Non-invasive methods for measuring data quality in general practice

Barry Gribben, ***Director RNZCGP Research Unit;*** **Gregor Coster,** ***Professor of General Practice, Department of General Practice and Primary Health Care, University of Auckland;*** **Mike Pringle,** ***Professor of General Practice, Department of General Practice, University of Nottingham, UK;*** **Jonathan Simon,** ***Medical Director, FirstHealth, Auckland.***

Abstract

Aim. To develop non-invasive methods of measuring the quality of data recorded in general practice.

Methods. Laboratory and pharmaceutical claims data from fourteen practices (44 doctors) from the FirstHealth network of general practices were examined to determine the extent to which valid minimum bounds on expected rates of diagnosis coding could be established. These were compared with recorded rates in patient notes to measure completeness of diagnosis recording. Data completeness was measured for demographic data and a marker for the accuracy of gender coding was developed from diagnosis data.

Results. Minimum rates of diagnosis could be established for asthma, diabetes (NIDDM and IDDM), ischaemic heart disease, hypothyroidism, bipolar affective disorder and Parkinson's disease. Minimum bounds for the number of patients requiring monitoring of warfarin and digoxin levels were also established. These expected minimum rates were combined with measures of completeness of age, gender, ethnicity and smoking data, and a gender coding accuracy measure, to produce a set of fourteen data quality indicators. Pass/fail thresholds on each indicator were set and each of the fourteen practices was scored on the number of passes they achieved. The scores ranged from three to nine out of fourteen passses.

Conclusions. Non-invasive data quality measures may be useful in providing feedback to general practitioners as part of a data quality improvement cycle. The sensitivity of this method will decline as data quality improves.

NZ Med J 2001; 114: 30-2

Over the past decade, the amount of information collected in general practice (GP) has increased significantly. Across New Zealand, at least 85% of GPs now use computers.[1] Most often, computers are used for accounts and the maintenance of age/sex registers,[2] but increasingly, doctors are using practice management systems for clinical notes, prescribing and laboratory test ordering, as well as practice administration, and many doctors routinely code diagnoses using Read codes. Since this information is used for an increasing range of purposes, from the claiming of subsidies to disease surveillance, the quality of the data recorded by GPs will come under increasing scrutiny. Internationally, there is widespread interest in the introduction of information technology in primary care and the NHS Executive in the UK is funding a major programme to support the collection and analysis of primary care data, led by one of the authors (MP).

In November 1998, the RNZCGP Research Unit in the Department of General Practice and Primary Health Care at the University of Auckland was contracted by FirstHealth to develop methods for measuring the quality of the data recorded by general practice. FirstHealth collects data from all practices. A Clinical Policy Committee determines the data to be collected. The data are collected automatically using Structured Query Language (SQL) 'queries'. We were supplied with data from fourteen practices (representing 44 GPs) in the PrimeHealth group of practices, in the Western Bay of Plenty. All practices used computerised clinical records, twelve using 'MedTech' and two using 'GPDat' The data collected are shown in Table 1.

Methods

Each practice dataset was examined for completeness. In the case of demographic data, it was expected that age, gender, ethnicity and smoking status data would be recorded for every patient. To assess the completeness of diagnosis codes, it was necessary to develop a method of estimating a rate of known diagnoses for each practice, and comparing recorded rates with these estimates.

Data for the same practices were downloaded from Health Benefits Limited (HBL) for all laboratory tests and prescriptions for which claims were submitted to HBL for the year April 1997 to March 1998. These data were merged on practice codes to create a set of tables that could be consulted to produce monthly totals of tests and prescriptions submitted for claims, by individual test or prescription item. Each diagnosis code in the practice dataset was considered to see if a lower limit on the number of patients with that diagnosis could be derived from the laboratory or prescribing data available. For example, to determine a minimum number of diagnoses for osteoarthritis that a practice should have recorded would require finding a pharmaceutical used exclusively for osteoarthritis, and calculating the average number of prescriptions per month issued by a practice. No such pharmaceutical exists, as the commonest medications for osteoarthritis, the non-steroidal anti-inflammatories, are used in many other conditions. However, for many conditions there is a very tight association between diagnosis and prescription; for instance, insulin is only prescribed for diabetes, and lithium is only prescribed for mania.

