

Conference paper

citation and similar papers at core.ac.uk

provided by Informatics in Primary Care (BCS)

An eight-step method for assessing diagnostic data quality in practice: chronic obstructive pulmonary disease as an exemplar

Edwin R Faulconer MB BS
House Officer

Simon de Lusignan MSc MRCP
Senior Lecturer

Primary Care Informatics, Department of Community Health Sciences, St George's Hospital Medical School, London, UK

ABSTRACT

Background Chronic obstructive pulmonary disease (COPD) is an important cause of mortality and morbidity. Its management is shifting from the secondary to the primary care setting. The quality of data is known to vary between practices, and individual practices need to be able to assess their data quality.

Objectives To measure the quality of diagnostic data in COPD.

Subjects 10 975 patients registered with a computerised general practice in the south of England, and 190 patients likely to have COPD.

Methods An eight-step method was developed: (1) research the expected prevalence of the diagnosis and define audit criteria; (2) find out how the diagnosis might be coded – look at the terminology and the codes presented by the computer interface; (3) examine the characteristics of the practice population; (4) calculate the prevalence and infer its reliability; (5) investigate the completeness; (6) accuracy; (7) currency and consistency; and (8)

calculate sensitivity and positive predictive value of the data.

Results The prevalence of COPD in the literature ranges between 3% and 10%. The coding for bronchitis and COPD is complex and it is easy to select an incorrect code. The test population is younger but of similar social class to the national average. The prevalence of COPD in this study was 1.3%. The data were incomplete and some were inaccurate; patients with COPD had to be identified from additional searches. The sensitivity of the use of the diagnostic code was 79%, and the positive predictive value 75.3%.

Conclusions The method provides a tool to help practices and localities assess their diagnostic data quality.

Keywords: computerised medical records, data quality, medical informatics, primary care, pulmonary disease: chronic obstructive, terminology, vocabulary: controlled – classification

Introduction

Chronic obstructive pulmonary disease (COPD) is an important cause of morbidity and mortality within the smoking population.^{1,2} It is predominantly a disease of males, though the incidence of the disease in women is increasing rapidly, probably due to

increased cigarette smoking.³ COPD encompasses both emphysema and chronic bronchitis and is characterised by an incompletely reversible limitation of the airway.⁴ Its severity is graded using spirometry, analysing forced expiratory volume in the first second

(FEV1) and forced vital capacity (FVC), and the FEV1:FVC ratio.⁵ These clinical measurements can be made in primary care, where increasing numbers of people with COPD will be monitored and treated.⁶ Incentives have been provided to accelerate the shift of management of COPD to primary care through the inclusion of a quality target for COPD monitoring in the new United Kingdom (UK) general practice contract.⁷ Best practice has also been reinforced through the production of national guidance for the management of COPD.⁸

The quality of COPD management in primary care will be measured using Read-coded data extracted from general practice (GP) computerised medical records. Such data will be measured by the Quality Management and Analysis System (QMAS) and will have financial rewards linked to it.⁹ There is every incentive for UK general practitioners to improve their data quality ahead of these assessments. The quality of computer data is measured in terms of completeness, accuracy, currency and consistency of recording.^{10,11} There is an emerging consensus that an indicator of data quality is best achieved by looking at three areas: the reliability of data, by comparison with existing databases; combined with calculating sensitivity and positive predictive value.^{12,13} With the exception of data quality probes, which rely on a Boolean comparison of two features (such as prescribing thyroxine *and* diagnosis of hypothyroidism), and the work of the PRIMIS Team, little has been written about how a practice might assess the quality of data about an individual diagnosis.^{14–17} In the case of COPD there is no single drug that is a surrogate marker of the disease, making data quality probes unsuitable; and within the time-scale available it was not possible to address this issue via a data quality programme or educational intervention, the usual way we would recommend improving data quality in practice.^{18,19}

Against the backdrop of a new quality-based contract, and the need to implement best practice, general practice diagnostic data need to be valid and reliable; we therefore developed a tool to assess data quality and used it to measure the quality of COPD diagnostic data.

Method

An overview of the eight steps of the methodology is shown in Table 1. The table also shows the sources of information required to assess diagnostic data quality. The knowledge required is broad. It includes an understanding of the disease, the clinical computer system and its user interface, the coding system or terminology it uses to record data, information about

the practice area, the nature of clinical practice, and how primary care professionals actually use the system.

