BMJ open
accessible medical research

Using free text information to explore how and when GPs code a diagnosis of ovarian cancer: an observational study using primary care records of patients with ovarian cancer

A Rosemary Tate,[1] Alexander G R Martin,[2] Aishath Ali,[1] Jackie A Cassell[1]

▸ Prepublication history for this paper is available online. To view these files please visit the journal online (http://bmjopen.bmj.com).

Received 18 November 2010
Accepted 21 January 2011

This final article is available for use under the terms of the Creative Commons Attribution Non-Commercial 2.0 Licence; see http://bmjopen.bmj.com

[1]Division of Primary Care and Public Health, Brighton and Sussex Medical School, Falmer, Brighton, UK
[2]Barts and the London NHS Trust, London, UK

Correspondence to
A Rosemary Tate;
r.tate@bsms.ac.uk

ABSTRACT

Background: Primary care databases provide a unique resource for healthcare research, but most researchers currently use only the Read codes for their studies, ignoring information in the free text, which is much harder to access.

Objectives: To investigate how much information on ovarian cancer diagnosis is 'hidden' in the free text and the time lag between a diagnosis being described in the text or in a hospital letter and the patient being given a Read code for that diagnosis.

Design: Anonymised free text records from the General Practice Research Database of 344 women with a Read code indicating ovarian cancer between 01 June 2002 and 31 May 2008 were used to compare the date at which the diagnosis was first coded with the date at which the diagnosis was recorded in the free text. Free text relating to a diagnosis was identified (*a*) from the date of coded diagnosis and (*b*) by searching for words relating to the ovary.

Results: 90% of cases had information relating to their ovary in the free text. 45% had text indicating a definite diagnosis of ovarian cancer. 22% had text confirming a diagnosis before the coded date; 10% over 4 weeks previously. Four patients did not have ovarian cancer and 10% had only ambiguous or suspected diagnoses associated with the ovarian cancer code.

Conclusions: There was a vast amount of extra information relating to diagnoses in the free text. Although in most cases text confirmed the coded diagnosis, it also showed that in some cases GPs do not code a definite diagnosis on the date that it is confirmed. For diseases which rely on hospital consultants for diagnosis, free text (particularly letters) is invaluable for accurate dating of diagnosis and referrals and also for identifying misclassified cases.

ARTICLE SUMMARY

Article focus

- How much information on ovarian cancer diagnoses is 'hidden' in the free text of primary care records?
- How accurate is the date of diagnosis based only on Read codes?
- How many cases might be misclassified if codes alone are used to identify diagnoses?

Key messages

- Free text contains much extra information on ovarian cancer diagnoses, including the dates on which the patient was investigated and diagnosed in secondary care.
- This information can be used to determine the date at which a diagnosis was notified to the GP and to identify cases that have not been coded.
- For certain disease areas, particularly where specialist care is involved, free text should be used to determine the extent of misclassification associated with both the (coded) date of diagnosis and identification of cases.

Strengths and limitations of this study

- An in-depth analysis of information relating to ovarian cancer diagnoses using free text records from a large primary care database.
- We did not have access to letters that had been scanned in as images, so will have missed some important information.
- We only looked at cases which had been assigned an unambiguous Read code for ovarian cancer and thus will have missed cases with no code or an ambiguous code.
- We ignored text that did not explicitly refer to the patient's ovaries and thus did not investigate pathways of care or symptoms. This is the topic of a separate study which is already underway.
- We only looked at ovarian cancer, and cannot say whether our findings can be generalised to other diseases.

INTRODUCTION

UK primary care databases provide a valuable source of information for research on disease epidemiology, drug safety and adverse drug reactions. The records in these databases contain a mix of coded and free text data.

BMJ Open: first published as 10.1136/bmjopen-2010-000025 on 23 February 2011. Downloaded from http://bmjopen.bmj.com/ on October 6, 2022 by guest. Protected by copyright.

Diseases, symptoms and clinical events are coded using 'Read' codes, enabling searching for clinical entities. Analyses of existing large-scale electronic patient records—most extensively collated in the form of large primary care datasets such as the General Practice Research Database (GPRD), The Health Improvement Network and QResearch—have almost exclusively exploited coded data. Such data are readily accessible to the classical methods of epidemiological analysis, once the complexities of defining and selecting a patient cohort have been overcome.

