PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Cohort profile: prescriptions dispensed in the community linked to the national cancer registry in England
AUTHORS	Henson, Katherine; Brock, Rachael; Shand, Brian; Coupland, Victoria; Elliss-Brookes, Lucy; Lyratzopoulos, Georgios; Godfrey, Philip; Haigh, Abigail; Hunter, Kelvin; McCabe, Martin; Mitchell, Graham; Monckton, Nina; Robson, Robert; Round, Thomas; Wong, Kwok; Rashbass, Jem

VERSION 1 – REVIEW

REVIEWER	Janick Weberpals German Cancer Research Center (DKFZ), Germany
REVIEW RETURNED	07-Feb-2018

GENERAL COMMENTS	The manuscript by Henson et al. describes the linkage of community pharmacy based prescription/dispensing data to the national cancer registry of England. This database linkage offers an exciting new opportunity for population-based pharmacoepidemiological research which has -to my knowledge- not existed before with data from the UK. This might become be a valuable resource for epidemiologist working on cancer drug associations. However, the manuscript could benefit by considering the following points: MAJOR 1. In my opinion the biggest limitation of the database so far is the very short available follow-up from April to July 2015 which makes it rather a cross-sectional database than a longitudinal. However, the authors state that more follow-up data will be available in due course. Maybe they could elaborate a bit more on that? And why does the database linkage not start earlier? 2. The authors use the words „prescription“ and „dispensing“ quite interchangeably throughout the manuscript and for readers it might be confusing to understand if the drug data is based on prescriptions or actual dispensings. This might make a difference for interpretation of results in terms of adherence, since a big chunk of prescribed drugs are usually never picked up at the pharmacy. The authors might want to be consistent in their phrasing throughout the manuscript. 3. In the abstract the authors should add the information on how the two databases are linked (personal identifier). 4. Table 1: What does „Completeness“ refer to? Including (overall) or excluding EPS? Maybe clarify this in a table footnote 5. In the description of the variables included in the drug prescription/dispensing database the BNF code is mentioned as a

	code/classifier for the unique identification of the active compound. In ATC coding the ATC code also implies the indication that the active compound is used for since it's hierarchically classified by pharmacological properties. Is that also the case for the BNF, which might be a solution for the authors' stated limitation that the drug indication might not be clear if approved for multiple indications? 6. As this resource is a very interesting opportunity due its whole population coverage, could the authors also explain a bit further if the database linkage might be used from researchers outside the UK as well (e.g. by handing in study proposals or similar (like it is done with CPRD data))? 7. Also a linkage to CPRD data might be of high interest; would this be possible using the personal identifier code? MINOR 3. Last point in „Strengths and limitations of this study" page 3: maybe rephrase the sentence to „For pharmacovigilance studies, one should be aware that dispensed drug information…." 4. Page 4 line 45: maybe replace „skilled" with „trained"?

REVIEWER	Kuo, Chang-Fu Associate professor of Rheumatology, Chang Gung Memorial Hospital, Taiwan., Honorary clinical associate professor, University of Nottingham,UK
REVIEW RETURNED	20-Feb-2018

GENERAL COMMENTS	This paper described the characteristics of prescription data linked to the England national cancer registry and the current and future research usage. In general, this cohort profile is comprehensive. I have several minor comments: 1. In the first paragraph of methods, the authors reported that half of the prescriptions were from FP10 or EPS. IS it possible that the linked data may miss prescriptions prepared by other methods? and what is the influence on the estimates in research using the data? 2. NCR contains information about death and causes of death. Any other relevant information that can assist research using prescription data? Can the authors prepare a brief description? The prescription data are from the GPs, any data from the secondary care? Some medications such as targeted anticancer drugs probably are not widely available in GPs. 3. Since the prescription data is near complete, is there any plan to expand the data linkage to other registries? such as MINAP?

VERSION 1 – AUTHOR RESPONSE

Comments by Referee	Authors Response	Location
Referee #1		
The manuscript by Henson et al. describes the linkage of community pharmacy based	Thank you for your considered comments.	N/A

prescription/dispensing data to the national cancer registry of England. This database linkage offers an exciting new opportunity for population-based pharmacoepidemiological research which has -to my knowledge- not existed before with data from the UK. This might become be a valuable resource for epidemiologist working on cancer drug associations. However, the manuscript could benefit by considering the following points:		
MAJOR		
1. In my opinion the biggest limitation of the database so far is the very short available follow-up from April to July 2015 which makes it rather a cross-sectional database than a longitudinal. However, the authors state that more follow-up data will be available in due course. Maybe they could elaborate a bit more on that? And why does the database linkage not start earlier?	We agree that the current biggest limitation is the time coverage of the prescriptions data, though as we state in the manuscript we are currently in the process of extending the linkage. Updated linked data will be available this year. We are in the process of finalising intermediate timescales, but the objective is to receive updated prescriptions data from NHSBSA on a quarterly basis, with an approximate lag to real-time of six months. We have adjusted the text to clarify this point. 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. The database linkage does not start earlier as NHS numbers are not available in the prescriptions data prior to April 2015. This was described on page 4 in the following text: “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	Page 7 lines 6-8