The accuracy of estimates for the number of patients with a given diagnosis relies upon the frequency with which a prescription is given for a given condition, and the number of possible agents that may be prescribed. For example, non-insulin dependent diabetes (NIDDM) may be managed with dietary modification only. Cases managed in this way will generate no prescriptions, and estimates of cases based on prescribing analysis will be low.

Table 1. Description of collected data.

Data items collected (numbers)	Diagnoses / Treatments requiring blood tests
Total list size Ages (in bands) Weight/Height/BMI recorded Ethnicity (Maori/Other/Absent) Smoking status (Present/Absent) in age bands Patients >65 with flu shot Patients >65 with diagnosis "flu" (H27) Asthmatics with Peak Flow recorded BP recorded in last 5 years	Group 1 – specific Read code: Warfarin (INR), Hypothyroidism (T4), Epilepsy (anti-convulsant), Polymyalgia Rheumatica (ESR), Lithium Therapy (Li) Diabetes (HbA1c), Hyperlipidemia, Blood disorder, Gout Group 2 – diagnosis + blood test recorded as necessary: Hypertension, Chronic Renal Failure, Chronic Hepatitis

Diagnoses searched for by Read code (numbers)
Hypertension, Asthma, NIDDM, IDDM, Heart Failure, CORD, Hypothyroidism, Hepatitis, Depression (reactive, endogeonous, with psychosis), schizophrenia, HIV, Chronic Renal Failure, Leukaemia, Ca colon, Ca lung, Ca breast, Ca cervix, Ca stomach, Ca prostate, Ca bladder, Ca brain, melanoma, BCC, SCC, Epilepsy, Parkinson's, Alzheimer's, Multiple Sclerosis, Motor Neurone Disease, Anxiety states, eating disorder, acute lumbar strain, RSI, methadone treatment, B12 deficiency, Gonorrhoea, Gastroenteritis, Strep pharyngitis, Chlamydia, Cataract, Glaucoma, Osteoarthritis, Rheumatoid arthritis, Meningitis (meningococcal, haemophilus), Measles, Rubella, Influenza

All data elements were tabulated by gender.

The estimate of the number of cases is further complicated by multiple agents. Some NIDDMs are managed with sulphonylureas, some with metformin (the only biguanide available in New Zealand), and some with insulin. Because of the possibility of being prescribed two agents, however, an estimate of the number of cases must use the most frequently prescribed agent or the sum of mutually exclusive agents – no one is prescribed two sulphonylureas. The estimate must then be interpreted as being a minimum, bound on the number of cases that are receiving prescriptions.

Similar analyses were applied to each of the conditions in the data downloaded from the practices. Estimates were constructed to be as valid as possible, and were therefore usually very conservative. In the best possible case, the estimates would logically significantly underestimate the actual numbers of cases of a given condition. The estimated rates of recorded diagnoses were then compared with the actual recorded numbers to see if these estimates could be used as useful measures of data quality at this stage in the development of diagnosis recording in general practice.

Results

After examining possible methods of data accuracy determination for each of the recorded data elements, we constructed the following set of data quality indicators. The indicators fell into three groups: measures of demographic data completeness (including smoking status recording), measures of diagnosis coding adequacy, and a gender by diagnosis composite index. The demographic data quality indicators are described in the top section of Table 2, and measure the completeness of gender, date of birth, ethnicity and smoking status recording. From the analysis of diagnoses, we selected a set of diagnosis data quality indicators. They have been chosen using two criteria. The first is that they are valid, that is, their construction leaves little doubt that they are measuring data quality. Typically, they do not rely upon any unreasonable assumptions and involve tight binding between diagnoses and pharmaceuticals or laboratory tests. The second is that they are sensitive. There is little point using an indicator that every practice passes already, as there is no further information to be gained in subsequent data collection. Neither is there any benefit in including measures based on diagnoses with low prevalence (for example, glaucoma, as evidenced by low prescribing rates) since the scoring of the indicator is problematic when there are no cases. However, indicators that no practice presently passes may be useful if we are certain the indicator is valid.