However, since clinicians can choose whether and how to code a consultation, an unknown amount of clinical data is in effect 'hidden' outside the coded data in the free text notes. Free text records, as distinct from coded records, may contain further information on diagnosis (which have been copied or imported from hospital letters) and are also likely to contain important information on the severity of symptoms or on additional symptoms which have not been coded. The degree to which clinical information is coded, and how this varies between by practitioner, practice, or type of clinical problem, is currently unknown. The aim of this study was to determine how much extra information on ovarian cancer diagnosis is recorded in the free text, how often this is recorded before the event date for which the diagnosis is coded and whether information from free text is needed for more accurate dating of diagnosis in research studies.

METHODS

Data

This study builds on previous work on dating of diagnosis,[1] where we used coded records from the GPRD; one of the largest primary care databases in the UK. The GPRD contains anonymised longitudinal data on a representative sample of the UK population. Records are being collected on over four million active patients who are registered for care in general practice from around 500 primary care practices throughout the UK. These records are created during consultations or processing correspondence, and are widely used in research on disease epidemiology, drug safety and adverse drug reactions.[2]

The target population consisted of all women between 40 and 80 years of age (inclusive) who were alive and registered with a GPRD contributing practice on 1 June 2002. From this population, all women with an incident diagnosis of ovarian cancer recorded during 1 June 2002 to 31 May 2007 (ie, with one of the Read codes: B440.00 (malignant neoplasm of ovary), B440.11 (cancer of ovary) or B44.00 (malignant neoplasm of ovary and other uterine adnexa)) were identified (n=1107). From these we chose 374 patients by randomly selecting one-third of the contributing practices. Of these, 344 patients were used for this study after excluding three cases with a previously ambiguous diagnosis before the study period and 27 patients who had been registered with the GP for <2 years before diagnosis. Full details of the sample selection procedures have been provided by Tate *et al*.[1]

At each consultation the GP may enter one or more Read codes into the computer system and, for each code entered, is given the option to add free text which will be associated with that code. Read codes were developed in the 1980s and are currently used for coding clinical events in primary care in the UK. Each code has an associated text description—for example, 'abdominal pain', 'ovarian cancer', 'letter from specialist', which is available on GP systems (usually as a drop down menu) to help them record the correct code. The GP also has the option to enter a date indicating when the event occurred—that is, the 'event date', if this differs from the date of the consultation. In this paper unless otherwise stated 'date' will refer to the event date.

For our study we obtained anonymised free text records for all consultations recorded during the 12 months before the date of the earliest Read code indicating a referral for, or suspicion of, ovarian cancer (date 4) and up to and including the date of definite diagnosis (date 1), where the dating scheme follows the definitions of our earlier paper—that is,

Date 1. Date of first definite diagnosis—(referred to in this paper as date of coded diagnosis). Earliest recorded (event) date of definite diagnostic code (Read codes as above).

Date 2. Date of first ambiguous diagnosis—Date 1, or, if present, the first date of an 'ambiguous' code (eg, 'cancer', 'carcinomatosis', with no previous cancer diagnosed in another site) if this occurs before, but within 2 years of, date 1.

Date 3. Date GP first knew, or suspected a diagnosis—Date 2, or, if present, first date of code indicating GP already knew of a cancer diagnosis if this occurs before but within 2 years of date 2 (eg, cancer care review, 'chemotherapy' with no previous cancer diagnosis).

Date 4. Date of first suspicion of, or first referral for, ovarian cancer—date 3, or, if present, the first date of a code for an investigation or referral to a gynaecologist if this is earlier than but within 12 months of date 3.

Dates 1 and 4 were different in 73% of the 344 ovarian cancer cases (67% of cases had tests or referrals before the diagnosis date). Full details of the codes which were used to define these dates are given in our earlier paper.[1]

Extraction of information on diagnosis from the free text records

To find information on ovarian cancer diagnoses in the free text, all free text records that referred to the ovary were identified by automatically extracting records containing the fragment 'ovar', ' ov', or ' ov.' (in either upper or lower case). A manual inspection of the results showed that all the matching strings referred to the ovary except for two referring to the drug 'Novartin'. These records were excluded as were six records that referred to a family history of ovarian cancer—that is, in the patient's mother or sister. Textual data recorded on

BMJ Open: first published as 10.1136/bmjopen-2010-000025 on 23 February 2011. Downloaded from http://bmjopen.bmj.com/ on October 6, 2022 by guest. Protected by copyright.

the date of coded diagnosis (ie, text that was associated with the Read code for ovarian cancer or other Read codes recorded on the same date) were then merged with these records. Textual records, together with their associated Read codes, were grouped chronologically by patient ID and if a number of free text records were available for a patient on the same (event) date these were counted as a single record.