1 Find out the expected prevalence of the diagnosis and define audit criteria

We performed a literature search using Medline to identify review articles about COPD, and the use of controlled vocabularies in primary care. We also looked up the Office of National Statistics data on the UK prevalence of COPD.²⁰ We then developed audit criteria based on this review. Data were extracted from the computer records under the following fields: gender; year of birth; smoking habit; history of asthma; last spirometry (date, location and results); recent hospital admission (for exacerbation of COPD); recent oral prescription of prednisolone (for an exacerbation of COPD). 'Recent' was defined as within 27 months to coincide with the length of time specified in the new General Medical Services (GMS) Contract.⁷

2 How the diagnosis is recorded

- (a) The terminology/coding system
- (b) The computer interface

The coding system most commonly used in the UK is the Read Terminology, Version 2 (referred to as the Read codes and the process of using them as coding).²¹ This system is hierarchical, consisting of parent and child codes. It is organised into chapters relating to body systems or parts of the clinical process (for example, symptoms, examination findings, and so on). The coding hierarchy and picking lists of terms and codes were investigated to identify which terms were prominent when a clinician typed the terms 'bronchitis', 'emphysema' or 'COPD'. The picking list displayed by the EMIS (Egton Medical Information System), the computer system used, is shown in Figure 1. The 'Clue' search engine was used to check the validity of the codes the EMIS system generated.^{22,23} We asked primary care clinicians how they used this interface and coded COPD patients.

3 Population denominator

The population denominator was calculated from the GP computer system. These data are plotted as an age–sex profile using five-year age bands. This can be done in one of two formats: either as a bar chart or as a population pyramid (see Figure 2). The English population age–sex profile was also plotted on the same graph.²⁴ This enables the pattern of the practice

Table 1 Eight steps to appraise data quality: their purpose and the information required

Title	Purpose	Source of information
1 Prevalence and audit criteria	The expected prevalence is needed for comparison with that of the practice, so that the reliability of the data can be inferred. It enables appropriate audit criteria to be developed.	<i>International/national:</i> Medline/Office of National Statistics (ONS) <i>Local:</i> Public health specialist, data quality facilitator or programme.
2 How the diagnosis is coded: (a) terminology/ coding system (b) computer interface	The same clinical concept might be represented in different parts of the coding scheme. The computer could bias coding by the codes in its picking lists or linked to templates for data entry. Sometimes computer systems have local codes.	<i>Users of the system:</i> users need to be asked how they code the diagnosis and related clinical concepts and data. <i>The clinical system:</i> this needs to be examined carefully to see how it might bias coding choices. <i>Reference copy of terminology:</i> relevant codes in the clinical system need to be checked against a reference.
3 Population denominator	If the denominator is incorrect, any calculation of prevalence, sensitivity or positive predictive value will be inaccurate. Social class and ethnicity data might also be useful.	<i>ONS:</i> 2001 census data for national population. <i>Clinical system:</i> we recommend visually plotting an age–sex profile and checking this against local + national comparators. <i>Users:</i> are usually aware of the social class and ethnicity factors locally.
4 Prevalence and its inference on reliability	Disease prevalence provides an indication of the reliability of the data; however, this requires comparison with the age–sex profile, social class and ethnicity of the practice population.	<i>Use data from sections 1 to 3 above:</i> inter-practice variation in recording levels is to be expected; however, if this variation cannot be explained (by, for example, high social class, low smoking prevalence), then the reliability of the data is in doubt.
5 Completeness	To see if every patient who has the condition has the diagnostic code.	<i>Clinical system:</i> look for surrogate markers of disease (especially numerical pathology results and drugs). <i>Users:</i> how do they code this diagnosis?
6 Accuracy	Have the patients with the diagnostic code really got the disease in question?	<i>Clinical system:</i> investigate whether other clinical data (including written records) confirm the diagnosis.
7 Consistency + currency	Are the data recorded within a reasonable time and in the same way?	<i>Clinical system:</i> extract time series data appropriate to the condition. <i>Users:</i> how often do they think data should be recorded?
8 Sensitivity + positive predictive value (PPV)	Sensitivity is the proportion of people with COPD who actually have a diagnostic code. PPV is the likelihood that a person with the diagnostic label actually has the disease.	<i>Clinical system:</i> searches based on users' comments inform us of the likely sensitivity and PPV of the diagnostic code used. As with prevalence the calculation requires a reliable denominator.

```

Entry : BRONCHITIS

Select option. <Return> to alter synonym : █

A Bronchitis unspecified H30
B Chronic bronchitis H31
C Wheezy bronchitis H302
D Chest infectn-unsp bronchitis H30-1
E Acute bronchitis and bronchiolitis H06
F Acute bronchitis or bronchiolitis NOS H06z
G H/O: bronchitis 14B3-1
H Acute wheezy bronchitis H060-1
I FH: Bronchitis/COAD 12D1
J Recurrent wheezy bronchitis H30-2
K Chronic wheezy bronchitis H3120-1
    
```

