

Collecting morbidity data in general practice: the Somerset morbidity project

Nicky Pearson, Jim O'Brien, Huw Thomas, Paul Ewings, Lesley Gallier, Alan Bussey

Abstract

Objective—To collect a valid, complete, continuous, and representative database of morbidity presenting to primary care and to use the data to help commission services on the basis of local need and effectiveness.

Setting—Computerised general practices in Somerset.

Methods—Participating general practices were selected to be representative of the district health authority population for general practice and population characteristics. All conditions presented at face to face consultations were assigned a Read code and episode type and the data were regularly validated. Data were sent by modem from the practices via a third party to the health authority each week.

Main outcome measures—Proportion of consultations coded and accuracy of coding.

Results—11 practices agreed to participate. Validations for completeness during April 1994 to March 1995 revealed that 96.4% of the records were coded; 94% of the 1090 records validated had appropriate episode types and 87% appropriate Read codes. The results have been used to help formulate the health authority's purchasing plans and have enabled a change in the local contracts for surgery for glue ear.

Conclusions—The project has shown the feasibility of establishing a network of practices recording and reporting the morbidity seen in primary care. Early indications are that the data can be useful in evidence based purchasing.

Introduction

An important responsibility of health authorities and general practitioners is to assess the health needs of their local populations and to commission services to meet these needs. This task is hampered by a lack of routinely available, up to date, and accurate information on incidence and prevalence of disease. Existing national sources of morbidity data, of which the most comprehensive are the national morbidity studies,¹⁻⁴ are not representative of smaller localities or health authority populations, may be out of date as the studies take place only decennially, and may be biased by the fact that only general practitioners who have volunteered contribute.

General practice records have already been recognised as a potentially rich source of morbidity data.⁵ Improved uptake of computerisation in primary care and technological advances in electronic data transfer make it possible to use general practice computer systems for assessing local health needs. However, attempts at collecting up to date and reliable information in other areas of Britain have had only limited success because of problems with completeness and accuracy of the data, Read codes, computer software, and information technology, particularly arising as a

result of the wide variety of commercially available general practice computer systems.⁶⁻⁸

The Somerset morbidity project aims to collect valid, complete, continuous, and representative data on disease presenting to primary care. The resulting database is readily obtainable; useful to purchasers, providers, and general practitioners; and could be reproduced in other health districts. We report on the data collection methods and analyse the data collected on three conditions in greater detail: glue ear, where the health authority has used the data to commission surgical services; asthma, which has generated interest among the general practitioner user group; and hypertension, where the local prevalence seemed to be significantly different from the prevalence found nationally.

Methods

To avoid the potential bias of using volunteer general practitioners we selected practices with a suitable computer system that were representative of the district health authority population for general practitioner, practice, and population characteristics. In 1993, when recruitment began, 97% of practices in Somerset were computerised and 64% were using a computer system suitable for our study (using Read codes and software compatible with our data extraction software). We recruited from these practices in four stages using a computerised model established for the purpose. The model initially generated all possible samples of practices that met the representativeness criteria within predefined acceptable limits. Those practices appearing most frequently in the samples were approached and recruited. As practices were recruited the sampling process was repeated.

We chose this method as it was systematic, repeatable by others, and maximised the chances of selecting a representative sample if a substantial number of practices declined to participate. However, only one practice we approached declined. We treated the resulting sample as a randomly selected sample of clusters of unequal size for the purpose of calculating confidence intervals for population characteristics using the method described by Kish.⁹

Each practice is paid a small fee, currently 40 pence a year for each patient registered with the practice, which covers the cost of generating the weekly report.

RECORDING AND EXTRACTING DATA

Practices record Read and episode type codes for each face to face contact between patients and practice medical and nursing staff. The episode types refer to whether the patient is consulting with the illness for the first time (F), attending for a new episode of a previously diagnosed illness (N), or for a follow up consultation (O).

The extraction software searches for each diagnosis for each patient during the time specified. It also prioritises episode types for each diagnosis: F over N, over O.