Classification scheme for the free text

The first 50 text records and their associated Read code descriptions were then examined by a gynaecological oncologist (AGRM). Since the major purpose of this study was to determine how often the GP recorded a definite diagnosis before coding it, we decided to err on the side of caution when classifying cases as 'definite'. A scheme for classifying a diagnosis recorded in the text records was developed as follows:

1. *Blank*: there is no information in the text relating to diagnosis of either a benign or malignant condition, or that ovarian cancer is suspected.
2. *Benign*: text indicating definite diagnosis of a benign condition—for example, ovarian (cyst)adenoma.
3. *Suspected*: text indicating that ovarian cancer is suspected but with no definite diagnosis yet—for example, 'possible' or 'probable' or 'highly likely'. Alternatively, a surgeon's report (or GP entry relating to the report) may describe, for example, the appearance of ovaries, presence of peritoneal spread, or ascites, which implies suspicion of ovarian cancer. Although surgeons can sometimes be very confident of the diagnosis based on operative appearances, and blood tests/radiology investigations and state the diagnosis as 'ovarian cancer', we classified text as a suspected cancer if there was no mention of histological or cytological confirmation.
4. *Ambiguous*: an ambiguous diagnosis—for example, 'tumour', which might be benign or might be a primary or secondary cancer, or 'metastatic cancer', which might be a primary or secondary ovarian cancer or another type of cancer.
5. *Secondary*: where the subject of the text (and Read code) is a documented primary malignancy of non-ovarian origin.
6. *Definite*: text indicating that a diagnosis of ovarian cancer has been confirmed. This confirmation had to have been made by a histological or cytological confirmation of ovarian cancer—for example, after surgery such as laparotomy, or cytology from ascitic fluid drainage. In the cases where the information relating to how the doctor arrived at the diagnosis is not there—for example, simply 'ovarian cancer', we assumed that it had been confirmed appropriately. We included borderline ovarian cancers in this category, although they are classed as semi-malignant.
7. *Negated*: text indicating specifically that there is no ovarian cancer (despite the Read code).

The full set of selected text records were then inspected and assigned a provisional classification by the first author (ART). These were checked by AGRM who reassigned any that had been incorrectly classified or which were not sufficiently clear cut for a non-specialist to classify. Text that had been recorded on the coded date of diagnosis was double-checked to see whether or not it confirmed the Read code for a diagnosis. Any Read codes for ovarian cancer which had no associated text, or text not relevant to the diagnosis, were assumed to be correct. In addition, each text record was classified as a either a 'letter' or 'GP notes', and if there was information on the stage or grade, this was recorded.

RESULTS

The total number of text records, in the specified time period, was 7860, representing 5777 consultations for 340 of the 344 patients. The median number of text records per patient was 19. Of these, 678 text records (representing 245 patients) were found to contain a reference to the patient's ovary. When these were merged with text recorded on the same date as the coded diagnosis the number of text records increased to 1007, representing 311 patients. The total number of text records, after combining text recorded on the same event date for each patient (and discarding any 'blank' text recorded on the date of diagnosis) was 706 (for 282 patients), 462 of which were recorded before the date of coded diagnosis (191 patients). The analysis was based on these 706 records.

After examining the text records it was clear that information about possible ovarian cancer can be recorded at several different stages of the diagnostic process (table 1). Approximately 25% of the textual data appeared to be electronic letters. These were much more detailed than the GP notes which were often quite terse with misspellings and abbreviations, as illustrated by some of the examples in table 1. However, not all the letters were available as electronic text; many of the records with a Read code indicating a letter just reported the result of the letter, and approximately 5% of free texts indicated that a letter was available elsewhere (eg, as a scanned letter on an image viewer).

According to our classification scheme, 64% of patients had text either recording or confirming a definite diagnosis and 32% had a 'probable' or 'ambiguous' diagnosis. Figure 1 depicts the text classifications in relation to the coded date of diagnosis.