Figure 1 Picking list that appears when the term 'bronchitis' is entered

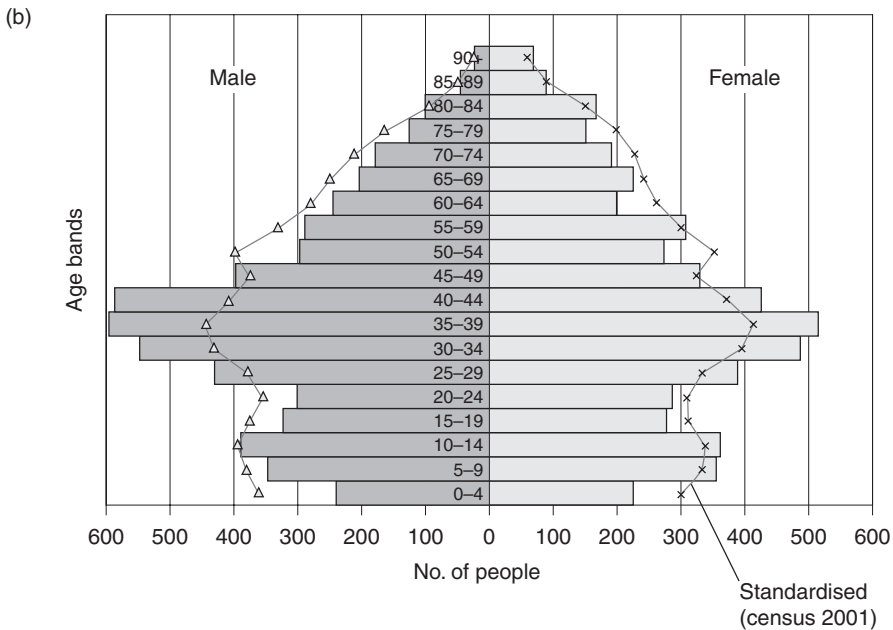
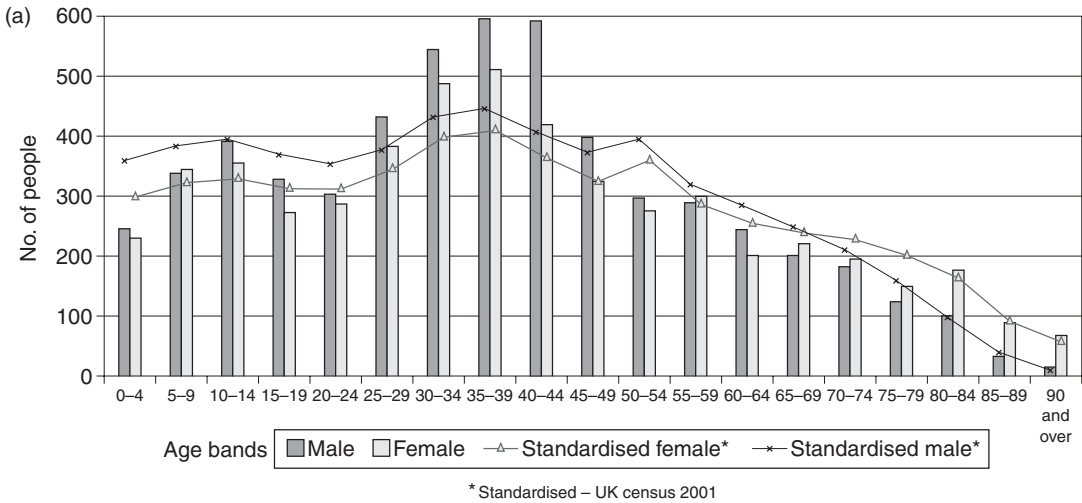


Figure 2 Age-sex profile of study population – data comparing practice and UK population using (a) a bar chart and (b) population pyramid

population to be identified, and to explore its likely influence on prevalence. Information about the social class and ethnicity of the practice relevant to the disease was also collected.

4 Prevalence and its inference on reliability

The prevalence of the disease should be calculated and the age–sex profile and other data held about the practice should enable some inference to be drawn about the reliability of the data. If comparisons are to be made between practices then the population needs to be standardised. This can either be done using the relevant national population or by using the European standard population. The latter has become the standard comparator used within the UK health service.²⁵

5 Completeness of the data

Additional searches were performed to find any patients who were likely to have a diagnosis of COPD, but were not in the audit result due to incorrect coding. Criteria used were: age over 45; repeat prescription for ipratropium or tiotropium; not coded as asthma; not coded as COPD. In addition a search was carried out on patients aged over 45 years who had been newly prescribed salbutamol or other beta-agonist since 1996, when a detailed review of all patients with possible COPD had been carried out.⁶

6 Accuracy of the data

The clinical records of the patients that the computer search identified as having a diagnosis of COPD were examined. This consisted of looking at all the free text in the computer notes to validate the diagnosis; if the information was not found then the written record was examined. Important factors in determining the accuracy of the coding were: age of patient; smoking status; history of asthma; current monitoring and treatment regimens.

7 Consistency and currency

Records of those with an accurate, coded diagnosis of COPD, but who did not have a coded FEV1 measurement in the last 27 months were examined. Free-text entries and hospital outpatient correspondence were read, and date and result of the patient's last spirometry were noted, if recorded.