	data."	
2. The authors use the words „prescription" and „dispensing" quite interchangeably throughout the manuscript and for readers it might be confusing to understand if the drug data is based on prescriptions or actual dispensings. This might make a difference for interpretation of results in terms of adherence, since a big chunk of prescribed drugs are usually never picked up at the pharmacy. The authors might want to be consistent in their phrasing throughout the manuscript.	Thank you for your comment, and on reflection we agree that this may be confusing. We have adjusted the text accordingly to 'dispensed prescriptions' throughout.	Various
3. In the abstract the authors should add the information on how the two databases are linked (personal identifier).	This has now been added. ", linked using a pseudonymised version of the patient's NHS number and date of birth"	Page 2 lines 12-13
4. Table 1: What does „Completeness" refer to? Including (overall) or excluding EPS? Maybe clarify this in a table footnote	Completeness is the number of items with a known value as a percentage of the total number of all items (FP10 and EPS). This has now been included as a footnote. [a] the number of dispensed prescription items with a known value as a percentage of all dispensed prescription items (both FP10 and EPS)	Page 7 lines 1-2
5. In the description of the variables included in the drug prescription/dispensing database the BNF code is mentioned as a code/classifier for the unique identification of the active compound. In ATC coding the ATC code also implies the indication that the active compound is used for since it's hierarchically	This is an interesting comment which we considered carefully. Unfortunately, it is not the case for the BNF as the administration schedule is not included. We took Zoledronic Acid (a bisphosphonate) as an example, as there are multiple indications for this drug. In the prescriptions data, there are a number of item types for Zoledronic Acid: Zoledronic Acid_I/V Inf 800mcg/ml 5ml Vl	N/A

classified by pharmacological properties. Is that also the case for the BNF, which might be a solution for the authors' stated limitation that the drug indication might not be clear if approved for multiple indications?	Zoledronic Acid_I/V Inf 40mcg/ml 100ml Zoledronic Acid_I/V Inf 40mcg/ml 100mlVl According to the BNF online, the indications and dose are as follows. **For Aclasta®** **Treatment of Paget's disease of bone** By intravenous infusion, For Adult 5 mg as a single dose, to be administered over at least 15 minutes, at least 500 mg elemental calcium twice daily (with vitamin D) for at least 10 days is recommended following infusion. **Treatment of postmenopausal osteoporosis and osteoporosis in men (including corticosteroid-induced osteoporosis)** By intravenous infusion, For Adult 5 mg once yearly as a single dose, to be administered over at least 15 minutes, in patients with a recent low-trauma hip fracture, the dose should be given 2 or more weeks following hip fracture repair; before first infusion give 50000–125000 units of vitamin D. **For Zometa® infusion** **Reduction of bone damage in advanced malignancies involving bone** By intravenous infusion, For Adult 4 mg every 3–4 weeks, to be administered over at least 15 minutes, calcium 500 mg daily and vitamin D 400 units daily should also be taken. **Hypercalcaemia of malignancy** By intravenous infusion, For Adult 4 mg for 1 dose, to be administered over at least 15 minutes.	

	Therefore, we are unable to match the information that we have in the linked data to the specific recommended doses and our stated limitation holds. We will, however, investigate further a linkage to the ATC classification.	
6. As this resource is a very interesting opportunity due its whole population coverage, could the authors also explain a bit further if the database linkage might be used from researchers outside the UK as well (e.g. by handing in study proposals or similar (like it is done with CPRD data))?	We are pleased to hear that the reviewer thinks this resource may have international interest. Researchers outside the UK can submit applications to the Office for Data Release (ODR), with appropriate safeguards. These have now been detailed in the manuscript. 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; and that appropriate ethical assurances are met. Whilst considering this, we have also amended the manuscript to include a reference to the General Data Protection Regulation (GDPR). (to be superseded by the General Data Protection Regulation (EU) 2016/679) which will take effect on 25th May 2018)	Page 15 lines 1-7 Page 14 lines 21-22
7. Also a linkage to CPRD data might be of high interest; would this be possible using the personal identifier code?	Currently this is not available, though it would be feasible as CPRD data is linked to cancer registration data. This linkage is governed by CPRD, and any extension to this linkage would need to satisfy requirements under the Common Law Duty of Confidentiality, Data Protection Act 1998 (soon to be superseded by General Data Protection Regulation (GDPR)), Caldicott Principles and adequate contract provisions between Public Health England, NHS Business Services Authority and CPRD (Clinical Practice Research DataLink).	N/A