Text classified as 'definite'

The majority of text records that were classified as definite were recorded on the same date as the patient's first ovarian cancer code (205 (60%) patients). However 74 (22%) patients had a 'definite' diagnosis recorded in the text before this date. The median (IQR) difference between the date that the diagnosis was coded and the date of the text for these 74 patients was 24 (8,67) days, with 34 (10%) of patients having a diagnosis more than 4 weeks before. Six patients had text stating that they had a recurrence of previously diagnosed ovarian cancer. All six of these had a previous definite diagnosis in the

BMJ Open: first published as 10.1136/bmjopen-2010-000025 on 23 February 2011. Downloaded from http://bmjopen.bmj.com/ on October 6, 2022 by guest. Protected by copyright.

Table 1 Typical situations and use of text to describe the diagnostic process for patients with ovarian cancer and number of free texts and patients with text (referring to the ovary) according to our classification scheme

Scenario	Example	Classification	Texts	Patients No (%)
Suggestive clinical feature but no mention of ovarian malignancy in text (before coded date)	Seen by GP. Suspected ovarian cyst or ovarian mass—for example, 'lump? ovary' but no mention of cancer	Blank	151*	105 (31)
Specific statement that clinical feature is benign	Text states that cyst or lump is benign or that there is no evidence (yet) of malignancy—for example, 'multiple fibroids', 'thought to have ovarian ca but histology so far has shown benign cyst', 'the curettings were benign'	Benign	7	7 (2)
Referred for investigation of possible ovarian cancer	Seen by GP. Symptoms and signs suspicious of possible ovarian cancer, so referred for urgent scan/blood tests or gynaecology outpatients appointment	Suspected	116	85 (25)
Diagnostic test indicates suspicion of ovarian cancer	Has had scan/blood tests and report (or GP entry relating to report) is suspicious of ovarian cancer			
Specialist's communication states that ovarian cancer very likely	Has been seen at gynaecology outpatient clinic, and consultant letter (or GP entry relating to letter) may state ovarian cancer diagnosis very likely			
Specialist communication after surgery indicates presumptive ovarian cancer, but histology/ cytology awaited	Has had surgery for probable ovarian cancer. Surgeon's report (or GP entry relating to report) may describe, **for example, appearance of ovaries, presence of peritoneal spread, ascites.** The surgeon may be confident of a diagnosis, but is still awaiting a histological (or cytological) confirmation			
A malignancy is suspected or confirmed, but site/origin not yet established	Cancer is suspected, or has been confirmed but the site of the cancer has not yet been established	Ambiguous	50	36 (10)
Metastatic cancer (non-ovarian origin)	Cancer is from another site—for example, 'metastic lobular carcinoma of the breast'	Secondary	5	1 (0)
Histologically or cytologically confirmed ovarian cancer	Histological or cytological confirmation of ovarian cancer has been made—for example, from surgery such as laparotomy, or cytology from ascitic fluid drainage	Definite	374	220 (64)
Text provides additional confirmatory evidence of ovarian cancer (eg, grade, spread)	The Read code for ovarian cancer is supplemented by extra information in the text such as 'grade III', 'sig metastsatic spread', 'chemotherapy'			
No further information available in free text on basis of the coded ovarian cancer diagnosis	Sometimes the information relating to how the doctor has arrived at the diagnosis is not there—for example, simply, 'ovarian cancer'			
Text on date of diagnosis excludes ovarian cancer as a diagnosis (eg, emendations indicating error, or discussion about cancer rather than diagnosis)	For example, 'this consultation has been changed as wrong diagnosis entered', 'Pt very concerned about possible cancer of ovary. … healthy eating and exercise discussed'	Negation	3	3 (1)

*'Blank' text that was recorded on the date of diagnosis was not included in this category.
The misspellings and abbreviations included in the GP notes have been retained in the table.

free text and one had a previous ambiguous code 'carcinoma in situ of the ovary', a year earlier.

Fourteen patients were classified as definite before the derived date of first suspicion or investigation for ovarian cancer (date 4), 10 when this differed from date 1, but only four of these (of which three were described in the text as a recurrent case) had a diagnosis in the text more than 4 weeks before date 4.

BMJ Open: first published as 10.1136/bmjopen-2010-000025 on 23 February 2011. Downloaded from http://bmjopen.bmj.com/ on October 6, 2022 by guest. Protected by copyright.