8 Sensitivity and positive predictive value

Data were collated and used to calculate the sensitivity of searching for patients with a diagnosis of COPD (that is, how many patients with COPD did the search strategy miss?). The positive predictive value was calculated to highlight the effectiveness of using coded data for COPD analysis (that is, the likelihood that someone with a code for COPD actually has the disease).

Results

1 Literature review and audit criteria

The information about COPD and data quality has been included within the introduction; and the audit criteria within the method. The expected prevalence rate is between 3% and 10%, though it is recognised to be under-recorded and under-treated.^{2,26} The objective of the study was to develop a generalisable method to assess diagnostic data quality and to use COPD as an exemplar to test the method.

2 How the diagnosis is coded

The patients' records at this practice are all computerised, and any correspondence is scanned in. The practice uses EMIS as its clinical computer system.²²

The Read codes for bronchitis, COPD and asthma are confusing, making it easy to miscode and create inaccurate diagnoses. All the respiratory chapter codes start with an H. However, bronchitis appears in two parts: H06 for acute and H3z for chronic. For a busy clinician it can be all too easy to select the wrong code (see Figure 1). The hierarchy of Read codes for COPD is such that several diagnoses fall under the broad heading of 'Chronic Obstructive Pulmonary Disease'. Figure 3 illustrates some of the diagnoses that may be found when searching the H3 hierarchy, and the potentially difficult choice that a general practitioner might have to make.

Several of the clinicians in the practice were confused about coding and the coding hierarchy. Several commented on the apparently absurd location of the H33 (asthma) code as a child code of H3 (bronchitis). The template for data entry was correct and linked to appropriate codes. Coding of smoking habit was also examined, and nearly half of the clinicians thought that 137: 'smoking status' was a code that indicated that a patient was an active smoker of some sort. They did not realise that selection of this code was ambiguous, as it contains child codes that included

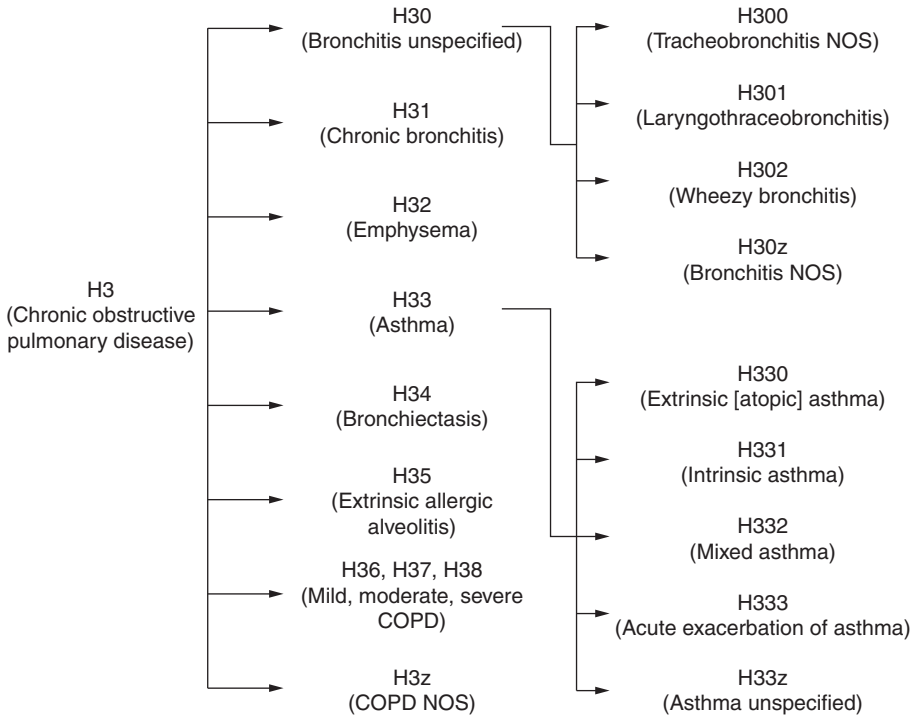


Figure 3 Part of the Read code hierarchy of H3 (chronic obstructive pulmonary disease)

never smoked (137I), passive smoker (137I), chews tobacco (137W), as well as codes for being a current and ex-smoker.

3 Population denominator

The audit was carried out in an undergraduate teaching practice, where ERF spent his final year GP attachment. The practice was involved in COPD research from 1995 to 1997.⁶ This is a six-whole time equivalent practice located in Guildford, Surrey. The practice population is 10 975. The age and sex profile can be seen in Figure 2, and is typical for the south of England. There is a dip in the number of older age-group children and those in their early twenties associated with people moving away for education, university and at the start of their working lives. There is then a 'bulge' in the graph, associated with the 1960s baby boom. Unusually for the UK, there is an excess of men between 40 and 54, associated with those who work in a high-cost housing area leaving their family elsewhere. Ethnicity coding is present for less than 5% of the practice population.

