

The capture of socioeconomic data in general practice

D M FLEMING

A McCORMICK

J CHARLTON

SUMMARY

Background. It is common practice to record the reasons why patients have an encounter with the practice, but the collection of socioeconomic data with which to link this morbidity data is less easy to achieve.

Aim. To describe the social enquiry used in the Fourth National General Practice-based morbidity study (1991–1992) and to consider its effectiveness for use in practice.

Method. Socioeconomic data were collected using a structured questionnaire administered by a trained interviewer. Data were provided by both consulting and non-consulting registered patients.

Results. The interview technique proved to be acceptable to patients, interviewers and general practitioners, simple to administer, and inexpensive to collect. Eighty-three per cent of the 502 000 people included in the study provided social and occupational data. Less than 1.5% of patients refused to be interviewed. Fifty-four of the 60 practices achieved the target level of 90% of registered patients being successfully interviewed.

Conclusion. A method of socioeconomic data collection based on that used in the 1991–1992 study would be of benefit for health care planning, allocation of resources, design of performance indicators and epidemiological research.

Keywords: consultation process; record content; socioeconomic factors; health status.

Introduction

KNOWLEDGE of the socioeconomic background and morbidity record of individual patients provides the basis of many aspects of health care planning and research in general practice. The population of every practice is different from every other. Consequently, the service required by patients registered with each practice differs. Practices need to know the relationship between morbidity and particular groups within their catchment population to plan appropriate care. Those practices with an interest in research will find that, if they can capture socioeconomic data about their patients and link this to the illnesses these patients suffer, they will be rewarded by data which they can analyse, thus enhancing interest in their work and being of benefit to patients.

In the UK, there is a system of universal patient registration with general practitioners. With the exception of medical emer-

gencies, the convention that access to secondary care is through referral by general practitioners is well accepted. The integration of primary and secondary care is evident following discharge from secondary care when relevant morbidity data are forwarded to practices by letter or electronic means. Thus, there exists the opportunity to consider the health problems of the community and its use of health services. Denominators by age and sex are available from the age–sex registers of practices. However, there is no universal way of linking different morbidities with particular groups of people in the community, such as the unemployed or those living alone, because socioeconomic information is not systematically recorded by practices.

Socioeconomic information has been collected and linked with morbidity data in four national surveys, each of a year's duration.^{1–6} In the 1955–1956 study, the only socioeconomic information collected, other than age and sex, was occupation.¹ In the 1971–1972 and 1981–1982 studies,^{2–5} morbidity was recorded in diagnostic indexes which were later linked by the Office of Population Censuses and Surveys (OPCS) to the age–sex register. Using identification from the register, a search was made at the OPCS for each patient's census form and the required socioeconomic data was extracted. This method, though largely successful, was time consuming and expensive. Thus, in the 1981–1982 study socioeconomic data collection was only attempted for patients registered with 25 of the 48 practices which collected morbidity data. Therefore, a different approach was used in the 1991–1992 study.⁶

The purpose of this report is to describe the methods employed in the study and to evaluate their usefulness by comparison with the methods employed for capturing socioeconomic data in earlier morbidity surveys.

Method

Sixty practices throughout England and Wales took part in the 1991–1992 study.⁶ There were 241 general practitioners working in these practices. Practices were recruited from volunteer practices, ensuring that each health region was represented. Each practice used computer software developed for the study by one of three computer companies. There were 502 493 patients registered with these practices at some time during the study year. Allowing for some patients being registered for only part of the year, these patients represented 468 042 person years at risk.

Between 1 September 1991 and 31 August 1992, morbidity data were recorded on computer for each face-to-face encounter between a patient on the age–sex register and a doctor or practice nurse. The data included the date, the diagnosis or diagnoses, the episode type(s), where the patient was seen, and whether and to which agency the patient was referred. Socioeconomic information was collected by trained interviewers about every person registered with the practice at any time during the study year, irrespective of whether they contacted the practice during the year. The method is described in detail in the study report.⁶ Morbidity data and socioeconomic data, which were also recorded on computer in the practice, were sent separately at the end of the study to the OPCS. At the OPCS, the two anonymized data sets were linked using unique patient identification numbers. Any data which were outside a specified range or were absent or

D M Fleming, PhD, FRCGP, Royal College of General Practitioners Birmingham Research Unit. A McCormick, FFPHM, medical statistician, and J Charlton, MSc, Hon Mem FPHM, statistician, Office of Population Censuses and Surveys, London.