Somerset Health Authority, Taunton TA2 7PQ

Nicky Pearson, senior registrar public health medicine

Jim O'Brien, consultant public health medicine

Paul Ewings, district statistician

Lesley Gallier, project coordinator

Alan Bussey, consultant public health medicine

King Edward Road Surgery, Minehead Huw Thomas, general practitioner

Correspondence to: Dr Pearson.

BMJ 1996;312:1517-20

Table 1—Characteristics of practices in Somerset morbidity project and whole district health authority

	% Of sample practices (n = 64 775)	% Of district health authority (n = 415 771)
District council:		
Mendip	5.2	10.6
Sedgemoor	28.4	24.5
Taunton Deane	21	23.2
West Somerset	10.1	7.6
South Somerset	35.3	34.2
Rural practice	35.8	32.4
Urban practice	64.2	67.6
No of partners:		
1-2	21.1	13.6
3-4	34.2	34.1
≥ 5	44.7	52.3
≥ 1 Female partner	73.3	66.6
Age/sex structure:		
Males:		
0-4	6.1	6.1
5-14	12.9	13
15-44	43.2	40.3
45-64	22.1	24
65-74	9.2	10
≥ 75	6.4	6.6
Females:		
0-4	5.6	5.6
5-14	11.2	11.5
15-44	39.1	37.3
45-64	21.9	23.1
65-74	11.5	11.6
≥ 75	10.7	11

Each diagnosis is therefore represented only once for a patient for the time concerned. Data are routinely extracted at the end of each week, the end of each quarter, and the end of each year. These searches produce different databases that enable disease processes with different epidemiological patterns to be studied optimally.

The data are sent electronically by modem to an independent collection agency (the Royal College of General Practitioners Birmingham Research Unit) and then on to the database held at the health authority. No identifiable patient details leave the practice, and the research unit gives data from each practice a code, which is known only to the project coordinator. Analyses are presented coded by practice, and each practice knows only its own code.

DATA QUALITY

Several systems are in place to ensure optimum data quality. Firstly, all general practitioners and practice staff are given initial training and regular updates. Secondly, the project coordinator visits each practice on a random day every three months to validate the data. The Read codes entered for a random 25 patient sample are checked against the paper notes to ensure that they are compatible, appropriate, and given correct codes and episode types. Thirdly, the computer carries out edit checks for rare diagnoses and those recorded outside the usual age and sex parameters—for example, gynaecological disorders in men, very rare diagnoses where a coding error is more likely than the disease—for example, tetanus immunisation wrongly coded as a case of clinical tetanus. Incorrectly coded items noted during the validation visits and as a result of the edit checks are brought to the practice's attention and the computerised records amended. Data integrity checks are also carried out to verify the electronic interchange process, and a general practice user group meets regularly to enable data feedback and interpractice comparisons and to agree common use of Read codes where appropriate across the participating practices.

Results

We recruited a final sample of 11 practices with a combined population of about 65 000. The practices were representative of the health authority for geographical spread, urban-rural balance, practice size, presence of a female partner, and age and sex structure (table 1). Table 2 compares the morbidity for the Somerset practices with those practices which participated in the fourth national morbidity study.⁴

Of the 4685 records validated for completeness during April 1994 to March 1995, 4516 (96.4%) of the records were coded. During the same period 1035 (94%) of the 1090 records validated had appropriate episode types and 948 (87%) appropriate Read codes. Of the 142 cases where the Read code was queried, most were not incorrectly coded but were either not coded consistently throughout the course of the illness or not coded in as much detail as was possible from the available information.

The distribution of illness across the disease categories in Somerset was similar to the national distribution (table 2). However, some differences with national data were apparent.