4 Prevalence and the inferred reliability of the data

A total of 190 patients were initially identified as having a diagnosis of COPD, a point prevalence of

18.6 per 1000 registered patients. After the inaccurate diagnoses were excluded and missed diagnoses added this was estimated to be 13.3 per 1000. This is low compared with the expected 3% to 10%, but compatible with the area's relative affluence, relatively small elderly population and low rate of cigarette smoking.

5 Completeness

A total of 38 more patients with COPD were missed from the initial search; these patients were identified by searching for those receiving treatment for COPD who lack a diagnosis. Only 10% (19/190) had FEV1 values within 27 months identified from the computer search (see Figure 4). Smoking data were present for 171/190 (90% of patients identified by the initial search). These data showed 38.6% (66/171) to be current smokers, 38.6% (66/171) to be ex-smokers, 17% (29/171) were coded as having never smoked tobacco and 5.8% (10/171) had the ambiguous higher order 'smoking status' code (137).

6 Accuracy

On inspection of the computerised records it was possible to exclude the diagnosis of COPD in 24.7% (47/190) patients because of an inaccurate diagnosis. These exclusions were: six due to acute bronchitis (H06) being coded as COPD (H3z); 16 due to an

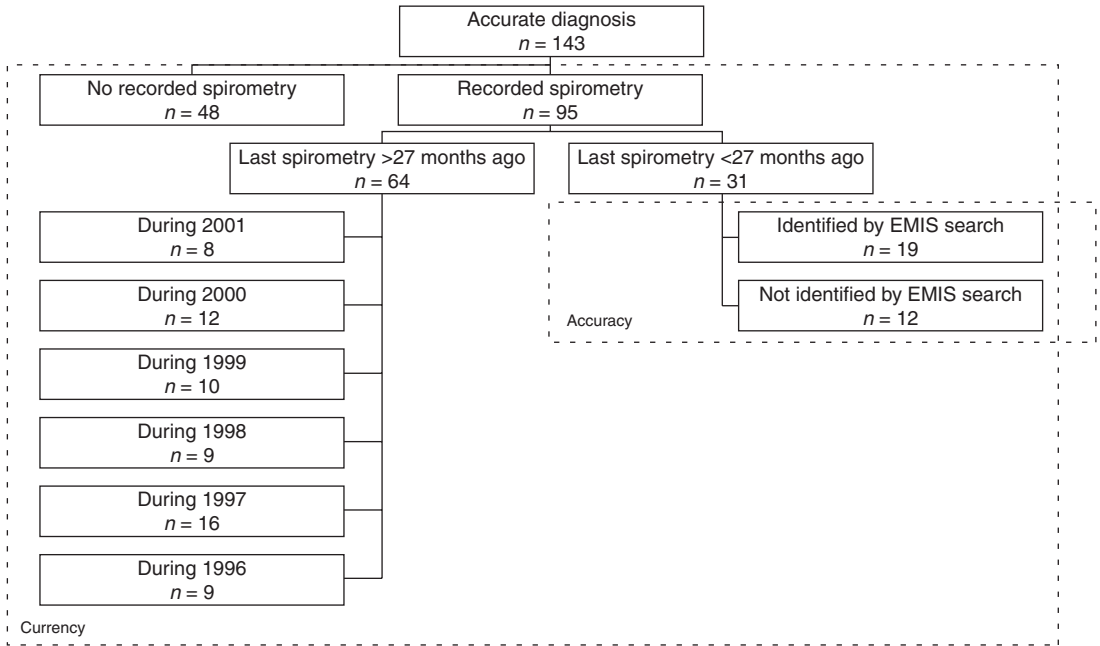


Figure 4 Currency of the FEV1 values for those with an accurately coded diagnosis of COPD

asthma coding and no evidence of concurrent COPD; 25 due to a single coding for COPD in the history, but no investigation or treatment data compatible with the diagnosis. After excluding the inaccurately diagnosed patients there remained 143 COPD patients. These have a mean age of 70 years (range 44–94

years). Figure 5 shows the completeness and accuracy of the audit search in diagrammatic form.

Those with COPD were more likely to be smokers or ex-smokers, and those recorded as having never smoked were much less likely to have COPD (Chi-squared test $P < 0.01$, see Table 2).

