07/2017]

[33] Public Health England. The Office of Data Release. Available from: https://www.gov.uk/government/publications/accessing-public-health-england-data/about-the-phe-odr-and-accessing-data [Accessed 21/07/2017]

[34] Great Britain. The Health Service (Control of Patient Information) Regulations 2002. Available from http://www.legislation.gov.uk/uksi/2002/1438/made?view=plain [Accessed 22/02/2018]

For peer review only

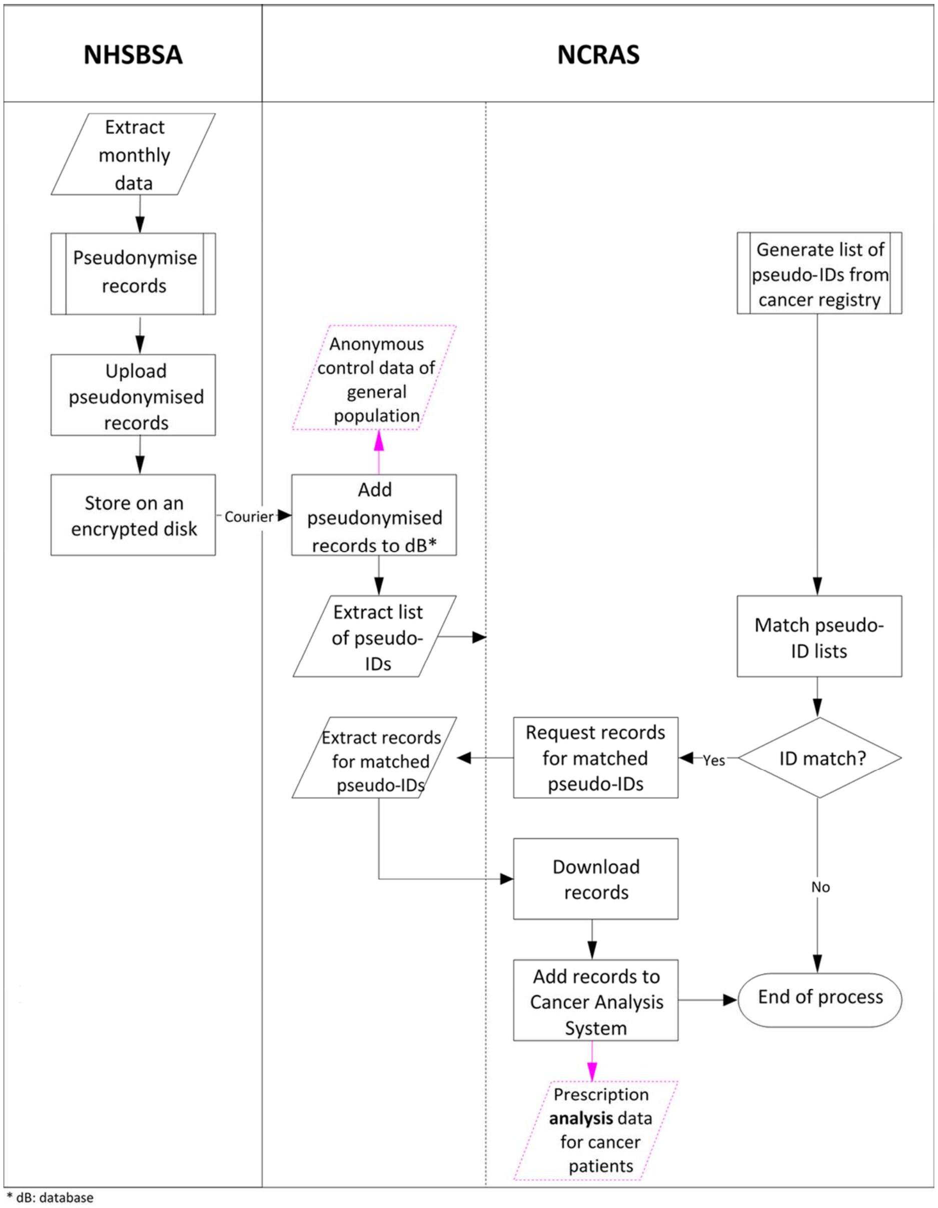

Figure 1: Pseudonymisation and linkage process for the prescription and cancer registry data

99x128mm (300 x 300 DPI)

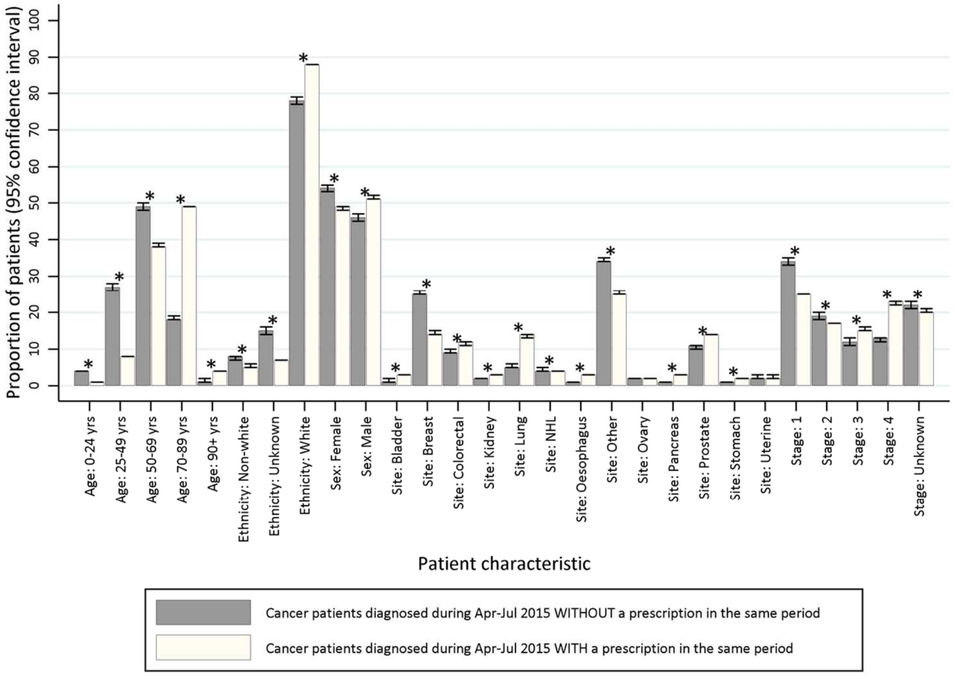

Figure 2: Representativeness of linked cancer registry and prescription data, as compared to a cohort of cancer patients using cancer registry data alone, by key patient and tumour characteristics (age at diagnosis, ethnicity, sex, tumour site and stage at cancer diagnosis)

121x88mm (300 x 300 DPI)

w only