Submitted: 14 March 1995; accepted: 8 September 1995.

© British Journal of General Practice, 1996, 46, 217–220.

incompatible with other data, such as an occupation with an inappropriate employment status, were identified by the OPCS and referred back to the practices for completion or correction.

Socioeconomic interview

Field workers were employed by the practices for the purpose of conducting the socioeconomic interview, coding the data where applicable and entering the data on computer. This method was chosen in preference to self-completed questionnaires because it was anticipated that these would require an unacceptable level of follow up. These workers received special training by OPCS staff in interview techniques and in coding occupation and country of birth (coding manuals were provided by the OPCS). They were also issued with a manual which stated the order of the questions to be asked and the exact wording to be used, to ensure standardization over the study population. Practices were also issued with a video film for the benefit of field workers who were unable to attend the training sessions.

An attempt was made to interview every patient on the practice age-sex register at any time during the study year. When adult patients were interviewed, they were asked, in addition to questions relating to themselves, questions about other members of their family who were also registered with the practice; these were called 'proxy' interviews. Children aged under 16 years were not interviewed, the information being obtained from their parents.

The majority of patients were interviewed on surgery premises. In the early months of the study period, patients in the waiting room were invited to an interview with the field worker while they were waiting. In later months, a planned programme of interviews was arranged with patients not yet interviewed or for whom information had not been obtained by proxy, in the surgery, in the patient's home or by telephone. The interviews were preceded by the distribution of letters explaining the purpose of the study and giving patients the opportunity to refuse to answer all or any of the questions. Sensitive questions were handled by using a card listing possible answers, allowing the interviewee to point to that which most clearly applied. Wherever possible, the list of alternatives was in the same format as printed on the 1991 census form.

Interviewers recorded the answers on an interview record form and later transferred the replies to a computer screen, the format of which resembled the form. Questions directed towards every patient included date of birth, sex, marital status, tenure of household, ethnic group, country of birth, and whether the patient or someone on his or her behalf (proxy) supplied the information. Information was not updated during the course of the study. If patients visited the surgery after a proxy interview

with them had taken place, any questions with missing answers were completed.

Patients aged 16 years and over were also asked whether they were the sole adult in the household and whether the household included children aged under 5 years, or between 5 and 15 years. They were asked whether they had smoked in the week before the interview. They were also asked about their economic position in the previous week and a year prior to interview (whether working full- or part-time, whether they were seeking work, were permanently sick and so on), their occupation and their employment status (e.g. manager or self-employed). The occupation and employment status were used by the OPCS to derive the patient's social class.

For patients aged under 16 years, proxy interviews included questions about household composition and parental occupation. Similar questions were asked about the partners of married or cohabiting women.

With few exceptions, these questions and the grouped categories for analysis were similar to or compatible with those used in the 1991 census. The replies to each question included an option for 'refused' or 'not known'.

Results

The main results of the 1991-1992 study have been published elsewhere.⁶ This section is restricted to those results relating to the capture of socioeconomic data.

Socioeconomic data were collected for 417 996 people, 83.2% of the 502 493 patients registered at some time during the survey year. In the third study (1981-1982), similar data were collected for 139 350 people, 80.2% of 174 316. Standardizing both numerator and denominator to the number of person years at risk increased these proportions by approximately 2% in both studies.

Of the 502 493 patients, 182 524 (36.3%) were interviewed and information about 235 454 (46.9%) was collected by proxy. Only 7381 patients (1.5%) refused to provide any item of socioeconomic information. A total of 63 003 people (12.5%) could not be contacted for interview — 22 308 (4.4%) were known to have moved away during the study year before an interview was attempted and 40 695 (8.1%) did not live at the address held by the practice. No reason was given by practices to explain why no interview took place for 14 131 patients (2.8%).