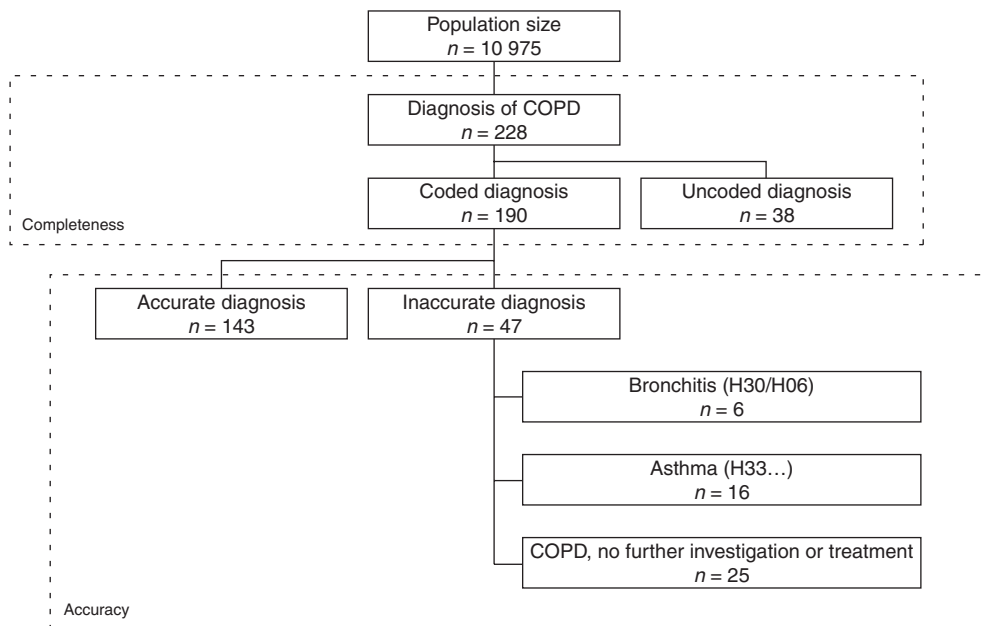


Figure 5 Completeness and accuracy of coded data search

Table 2 Cross-tabulation of smoking habit and likelihood of a computer diagnosis of COPD diagnosis being correct

			COPD computer diagnosis		Totals
			Diagnosis correct	Diagnosis incorrect	
Smoking habit	Current smoker	No.	51	15	66
		%	41.1	31.9	38.6
	Ex-smoker	No.	53	13	66
		%	42.7	27.7	38.6
	Never smoked	No.	14	15	29
		%	11.3	31.9	17.0
	Ambiguous high-level codes	No.	6	4	10
		%	4.8	8.5	5.8
Totals		No.	124	47	171
		%	100.0	100.0	100.0

7 Consistency and currency

Of the 143 accurately coded patients, 31 (21.7%) had FEV1 measurements within the recommended 27-month period. A further 64 (44.8%) had a value more than 27 months old, and the remainder had no recorded values. Figure 4 represents this diagrammatically, and Figure 6 shows the time since last FEV1 measure.

Clinicians were divided about what was appropriate currency for these data. Some felt that an FEV1 and FVC should be recorded once for diagnostic purposes, or an indication given that the diagnosis was based on spirometry performed in secondary care. If a patient had few exacerbations, and was well, there was little enthusiasm for prioritising spirometry over other clinical tasks; but they would comply to fit with their contract. It was clear that the different general practitioners used the diagnosis of COPD differently.

Some were likely to make the diagnosis clinically, especially in a wheezy smoker without a history of asthma or atopy; others were much more reticent. All believed smoking should be recorded, and opinions were divided on spirometry. Prospectively, new COPD diagnoses would only be made based on spirometry.

8 Sensitivity and positive predictive value

The sensitivity (that is, the proportion of those who really have COPD who have a diagnostic code) is 79.0%. The positive predictive value (that is, the likelihood that someone with the diagnostic code actually has the disease) is 75.3%. The figures and the calculation are shown in Table 3.

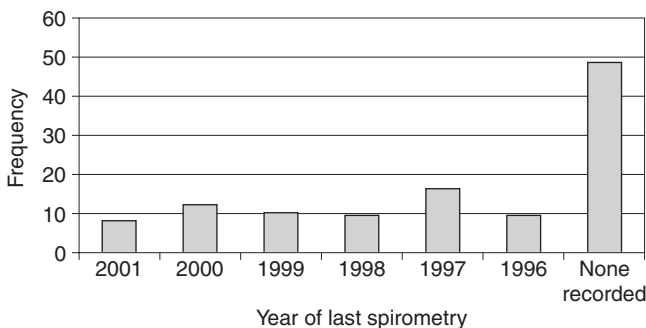
**Figure 6** Time since last spirometry

Table 3 Calculation of Sensitivity and Positive Predictive Value

		'True' diagnosis of COPD		Total
		Yes	No	
Read code diagnosis of COPD	Positive	143 (a)	47 (b)	190
	Negative	38 (c)	10 771 (d)	10 785
Total		181	10 818	10 975
Sensitivity = (a)/((a) + (c))				79.0%
Positive predictive value = (a)/((a) + (b))				75.3%

Discussion

Principal findings

The diagnostic data available has a sensitivity and positive predictive value that suggests it is accurate roughly three-quarters of the time. Sins of commission were as big a problem as sins of omission (that is, coding patients with a disease incorrectly was as big a problem as failing to code patients whom other data suggest had the disease). The weaknesses of these data were their incompleteness, inaccuracy and lack of currency. The idiosyncratic nature of the coding hierarchy in this domain, and lack of clarity in the interface in the clinical computer system between acute and chronic bronchitis, contributed to the data quality being poor.