Table 2—Rate of disease per 100 000 patients (95% confidence interval) in Somerset morbidity project April 1994-March 1995 and national survey 1991-2⁴

ICD-9 category	Males		Females	
	Somerset	National	Somerset	National
Infectious/parasitic diseases	11 194 (9 368 to 13 023)	11 370	13 924 (12 035 to 15 813)	16 500
Neoplasms	2 473 (2 250 to 2 697)	1 900	3 695 (3 154 to 4 237)	2 860
Endocrine/nutrition/metabolic diseases	3 346 (2 903 to 3 790)	3 050	4 326 (3 702 to 4 951)	4 460
Blood diseases	617 (420 to 814)	490	1 609 (1 333 to 1 885)	1 430
Mental disorders	5 440 (3 761 to 7 120)	5 030	8 442 (6 797 to 10 087)	9 440
Nervous system/sense organ diseases	16 172 (14 725 to 17 622)	15 380	19 305 (17 825 to 20 785)	19 190
Circulatory system diseases	10 359 (8 788 to 11 931)	8 390	12 902 (11 082 to 14 722)	10 200
Respiratory system diseases	24 708 (22 420 to 27 000)	27 220	29 563 (26 580 to 32 458)	34 040
Digestive system diseases	6 806 (5 956 to 7 657)	7 570	8 229 (7 114 to 9 344)	9 710
Genitourinary system diseases	3 657 (3 328 to 3 987)	3 590	15 654 (14 274 to 17 035)	18 760
Pregnancy/childbirth/puerperium	0	0	1 467 (985 to 1 948)	2 110
Skin/subcutaneous tissue diseases	12 653 (11 729 to 13 579)	12 710	15 817 (14 837 to 16 798)	16 310
Musculoskeletal/connective tissues	13 736 (12 483 to 14 990)	12 950	18 120 (16 452 to 19 789)	17 380
Congenital anomalies	472 (395 to 549)	540	444 (336 to 553)	520
Perinatal conditions	59 (24 to 93)	120	86 (62 to 110)	130
Symptoms/signs/ill defined conditions*	25 719 (16 910 to 34 534)	12 180	38 373 (26 715 to 50 032)	17 910
Injury and poisoning	13 810 (11 617 to 16 005)	13 560	12 605 (10 580 to 14 631)	14 230

*The high Somerset rate for the symptoms, signs, and ill defined conditions is artefactual and is due to default coding in the GP system software which is only picked up by the Somerset extraction software.

Table 3—Incidence (rate/100 000 children) of glue ear in Somerset morbidity project April 1994-March 1995

Age (years)	Incidence (95% confidence interval)
< 1	1122
1-4	2730
5-14	1455
0-14	1777 (1337 to 2216)

Read codes for glue ear: F511, F512, F513, F521, F522, F523, F59, F622, F623, F633, F62Z, F632, F68.

GLUE EAR, ASTHMA, AND HYPERTENSION

Table 3 shows the incidence of glue ear in 0-14 year olds during April 1994 to March 1995 (using all the Read codes which participating general practitioners are currently selecting for glue ear). The incidence of new cases was 17.7 per 1000 0-14 year olds.

Table 4 shows the age and sex breakdown of patients with asthma from Somerset compared with national data. The prevalence of asthma in Somerset is higher than the prevalence nationally for both males and females, with children aged 5-14 years and young adults from 15-24 years having the highest prevalence. Nationally, however, the 0-4 age group has the highest prevalence.⁴ There was wide variation in prevalence of asthma between participating practices (table 5),

practice F having over twice the number of asthmatic patients as practice A after standardising for age and sex.

We found a much higher prevalence of hypertension in Somerset than in the national study (table 4). The variation between practices of the prevalence of hypertension was also large.

Discussion

VALIDITY OF DATA COLLECTED IN GENERAL PRACTICE

There are obvious difficulties in collecting data in general practice, particularly that in the short time available for the average consultation it is not feasible to ask general practitioners to adhere to exact, previously defined diagnostic criteria for all illnesses they see. Thus, some of the interpractice variation is undoubtedly not true variation in morbidity but due to general practitioners' different diagnostic practices. However, much prescribing and many referrals to secondary care are triggered by these working diagnoses and therefore the interpractice variation has practical relevance.