Some patients were unwilling or unable to answer individual questions about themselves or another member of the family (Table 1). Among these questions, that about economic position presented the greatest problem. Only 44 people refused to answer this question, but 4494 did not know the answer (1586 about their own economic position). The largest number of refusals

Table 1. Percentage of patients refusing to answer and for whom 'not known' was entered, for selected questions, by type of interview.

Question	Percentage of respondents (n = 182 524/235 454)					
	Refused		Not known		Total	
	Personal	Proxy	Personal	Proxy	Personal	Proxy
Marital status	0.01	0.00	0.02	0.02	0.03	0.03
Tenure of housing	0.06	0.05	0.07	0.08	0.13	0.13
Ethnic group	0.16	0.18	0.27	0.52	0.43	0.70
Smoked in last week	0.01	0.01	0.05	0.67	0.06	0.68
Economic position						
Last week	0.01	0.01	0.87	1.24	0.88	1.24
One year ago	0.03	0.02	0.09	0.65	0.12	0.67

n = number of personal interviews/proxy interviews.

Table 2. Percentage of patients interviewed by 54 practices ranked by total percentage of patients interviewed.

	% of patients		
	Fourth-ranked practice	Median value	57th-ranked practice
Personal interview	30.5	37.2	44.9
Proxy interview	35.2	47.2	56.4
Total	66.7	84.5	95.8
Refused interview	0.1	0.5	4.2
No interview			
Known to have moved	0.5	3.9	10.4
Unable to be contacted	0.5	4.6	17.8
No reason given	0.0	0.5	11.5

occurred in relation to ethnic origin. A relatively large proportion of respondents did not know whether other members of the family had smoked in the previous week. Few refused to answer the question on smoking. While there was little difference between the refusal rate by type of interview, the number of 'not known' answers was much greater for proxy interviews than for direct interviews.

The degree of practice variability in the proportion of interviews achieved by the middle 90% of practices (54 practices) is shown in Table 2. Practices were ranked (low to high) according to a variety of interview achievements. For each of the achievements, we present the results in the 4th- and 57th-ranked practices, providing an indication of the range of the results in 54 out of the 60 practices. We also present the median practice achievement. Of the total interviews achieved, 4th-lowest ranked practice achieved only 66.7%, whereas the 57th ranked achieved 95.8%. Three out of the 60 practices obtained data on less than 66% of the practice patients, while 18 practices interviewed more than 90% of patients registered at any time during the year. The study target was for 90% of the patients registered at the beginning of the study to be interviewed — 54 out of the 60 practices achieved this target. The mean ratio of proxy to personal interviews was 1.3; 90% of practices ranged between 0.75 and 1.5. Rates of refusal to be interviewed were low, with a mean for the 60 practices of 1.5% — in 52 practices, the rate was less than 2%, but three practices exceeded 14%. In 13 practices, the percentage of patients who could not be interviewed because they were known to have moved or were not at the address held by the practice exceeded 20%.

Costs

The costs of socioeconomic data capture were estimated in pilot studies conducted in four practices in 1986-1987 and provided the basis for remunerating practices during the study. It was estimated that all necessary organization and data entry would involve processing data at a rate of 4.5 patients an hour and that, on average, each face-to-face interview would yield data for 2.25 patients. Extrapolation of these data suggested that a full-time worker would be able to acquire data for a practice of 7500 persons during the study year. These estimates proved reliable for this study as each interview yielded on average 2.3 patient data sets. If the interviewer post is costed at £10 000 per year, the cost is approximately £1.30 per patient.