The set of indicators has been constructed with a view to including measures of the accuracy of coding of the most prevalent chronic conditions when possible. Fortunately, reasonably strong indicators are available for diabetes and ischaemic heart disease (IHD).

Table 2. Data Quality Indicators.

Name of data quality indicator	Description
Demographic data indicator	
Gender	% records with gender entered
DOB (date of birth)	% records with date of birth entered
Ethnicity	% records with ethnicity entered
Smoking	% records (aged 16 or over) with smoking status entered
Diagnosis type indicators	***Criterion for 'pass' score***
Asthma	exceed maximum monthly prescriptions for greater of (average number of bronchodilator scripts per month) and (average number of inhaled steroid scripts per month)
NIDDM	exceed (average number oral hypoglycemic scripts per month)
IDDM	exceed (average number insulin scripts per month)
IHD	exceed (average number nitrate scripts per month)
Hypothyroidism	exceed (average number thyroxine scripts per month)
Bipolar	exceed (average number lithium scripts per month)
Parkinsons	exceed (average number levodopa scripts per month)
Warfarin	exceed (average number of INR tests per month) / 8
Digoxin	exceed (annual number of digoxin levels) / 2
Gender coding indicator	
Sexdiag	No gender errors for ca cervix, ca ovary or ca prostate.

Because of the weak linkages between the three datasets (practice data, laboratory test claims data and prescription claims), some of these indicators are only useful as measures of relatively serious under-recording of diagnoses. We have had to construct indicators that use conservative criteria to ensure high validity. It will be possible to significantly improve estimates of diagnosis rates if and when script/patient or laboratory test/patient linking becomes available. The proposed set is shown in the middle section of Table 2. The last entry in Table 2 describes a composite 'gender by diagnosis' measure. All practices have passed this index in our analysis, however, it has been included because it is a

different type of indicator to the previous two. The indicator is passed if there are no cases of male gender recorded for diagnosis of cancer of the cervix or cancer of the ovary and no cases of female gender for cases of cancer of the prostate. Breast cancer is not included because of the possibility of male gender (1% of cases of breast cancer).

To simplify reporting of results, we have chosen to use a pass/fail scoring system for tracking the quality of both demographic and diagnosis data over time. This requires the setting of thresholds for the demographic data items. We have chosen 100% as the threshold in the first instance, but this might be relaxed for ethnicity coding (to say 95%).

A practices data quality score is then simply the sum of their passes, to give a score out of 14. This is scaled to 100 for reporting purposes. The score for a network of practices is then the average value of the practices' scores, weighted by practice size.

Table 3 describes the performance of each practice on the 14 indicators, and gives a percentage score for each practice, ranging from 3/14 to 9/14.

Table 3. Data quality summary table.

Practice	Gender	DOB	Ethnicity	Smoking	Asthma	NIDDM	IDDM	IHD	Hypothyroidism	Bipolar	Parkinson's disease	Warfarin	Digoxin	Sexdiag	Total %
1	Y	Y	N	N	Y	N	N	Y	Y	Y	Y	N	Y	Y	64%
2	Y	Y	N	N	Y	N	Y	Y	N	Y	N	N	Y	Y	57%
3	Y	Y	N	N	Y	N	N	Y	N	Y	Y	Y	N	Y	57%
4	Y	Y	N	N	N	N	N	N	Y	N	N	N	N	Y	28%
5	Y	Y	N	N	Y	N	Y	Y	Y	N	Y	N	Y	Y	64%
6	Y	Y	N	N	Y	N	N	Y	N	Y	N	N	Y	Y	50%
7	Y	Y	N	N	Y	N	N	Y	N	N	Y	Y	N	Y	50%
8	Y	Y	N	N	Y	N	N	Y	N	Y	Y	Y	N	Y	57%
9	Y	Y	N	N	N	N	N	N	N	N	N	N	Y	Y	28%
10	Y	Y	N	N	Y	N	N	N	N	Y	Y	N	N	Y	42%
11	Y	Y	N	N	Y	N	N	Y	N	N	Y	Y	Y	Y	57%
12	Y	Y	N	N	N	N	N	N	N	N	N	N	N	Y	21%
13	Y	Y	N	N	Y	N	N	Y	Y	Y	Y	Y	N	Y	64%
14	Y	Y	N	N	N	N	N	N	N	N	N	N	N	Y	21%