Our system is a practical non-disruptive method of gathering data across the range of disease in the sector of the health service where most people consult. Useful information on the incidence of common disorders has previously been gathered from general practitioners' diagnoses,¹⁰ and the approach of using general

Table 4—Prevalence (rate/100 000 patients) of asthma and hypertensive disease in Somerset morbidity project April 1994 to March 1995 and national survey 1991-2⁴

Age group	Asthma*				Hypertensive disease†			
	Somerset		National		Somerset		National	
	Males	Females	Males	Females	Males	Females	Males	Females
0-4	6 152	3848	9940	7220	0	0	0	0
5-14	10 011	9170	8600	6450	0	0	0	0
15-24	4 784	8005	3960	4590	43	144	90	130
25-44	2 627	4190	2580	3340	1 009	1 025	1 030	1 010
45-64	2 507	3346	2600	3420	8 184	9 841	7 670	8 670
65-74	4 826	4161	3720	4120	19 039	22 725	14 880	17 820
≥75	5 256	3562	3260	3180	18 202	22 719	9 020	15 230
All ages (95% confidence interval)	4475 (4017 to 4934)	4954 (4297 to 5612)	4290	4220	5067 (3859 to 6276)	7467 (5915 to 9020)	3590	4820

*Read5 code : H33, Read4 code : H34.

†Read5 code: G2, Read4 code: G3.

Table 5—Variation in prevalence (rate/100 000 patients) of disease among practices in Somerset morbidity project

Practice	Asthma								Hypertensive disease (95% confidence interval)	
	Males				Females				Males	Females
	First	New	Review	Total (95% confidence interval)	First	New	Review	Total (95% confidence interval)		
A	683	299	1559	2541	665	570	1707	2942	4623	7440
B	887	1323	1531	3740	1128	1165	1566	3859	5965	7349
C	1118	844	3092	5053	1105	800	3267	5171	5201	7045
D	625	414	3774	4813	691	381	2837	3908	8136	12209
E	1977	2232	2394	6603	1650	2883	3011	7544	5944	8386
F	1751	1688	1330	4769	1119	2031	1115	4265	3966	5021
G	932	1313	2252	4496	913	1577	2372	4862	3773	6565
H	765	404	2413	3582	1070	679	2468	4217	4230	6735
J	889	1413	2042	4343	782	1274	3306	5362	7169	9652
K	476	682	2428	3587	1046	935	2894	4875	2322	4343
L	1289	1693	1798	4780	1959	2084	2185	6227	4777	6929
Total	1052	1141	2282	4475 (4017 to 4934)	1182	1360	2412	4954 (4297 to 5612)	5067 (3859 to 6276)	7467 (5915 to 9020)
National data ⁴	3000*		1290	4290	2950*		1270	4220	3590	4820

*First and new cases.

practitioners' working diagnoses has been defended in the context of the weekly returns service.¹¹ We believe we have improved on previous methods of collecting data in primary care by concurrently removing the potential bias of volunteer practices, incorporating local representativeness, recording across the whole range of disease, using a range of validation procedures, and making data available promptly.

Our data were similar to those reported in the fourth national morbidity study, which is reassuring given that other authors have compared their general practice data with data from other sources to assess external consistency.¹²

IMPORTANCE OF LOCAL DATA

Locally valid and representative data on morbidity are vital to purchasers of services if evidence based commissioning is to be realised. Glue ear is one example of how these data have been used. The recent publication of surgical treatment rates for glue ear across the United Kingdom showed that surgical intervention rates varied from four per 1000 children in the lowest region to nine operations per 1000 in the highest region. National effectiveness studies have shown that about half of episodes of glue ear will resolve spontaneously in three months and three quarters will resolve by six months with little recurrence.¹³ Therefore, an appropriate surgical intervention rate for glue ear should be roughly 25% of new episodes. The incidence of new episodes of glue ear in Somerset was 17.7 per 1000 0-14 year olds, suggesting that the surgical intervention rate should be about 4.4 per 1000. The actual surgical intervention rate was 6 per 1000 children. This indicates that some children in Somerset may be receiving inappropriate surgery. Clearly the figure of 4.4 per 1000 is not exact and there could be variation around it, but the combination of widely accepted effectiveness information with local valid data on the disease enabled a modification to the contract to purchase reduced levels of surgery and increased watchful waiting.