	As this has not been discussed with the relevant organisations, we did not feel that it was appropriate to include it in the main manuscript.	
MINOR		
3. Last point in „Strengths and limitations of this study“ page 3: maybe rephrase the sentence to „For pharmacovigilance studies, one should be aware that dispensed drug information….“	Thank you. The text has been amended accordingly.	Page 3 line 18
4. Page 4 line 45: maybe replace „skilled“ with „trained“?	We have amended the text to ‘trained’	Page 4 line 20
Referee #2		
This paper described the characteristics of prescription data linked to the England national cancer registry and the current and future research usage. In general, this cohort profile is comprehensive. I have several minor comments:	Thank you for your supportive and constructive comments.	N/A
1. In the first paragraph of methods, the authors reported that half of the prescriptions were from FP10 or EPS. IS it possible that the linked data may miss prescriptions prepared by other methods? and what is the influence on the estimates in research using the data?	Thank you for your comment. To clarify, in the linked data at least 24% of the dispensed prescription items per month are from EPS. The remaining items are from FP10. The proportion of prescriptions dispensed as EPS rather than FP10 has increased to approximately 50% in 2018, therefore increasing the quality of the data captured, but not the completeness. This was discussed in the strengths and limitations section, but we have now provided further clarification. 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.	Page 13 line 14

	To our knowledge, there are a few additional situations where drugs would be prescribed, but NHS Prescriptions Services wouldn't capture it. - Dispensed prescriptions without an NHS number recorded on the prescription. This is approximately 10% of all dispensed prescriptions. We discussed the impact of this missing information in the Data Quality section. - If the prescription was dispensed by a private pharmacy. We are unaware of published statistics that document the proportion of private prescriptions dispensed, but the proportion of private GP consultations are around 3% of all GP consultations, therefore we would assume that the proportion of private prescriptions dispensed would be equivalently low. - If the prescriptions was dispensed in prison. This information is held by NHS England (Health and Justice). The linked data will however cover prescriptions that were written in prison, but dispensed in the community setting. The overall impact of missing prison prescribing information is minimal as the prison population is 0.1% of the population. However, one must consider the impact for studies of certain drug groups, where the prison population prevalence is disproportionally high, for example mental health. - If the items were prescribed in England but dispensed in Wales, Scotland, Northern Ireland and Isle of Man. - If the items were prescribed but not presented for dispensing to a community pharmacy, or not submitted to NHS Prescription Services by the dispenser. - If the prescription was prepared and dispensed in secondary care, or another healthcare institutions. However, the aim of this partnership and data linkage was to understand primary care prescribing activity among cancer patients, and so we do not feel that this is a limitation. We have included sentences (with appropriate references) to reflect this in the limitations section. 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 would not be reimbursed.	Page 12 lines 20-25

	[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]	
2. NCR contains information about death and causes of death. Any other relevant information that can assist research using prescription data? Can the authors prepare a brief description? The prescription data are from the GPs, any data from the secondary care? Some medications such as targeted anticancer drugs probably are not widely available in GPs.	Thank you for this point. NCRAS holds a number of datasets, which are either collected by NCRAS or by NHS Digital. These datasets are all linked at either the patient or tumour level. This includes Hospital Episode Statistics (HES) which details inpatient, outpatient and accident and emergency hospital admissions, and the Systemic Anti-Cancer Therapy (SACT) dataset, which details all chemotherapy drugs. We have now included a sentence detailing this, with appropriate references. The 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]. [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	Page 12 lines 13-16

	22/02/2018]	
3. Since the prescription data is near complete, is there any plan to expand the data linkage to other registries? such as MINAP?	As detailed above, the cancer registry holds a wealth of linked data resources. Linkage to other registries and datasets is considered, within the appropriate information governance framework, on an ongoing basis to facilitate our organisational objectives. However, there are no additional partnerships underway that would be relevant to report in this manuscript.	N/A
Editor		
Please make it more clear in your manuscript that there are no findings yet to date.	Thank you for your comment. We have now added a sentence to clarify that the findings have not yet been published. This work has not yet (as of March 2018) been published in peer-reviewed journals.	Page 11, lines 17-18
Please include Figure legends at the end of your main manuscript.	These are now included. 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)	Page 16, lines 1-6

VERSION 2 – REVIEW

REVIEWER	Kuo, Chang-Fu Division of Rheumatology, Orthopaedics and Dermatology, School of Medicine, University of Nottingham
REVIEW RETURNED	26-Mar-2018

GENERAL COMMENTS	The author response is satisfactory. I have no more comment.