Implications for practice

Practitioners should be more aware of the coding hierarchy and take care not to label patients inappropriately with a code that means they will be included in the practice COPD population. Hospital outpatient letters and discharge summaries need to include the results of any relevant tests so that the results can be coded in GP electronic records. Contractual changes are going to result in much improved consistency in diagnostic data recording in COPD; however, this may distort coding by reducing primary care professionals' willingness to make a clinical diagnosis.⁷

Limitations of the method

The criteria used to search for COPD patients who did not have the appropriate diagnostic code might not have been exhaustive. The population size of the GP surgery used was such that it was not feasible to look through all of the records individually, and it is acknowledged that the apparent prevalence of COPD could have been higher if this had been done. COPD, FEV1 and smoking results were the only variables that have been analysed.

The study was conducted in a single practice which is relatively advanced in its use of information technology; the findings therefore may not be generalisable to practices with lower levels of data recording unless there is time to carry out exhaustive searches of the written records. We considered extracting free-text entries from the computerised record, and then searching these records using a standard spreadsheet package for key words; for example, 'FEV1'. It is also possible to run exported free text through more sophisticated natural language processing (NLP) packages such as CliniViewer, or to use other more sophisticated methods to explore clinical data (for example, KNAVE-II and ISABEL).²⁷⁻²⁹ We rejected this approach as our aim was to produce a tool that could be readily used by a non-expert; however, this decision should be reconsidered as this technology becomes more accessible to ordinary users.

Comparison with literature

Little has been published about data quality in COPD. However, studies of the prevalence of COPD find airways obstruction in as many as 24.3% of smokers aged over 40, and with severe obstruction in over 5%.³⁰ It would appear likely that prevalence of COPD is underestimated. The sensitivity and positive predictive values obtained for COPD in this study are much lower than most of those obtained by Hassey *et al* in 2001; however, they did not include COPD in their basket of diagnoses, perhaps because of some of the associated coding difficulties.¹² Although the literature on what makes a data quality programme effective has been reviewed, little has been written about how to go about examining diagnostic data quality at the individual practice or locality level, other than the work done by the PRIMIS Team.^{19,31} Brown and Warmington have used data quality probes extensively in their own practice to improve data quality.¹⁴ We could not identify ways that that approach could be used here because of the lack of a drug or widely-recorded test result that is readily associated with the diagnosis. Finding people who have never smoked with COPD is not unusual, and our proportion of those who had never smoked (11.5%) was less than that found by Meyer *et al* (16.7%).³²

Call for further research

Other disease areas and clinical systems should be examined to see if the problems are limited to one clinical system and one specific respiratory disease. Further research should investigate the influence of financially incentivised quality targets on clinical coding.

Conclusions

COPD is a diagnosis which is not ideally situated within the coding hierarchy, and without well-recorded surrogate markers, yet it is one where there is scope for quality improvement. The inclusion of COPD in the UK's new quality-based contract for general practice will ensure its prominence in general practitioners' minds. This method provides a tool for the measurement of data quality in a single practice, taking into account the broad range of factors which might affect data quality.

ACKNOWLEDGEMENTS

We would like to thank the staff and partners at Woodbridge Hill Surgery who gave up their time to

comment on their coding of COPD, and for their help with this study. Jeremy van Vlymen prepared the age-sex profile graphs and made helpful comments on the manuscript. The paper was presented at the Annual Scientific Meeting of the Primary Health Care Specialist Group of the British Computer Society; helpful comments and feedback given at the meeting have been used to improve the paper.³³