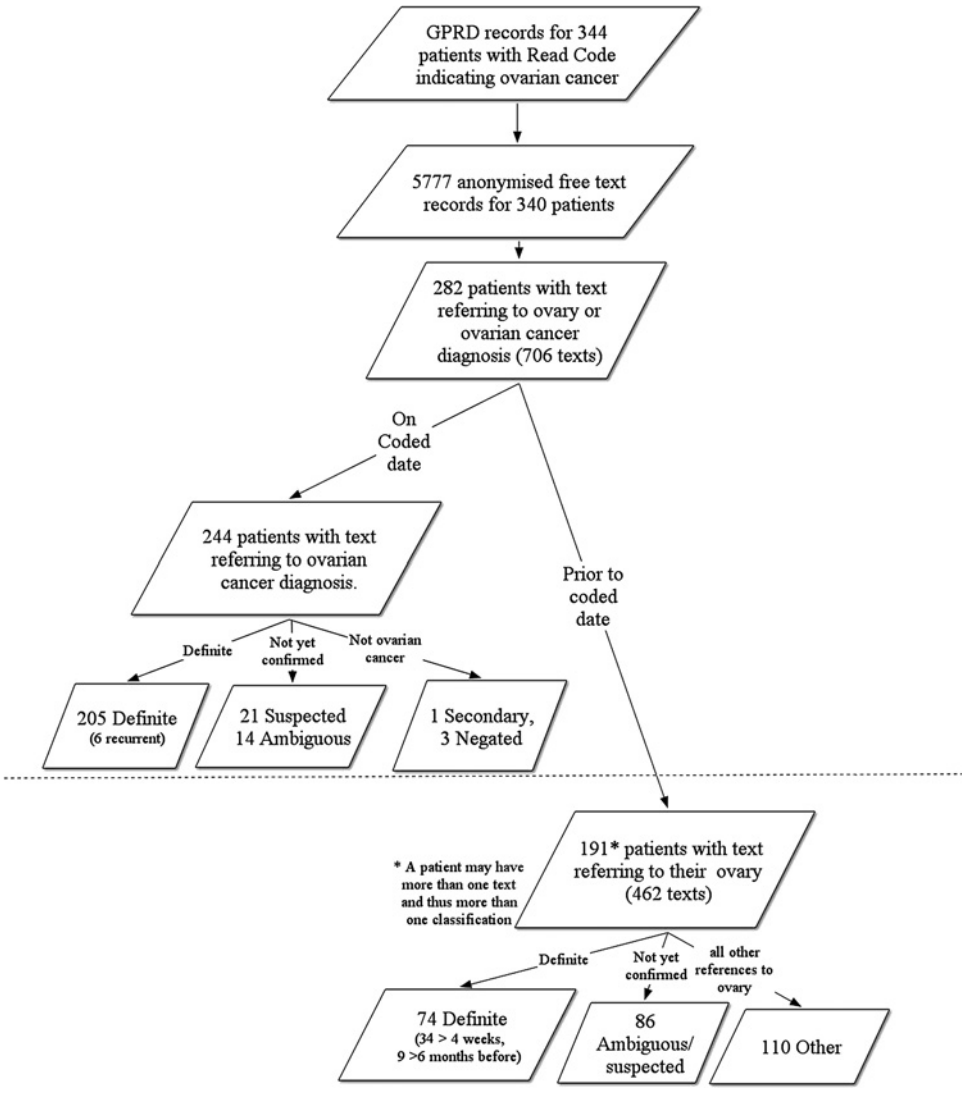

Figure 1 Number of patients and text classifications with event dates on and prior to the coded date of diagnosis.

Approximately two-thirds of the texts indicating a definite diagnosis had been entered in association with a Read code that was not ovarian cancer (including 55 texts recorded on date 1). Approximately one-third of these had a code indicating a visit to an oncologist or gynaecologist or a 'letter from specialist' or similar. The other third were associated with a variety of different codes—for example, a code for an operation or hospital discharge letter, a generic cancer code (eg, 'adenocarcinomas') or a code which bore no relation to the diagnosis at all—for example 'excepted from diabetes quality indicators' (four patients), 'fracture of neck of femur' (one patient).

Twenty-eight per cent of patients had information on the stage, grade or spread of the tumour.