Discussion

Socioeconomic factors are important variables in relation to health and the use of general practice services. They may in future contribute to the design of performance indicators in general practice.⁷ They are necessary to assist in appropriate budget-

ing by practices, which are taking increasing responsibility for providing care for specified populations.⁸ Morbidity and socioeconomic data linked in individual patients are essential for analysis of the social components of health. Linkage is an advantage in individual practices for epidemiological research and it is only as a result of the analysis of linked data that other methods can be employed to estimate morbidity based on the socioeconomic characteristics of other areas.^{9,10} Socioeconomic data collected about individuals provide a means to alert general practitioners to those with special needs such as the unemployed, who are more likely to suffer from mental problems, or smokers, or children living with only one adult, who are more likely to consult for injury and poisoning. In general, there is a need for accurate methods of assessing primary health care needs.¹¹

The method used in the 1991-1992 study proved acceptable to patients and workable in the practices. The socioeconomic data collected indicated that the study sample was, in general, representative of the population of England and Wales. When compared with the results of the 1991 national census, the proportions were similar by age, sex, marital status, tenure of housing, economic position, and whether they lived in an urban or rural area. There were small differences by social class, ethnic origin and the proportion who lived alone. This general agreement on a sample of this size provides some evidence that data collection by this method is reliable.

The same proportion of the population as identified by the general household survey (29%) said that they had smoked in the previous week,¹² suggesting that the answers were generally honest. In addition, clearly, data collected from individuals about themselves were more satisfactory than data provided by proxy. The small proportion of 'not known' answers obtained at proxy interviews is unlikely to bias the results. The cost saving in conducting proxy interviews provides sufficient justification for using this method.

There were some deliberate omissions in the 1991-1992 study in the interests of simplification and cost saving. Because this was primarily a morbidity study, data were not collected about members of the family or household who were not registered with the practice except where they related to registered patients. There was also no link made between members of one family or household with the loss of interesting epidemiological information, such as any possible association between childhood morbidity and passive smoking.

Lessons can be learnt from the variation between practices in the percentage of interviews achieved. This study was a 'one-off'. During continuous routine data collection, staff may become more experienced in data collection and in methods of obtaining patients cooperation. However, it is important that standards are maintained and this may require training reinforcement. It is also important that there is standardization in methods between practices if the data are to be used for comparisons, and for estimates of expected morbidity and use of primary health care services. During a continuous study, it would be necessary to decide the frequency with which different questions would need to be asked; for example, a woman may get married, have a baby, become unemployed and change the type of housing in which she is living between one birthday and the next. In three practices, more than 14% patients refused to be interviewed, compared with the practice mean of 1.51%. It is possible that greater care in recruiting suitable interviewers and follow up by personal interview of 'not known' answers given by proxy would prove worthwhile.

There are many ways in which the method could be adapted to provide data for different purposes. Collecting socioeconomic data within a practice from a random sample of all patients or from a sample selected by criteria such as age or diagnosis could

achieve acceptable results and would reduce the cost. In addition, the range of questions asked in the 1991–1992 study could be reduced, asked over a limited period or collected by self-completed questionnaires instead of by interview. However, any sample would need to be a random sample of all patients on the register, and not only those for whom contact is easy. The use of computers in general practice makes selection of a random sample, and recording and analysis of data much easier than in the past. However, computerization is not at present without its problems, owing to the diversity of systems used.¹³

Methods have been described to estimate the expected numbers of people with morbidity in a small area, such as a local authority area or general practice population, in which survey data from different sources are synthesized.⁶ However, these methods require a knowledge of the socioeconomic structure of each practice which is only available from within the practice itself. By linking postcodes of patients on practice registers with census data at the postcode level, the social class¹⁰ and age¹⁴ distributions within practices have been estimated with unreliable results. Use of a self-administered questionnaire to determine occupation and derive social class was more encouraging.¹⁰ A postal questionnaire was sent to each household in order to derive social class at an approximate cost of £0.81 per questionnaire. In the study reported here, the labour including interview and data entry (costed at £10 000 per annum) is equivalent to £1.30 per interview. The personal interview provides much more comprehensive information and is particularly important for obtaining reliable information about the socially disadvantaged, who have greater difficulty with questionnaires. However, study methods must be geared to the aims, and for the fourth morbidity survey, a high emphasis was placed on good-quality socioeconomic data. The linkage of individual socioeconomic data with morbidity in a population exceeding 400 000 has been a remarkable achievement. We are not aware of any comparable study approaching this size having been conducted elsewhere.

The results of the 1991–1992 study can be used to estimate morbidity within a population such as that in a health district for which the socioeconomic characteristics are known. This is done by applying morbidity rates from the study to the local population, using either univariate or multivariate analysis. However, to estimate requirements for influenza immunization among elderly people in communal establishments, for example, a practice would need to know the number of people by age group in such establishments. This information can only be collected within the practice.