REFERENCES

- 1 Chung F, Barnes N, Allen M *et al*. Assessing the burden of respiratory disease in the UK. *Respiratory Medicine* 2002; 96:963–75.
- 2 Calverley PM and Walker P. Chronic obstructive pulmonary disease. *Lancet* 2003;362:1053–61.
- 3 Soriano B, Maier WC, Egger P *et al*. Recent trends in physician-diagnosed COPD in women and men in the UK. *Thorax* 2000;55:789–94.
- 4 Celli BR, Cote CG, Marin JM *et al*. The body-mass index, airflow obstruction, dyspnea, and exercise capacity index in chronic obstructive pulmonary disease. *New England Journal of Medicine* 2004;350:1005–12.
- 5 Enright PL and Kaminsky DA. Strategies for screening for chronic obstructive pulmonary disease. *Respiratory Care* 2003;48:1194–201.
- 6 Wells S and de Lusignan S. Does screening for loss of lung function help smokers give up? *British Journal of Nursing* 2003;12:744–50.
- 7 NHS Confederation. *The New General Medical Services Contract*. www.nhsconfed.org/gms/gms_contract_documentation.asp.
- 8 National Institute for Clinical Excellence. *Chronic Obstructive Pulmonary Disease. Clinical Guideline 12*. London: National Institute for Clinical Excellence, February 2004. www.nice.org.uk/Docref.asp?id=106421
- 9 NHS National Programme for Information Technology. *QMAS (Quality Management and Outcomes) Bulletin*. www.dh.gov.uk/assetRoot/04/08/42/59/04084259.pdf
- 10 Pringle M, Ward P and Chilvers C. Assessment of the completeness and accuracy of computer medical records in four practices committed to recording data on computer. *British Journal of General Practice* 1995;45:537–41.
- 11 Williams JG. Measuring the completeness and currency of codified clinical information. *Methods of Information in Medicine* 2003;42:482–8.
- 12 Hassey A, Gerrett D and Wilson A. A survey of validity and utility of electronic patient records in a general practice. *British Medical Journal* 2001;322:1401–5.
- 13 Thiru K, Hassey A and Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. *British Medical Journal* 2003;326:1070–7.
- 14 Brown PJ and Warmington V. Info-tsunami: surviving the storm with data quality probes. *Informatics in Primary Care* 2003;11:229–33; commentary 234–7.
- 15 Teasdale S and Bainbridge M. Improving information management in family practice: testing an adult learning model. *Proceedings of the AMIA Annual Fall Symposium* 1997;19:687–92.
- 16 Teasdale S. Supporting change management with PRIMIS. *Informatics in Primary Care* 2002;10:221–6.

- 17 Horsfield P and Teasdale S. Generating information from electronic patient records in general practice: a description of clinical care and gender inequalities in coronary heart disease using data from over 2 million patient records. *Informatics in Primary Care* 2003;11:137–44.
- 18 de Lusignan S, Hague N, Brown A and Majeed A. An educational intervention to improve data recording in the management of ischaemic heart disease in primary care. *Journal of Public Health* 2004;26(1):34–7.
- 19 PRIMIS (Primary Care Information Services): www.primis.nhs.uk
- 20 National Statistics. *Key Health Data from General Practice*. www.statistics.gov.uk/CCI/nscl.asp?ID=6443
- 21 NHS Information Authority. Clinical Terminology Service; www.nhsia.nhs.uk/terms/pages/readcodes_intro.asp
- 22 EMIS (Egton Medical Information System): www.emis-online.com
- 23 Clinical Information Consultancy. *CLUE Character Based Hierarchy Engine*. www.clininfo.co.uk/main/clue.htm
- 24 National Statistics. *Census 2001: England population*. www.statistics.gov.uk/census2001/pyramids/pages/64.asp
- 25 Commission for Health Improvement. *NHS Performance Indicators Direct Standardisation Methodology*. www.chi.nhs.uk/Ratings/Downloads/direct_standardn_meth.pdf
- 26 Halbert RJ, Isonaka S, George D and Iqbal A. Interpreting COPD prevalence estimates. What is the true burden of disease? *Chest* 2003;123:1684–92.
- 27 Liu H and Friedman C. CliniViewer: a tool for viewing electronic medical records based on natural language processing and XML. *Medinfo 2004 Proceedings* 2004: 639–43.
- 28 Martins SB, Shahar Y, Galperin M *et al.* Evaluation of KNAVE-II: a tool for intelligent query and exploration of patient data. *Medinfo 2004 Proceedings* 2004:648–52.
- 29 Ramnarayan P, Tomlinson A, Kulkarni G, Rao A and Britto J. A novel diagnostic aid (ISABEL): development and preliminary evaluation of clinical performance. *Medinfo 2004 Proceedings* 2004:1091–5.
- 30 Zielinski J and Bednarek M. Know the age of your lung study group: early detection of COPD in a high-risk population using spirometric screening. *Chest* 2001;119: 731–6.
- 31 de Lusignan S and Teasdale S. The features of an effective primary care data quality programme. In: Bryant J (ed). *Current Perspectives in Healthcare Computing, 2004: proceedings of HC 2004*. Swindon: British Computer Society, 2004:95–102.
- 32 Meyer PA, Mannino DM, Redd SC and Olson DR. Characteristics of adults dying with COPD. *Chest* 2002; 122:2003–8.
- 33 Primary Health Care Specialist Group of the British Computer Society: www.phcsg.org

CONFLICTS OF INTERESTS

None.

ADDRESS FOR CORRESPONDENCE

Simon de Lusignan
Senior Lecturer, Primary Care Informatics
Department of Community Health Sciences
Hunter Wing
St George's Hospital Medical School
London SW17 0RE
UK
Tel: +44 (0)20 8725 5661
Fax: +44 (0)20 8767 7697
Email: slusigna@sghms.ac.uk

Accepted June 2004 (conference paper)