Other classifications

Thirty-nine patients (11%) were classified as not having a definite diagnosis of ovarian cancer on or before the date of coded diagnosis (figure 1). The text records of four patients indicated that there was definitely no ovarian cancer: one had a metastasis in the breast, two a cancer in another site and one was worried she might have cancer, but subsequent coded records for this patient showed no further indication of cancer, so we assumed this had been coded in error. The other 35 patients had only a suspected or ambiguous diagnosis. Examination of subsequent codes for these patients indicated that all these patients did indeed have cancer (19 died or went into a hospice, 15 had a subsequent (repeated) code for ovarian cancer and most had codes for chemotherapy or cancer care).

Thirty-one patients had an ambiguous, or suspected diagnosis before date 4 (17 where date 1 and date 4 were different); the median (IQR) difference was 13 (6, 25) days, with five (3 where date 1 and date 4 were different) patients having this classification more than 4 weeks previously.

DISCUSSION

This analysis shows that primary care records hold a large amount of free text containing information on ovarian cancer diagnoses. The majority (82%) of the 344 patients had free text relating to their ovarian pathology and 64% had free text confirming an ovarian cancer code. However, the information in the text was not

BMJ Open: first published as 10.1136/bmjopen-2010-000025 on 23 February 2011. Downloaded from http://bmjopen.bmj.com/ on October 6, 2022 by guest. Protected by copyright.

always reflected in the codes. In some cases a 'definite diagnosis' in the text field preceded the coded diagnosis, with 22% of patients having a 'definite' classification of ovarian cancer recorded in the text before the date of the first ovarian cancer code. Half of these cases had a 'definite' classification more than 24 days before the coded date. However, only 10 of these diagnoses occurred before our derived date for suspicion of, or referral for, ovarian cancer. A number of other inconsistencies were identified using the free text: four patients did not have ovarian cancer at all and six were recurrences of a much earlier ovarian cancer not evident from the codes.

The delay between the GP recording the diagnosis in the text and coding might be explained partly by incorrect entry of dates or by the administrative practices of the surgery. For example, practice staff code and date the letters when they arrive and the GPs assigns the cancer code at a later date. This latter supposition was supported by the fact that approximately 50% of 'definite' classifications recorded before the coded date came from the text of letters (as opposed to about 25% recorded on the date of diagnosis).

There is no Read code for a possible, probable or highly likely diagnosis, and this may explain why 10% of the patients classified as having only a suspected or ambiguous diagnosis nevertheless had an ovarian cancer code. Conversely, 'definite' diagnoses in free text were often associated with a very general Read code (eg, 'letter from specialist'). Since most studies of diseases are based on the codes, wrongly or uncoded cases will lead to incorrect estimates of the incidence of the disease and will also have an impact on case selection. In addition, incorrectly entered dates of the notification of diagnosis will have an impact on the findings of studies on the incidence of symptoms and delay before diagnosis. How much difference this will make will depend on the disease and also on the time period used for calculating incidence. For this dataset, redefining the (coded) date of diagnosis using the free text did not have much effect on estimates of delay or incidence of symptoms (data not shown), but this might not be the case for other diseases.

To our knowledge this is the only work which explicitly explores and reports dating of diagnoses in GP records using the textual part of the record. Other studies creatively used code lists—for example, our previous study,[1] or GP questionnaires—for example, the study of Hammad *et al*,[3] to investigate the accuracy of the coded date. A handful of studies have used free text to verify clinical conditions in combination with codes; a few of them have used free text to identify cases[4 5] while the majority have used free text to verify and validate coded information.[3 6–9] However, with the exception of the study by Wurst *et al*,[7] most studies have little or no detail on the process of selection and review of the free text that was used.

In this study we looked only at cases that had an unambiguous diagnosis code for ovarian cancer, so we will have missed any that had not been coded or which had ambiguous codes. Our recent comparison with the cancer registries[10] indicated that around 9% of cases may not have been coded and thus missed in this way. Another limitation was that we did not have access to letters that were not available in electronic text format, which might have contained important information that was not relayed by the GP. We estimate (data not shown) that approximately 20–25% of hospital letters for these patients will not have any information on their content (either a code or text) entered on the date that they were received. It is likely that many of these letters will contain information on diagnosis, particularly for cancer, where a specialist will always make the diagnosis. However, with the increased transfer of electronic records and even sharing of hospital records this problem is likely to be resolved in the near future.