Y = yes; N = no.

Discussion

Although the estimates of numbers of diagnoses were very conservatively constructed, it was interesting to see how many practices did not record the expected number of diagnoses. This suggests that there is significant room for improvement in diagnosis recording in these practices, which is not at all surprising given the relative short time that diagnosis coding has been available. More importantly, the low scores obtained by most practices indicates that the above method of measuring data quality has some practical use in providing practices with feedback on their levels of diagnosis recording.

One possible method for determining data completeness would have been by a survey of a random selection of each doctor's notes. However, this would have been time consuming, very expensive, and not useful as a practical tool for ongoing monitoring of data quality. By using claims data, coupled with automatic downloads from practice computers, data quality feedback can be given with no requirement for any extra work by the practice team, other than running a programme.

We conclude that non-invasive data quality measures may be useful in providing feedback to GPs as part of a data quality improvement cycle. The sensitivity of this method will decline as data quality improves. However, the potential for constructing more sensitive measures of data quality will improve substantially when laboratory testing and prescribing can be allocated to individual patients.

Acknowledgements. We are grateful for the assistance of the following people in data collection and preparation: Karen Keberle, Clinical Informatics Analyst, First Health; Jane Millington, Manager, Quality and Performance, First Health; Sara Williams, Senior Business Analyst, First Health.

Correspondence. Barry Gribben, Director RNZCGP Research Unit, Department of General Practice and Primary Health Care, University of Auckland, PO Box 92019, Auckland. Fax: (09) 834 2970; Email: b.gribben@auckland.ac.nz

1. Thakurdas P, Coster G, Gurr E, Arroll B. New Zealand general practice computerisation; attitudes and reported behaviour. NZ Med J 1996; 109: 419-22.
2. Gribben B, Bonita R, Broad J et al. Geographical variations in the organisation of general practice. NZ Med J 1995; 108: 361-3.

Almost a decade after it was found that folic acid could prevent the development of neural tube defects such as spina bifida, three researchers have received the Kennedy Foundation international award for scientific achievement for the discovery.

The award was made to Professor Nicholas Wald of Barts Hospital, Professor Richard Smithells, emeritus professor of paediatrics at Leeds University, and Dr Andrew Czeizel of the Budapest National Institute of Hygiene for key studies that led to the eventual identification of the beneficial effects of folic acid.

The award came four years after the United States acted on the research and began to fortify flour with folic acid. The UK Department of Health last month launched a three month consultation before deciding whether to require manufacturers to add folic acid to flour in the same way.

For Professor Wald, however, the evidence for fortification has been overwhelming for some time. "There is now conclusive evidence that if women consume more folic acid there is a substantial reduction in the changes of having a baby with spina bifida," he said. "Women need to take it before they become pregnant, so putting it in flour is the most effective way of preventing this disorder in the population, and the Americans have been doing it since 1997."

Roger Dobson, BMJ 2000; 321: 400.

"Basic research" said Werner von Braun "is what I'm doing when I don't know what I'm doing." Clearly this was an alarming statement, coming as it did from someone whose main research interest was lighting matches under huge quantities of explosive fuel. I suppose you could tell he was good at it, because he was still around to propose the next development, and the next… But, perhaps his remark had an effect. Certainly, 50 years or so later, a paraphrase would read rather differently – perhaps something along the lines of "Basic research is something that I can't get funding for unless I already know the answer and have already done half the experiments".

John A Lee, The Lancet Oncology May 2000; 24.