The purposes for which data are to be collected must be defined before deciding which data should be collected. The method used will be determined by those purposes. For some purposes, known morbidity rates can be applied to estimated populations. For others, socioeconomic data will need to be collected within the practice about each patient, so that the morbidity can be linked to defined populations, and individuals with socioeconomic characteristics of importance can be identified, and appropriate care or advice given. The method of data collection used in the 1991–1992 study would appear to be suitable for the production of a valuable database in general practices.

References

1. Office of Registrar General. *Morbidity statistics from general practice, 1955–56 (volumes I–III). Studies on medical and population subjects no 14.* London: HMSO, 1958.
2. Royal College of General Practitioners, Office of Population Censuses and Surveys and Department of Health and Social Security. *Morbidity statistics from general practice. Second national study, 1971–72. Studies on medical and population subjects no 36.* London: HMSO, 1979.

3. Royal College of General Practitioners, Office of Population Censuses and Surveys and Department of Health and Social Security. *Morbidity statistics from general practice. Second national study, 1970–71. Socioeconomic analysis. Studies on medical and population subjects no 46.* London: HMSO, 1982.
4. Royal College of General Practitioners, Office of Population Censuses and Surveys and Department of Health and Social Security. *Morbidity statistics from general practice. Third national study, 1981–82. Series MB5 no 1.* London: HMSO, 1986.
5. Royal College of General Practitioners, Office of Population Censuses and Surveys and Department of Health. *Morbidity statistics from general practice. Third national study: socioeconomic analysis 1981–82. Series MB5 no 2.* London: HMSO, 1990.
6. Royal College of General Practitioners, Office of Population Censuses and Surveys and Department of Health. *Morbidity statistics from general practice. Fourth national study, 1991–92. Series MB5 no 3.* London: HMSO, 1995.
7. Majeed FA, Voss S. Performance indicators in general practice. *BMJ* 1995; **311**: 210–221.
8. Ben-Shlomo Y, White I, McKeigue PM. Prediction of general practice workload from census based social deprivation scores. *J Epidemiol Community Health* 1992; **46**: 532–536.
9. Crombie DL. *Social class and health status: inequality or difference. Occasional paper 25.* London: Royal College of General Practitioners 1984.
10. Ward P, Morton-Jones AJ, Pringle MAL, Chilvers CED. Generating social class data in general practice. *Public Health* 1994; **108**: 279–287.
11. Shanks J, Kheraj S, Fish S. Better ways of assessing health needs in primary care. *BMJ* 1995; **310**: 480–481.
12. Office of Population Censuses and Surveys. *General household survey, 1992. Series GHS no 23.* London: HMSO, 1994.
13. Wilson AE, Pollock C, Weekes 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**: 227–230.
14. Scrivener G, Lloyd DCEF. Allocating census data to general practice population: implications for study of prescribing variation at practice level. *BMJ* 1995; **311**: 163–165.

Acknowledgements

We acknowledge the contributions of the participating practices, particularly the field workers assigned to the task of collecting socioeconomic data. We include with them Karen Ward, who was responsible for coordinating the training programme.

Address for correspondence

Dr DM Fleming, Royal College of General Practitioners Birmingham Research Unit, 54 Lordswood Road, Harborne, Birmingham B17 9DB.

Vale of Trent Faculty Royal College of General Practitioners

A Reading & Reference Book List for General Practice Compiled by Dr Brendan Jacobs FRCGP

This sixty page listing of books selected for GPs was published on 1st November 1995.

Contents include sections on 'the specialities' plus education, general practice, practice management and various topics of general interest etc.

Books are described by title, author, publisher, ISBN number, price, year of publication and the number of pages.

To order, contact: J Baily, Vale of Trent Faculty, RCGP, Department of General Practice, Medical School, Queen's Medical Centre, Nottingham NG7 2UH.

Fax: 0115 9709389. Price: £5.00 per copy, payable with order please. Cheques to be made payable to Vale of Trent Faculty, RCGP.