CONCLUSIONS

This study gives an in-depth insight into the extra information that is contained in the free text part of records relating to a suspected or confirmed diagnosis of ovarian cancer. A large amount of information in free text is available that modifies the coded date or discloses incorrect classification as a case, even for 'hard' outcomes such as ovarian cancer, which is considered well documented in primary care records. This shows that (*a*) the quality of information in primary care records is better than one might think, but (*b*) free text needs to be routinely explored to take advantage of this quality information. It is likely that the proportion of information concealed in free text will be greater for less 'hard' outcomes in certain disease areas. We are therefore working with natural language processing experts to find ways of extracting relevant information automatically (and therefore more cost effectively) from large volumes of text.

Acknowledgements We thank the PREP team and John Parkinson for helpful comments and discussions.

Funding This work was supported by the Wellcome Trust (086105/Z/08/Z). Access to the General Practice Research Database was funded through the Medical Research Council's licence agreement with Medicines and Healthcare Products Regulatory Agency (MHRA). The authors were independent from the funder and sponsor, who had no role in the conduct, analysis or the decision to publish. This study is based in part on data from the Full Feature General Practice Research Database obtained under licence from the UK MHRA. However, the interpretation and conclusions contained in this study are those of the authors alone.

Competing interests None.

Ethics approval Access to the dataset was approved by the Independent Scientific Advisory Committee (protocol 07 069).

Contributors ART conceived and wrote the paper, read and classified the free text, and carried out the subsequent analysis. AGRM devised the classification scheme, read through and classified the free text, wrote part of the paper, and provided expert advice. JAC was involved in the conception and writing of the paper. AA participated in writing the paper (including the literature review) and assisted with data management. All authors had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. ART is the guarantor.

Provenance and peer review Not commissioned; externally peer reviewed.

BMJ Open: first published as 10.1136/bmjopen-2010-000025 on 23 February 2011. Downloaded from http://bmjopen.bmj.com/ on October 6, 2022 by guest. Protected by copyright.

Data sharing statement Codelists and statistical code available from the corresponding author.

REFERENCES

1. Tate AR, Martin AG, Murray-Thomas T, *et al*. Determining the date of diagnosis—is it a simple matter? The impact of different approaches to dating diagnosis on estimates of delayed care for ovarian cancer in UK primary care. *BMC Med Res Methodol* 2009;9:42.
2. GPRD. Excellence in public health research. 2010. http://www.gprd.com.
3. Hammad TA, McAdams MA, Feight A, *et al*. Determining the predictive value of Read/OXMIS codes to identify incident acute myocardial infarction in the General Practice Research Database. *Pharmacoepidemiol Drug Saf* 2008;17:1197–201.
4. Masso Gonzalez EL, Johansson S, Wallander MA, *et al*. Trends in the prevalence and incidence of diabetes in the UK: 1996–2005. *J Epidemiol Community Health* 2009;63:332–6.
5. Martinez C, Assimes TL, Mines D, *et al*. Use of venlafaxine compared with other antidepressants and the risk of sudden cardiac death or near death: a nested case-control study. *BMJ* 2010;340:c249.
6. Stowe J, Andrews N, Wise L, *et al*. Investigation of the temporal association of Guillain-Barre syndrome with inuenza vaccine and inuenzalike illness using the United Kingdom General Practice Research Database. *Am J Epidemiol* 2009;169:382–8.
7. Wurst KE, Ephross SA, Loehr J, *et al*. The utility of the general practice research database to examine selected congenital heart defects: a validation study. *Pharmacoepidemiol Drug Saf* 2007;16:867–77.
8. Ruigomez A, Martin-Merino E, Garcia Rodriguez LA. Validation of ischemic cere-brovascular diagnoses in the health improvement network (THIN). *Pharmacoepidemiol Drug Saf* 2010;19:579–85.
9. Charlton RA, Weil JG, Cunnington MC, *et al*. Identifying major congenital malformations in the UK General Practice Research Database (GPRD) a study reporting on the sensitivity and added value of photocopied medical records and free text in the GPRD. *Drug Saf* 2010;33:741–50.
10. Tate AR, Nicholson AC, Cassell JA. Are GPs under-investigating older patients presenting with symptoms of ovarian cancer? Observational study using General Practice Research Database. *Br J Cancer* 2010;102:947–51.

BMJ Open: first published as 10.1136/bmjopen-2010-000025 on 23 February 2011. Downloaded from http://bmjopen.bmj.com/ on October 6, 2022 by guest. Protected by copyright.