Like many common surgical conditions, services for patients with glue ear have traditionally been supply led rather than need driven. When local providers have been challenged to explain high intervention rates they have often asserted that they result from high local incidences of the condition rather than a high inclination for intervention. Until now such arguments could be countered only by conducting expensive and time consuming community based surveys. Our data enabled more informed debate over the contract for glue ear services.

The general practitioner user group also believe they will be able to use the data on asthma, which showed wide interpractice variation. They intend using the relative proportions of the incidence of individuals developing asthma for the first time, the incidence of acute asthma attacks by those known to have asthma, and attendances for ongoing management as a way of gauging the effectiveness of their management of chronic asthma. Practices whose asthma management clinics are functioning optimally would have relatively fewer attendances for acute attacks than ongoing management.

CONCLUSIONS

The availability of up to date, representative, and accurate data from primary care on all disease in Som-

Key messages

- General practice computerised records are a rich source of data on morbidity
- This study shows that general practitioners will contribute to a morbidity database if given financial support and confidentiality is safeguarded
- Validation procedures ensure that the data are high quality
- Information from the database can be used to purchase health care based on the population's needs as well as to monitor and improve the health of the population

Somerset has allowed general practitioners, public health doctors, and commissioners to begin to work together to pursue a goal of primary care led purchasing. Databases such as ours offer an opportunity to use peer reviewed evidence of effectiveness in a local context, which we hope to pursue further. They are a useful addition to available information about the health of our local population and may enable us to monitor the effect of evidence based purchasing on the diseases seen in primary care.

We thank all participating general practitioners and staff in the Somerset morbidity project, Dr Douglas Flemming and his staff at the Weekly Returns Service for advice and support, and Dr Anna McCormick for software for computerised edit checks.

Funding: South Western Regional Health Authority Research and Development Directorate (grant No 275), Department of Health Information Management Group, and Somerset Health Authority.

Conflict of interest: None.

- 1 Logan WPD, Cushion AA. *Morbidity statistics from general practice*. Vol 1. *General studies on medical and population subjects*. London: HMSO, 1958: (MB5, No 14.)
- 2 Royal College of General Practitioners, Office of Population Censuses and Surveys, Department of Health and Social Security. *Morbidity statistics from general practice. Second national study 1970-71. Studies in medical and population subjects*. London: HMSO, 1974. (MB5, No 26.)
- 3 Royal College of General Practitioners, Office of Population Censuses and Surveys, Department of Health and Social Security. *Morbidity statistics from general practice. Third national study 1981-1982*. London: HMSO, 1986. (MB5.)
- 4 McCormick A, Flemming D, Charlton J. *Morbidity statistics from general practice. Fourth national study 1991-1992*. London, HMSO, 1995. (MB5.)
- 5 Pringle M, Hobbs R. Large computer databases in general practice. *BMJ* 1991;312:741-2.
- 6 Smith N, Wilson A, Weeks T. Use of Read codes in development of a standard data set. *BMJ* 1995;311:313-5.
- 7 Wilson AE, Pollock C, Weeks T, Dowell A. Can general practice provide useful information? Evaluation of a primary health care information project in northern England. *J Epidemiol Community Health* 1995;49: 227-30.
- 8 Department of General Practice. *GPASS data evaluation project. Annual report. 1992-1993*. Aberdeen: University of Aberdeen, 1993.
- 9 Kish L. *Survey sampling*. New York: Wiley, 1965:187-90.
- 10 Fleming DM, Norbury CA, Crombie DL. *Annual and seasonal variation in the incidence of common disease*. London: Royal College of General Practitioners, 1991. (Occasional paper 53.)
- 11 Pereira Gray D. Preface. In: Fleming DM, Norbury CA, Crombie DL. *Annual and seasonal variation in the incidence of common diseases*. London: Royal College of General Practitioners. 1991 (Occasional paper 53.)
- 12 Boydell L, Grandider H, Rafferty C, McAteer C, Reilly P. General practice data retrieval: the Northern Ireland project. *J Epidemiol Community Health* 1995;49(suppl 1):22-5.
- 13 The treatment of persistent glue ear in children. *Effective Health Care* 1992;4:3.

(Accepted 25 April 1996)