Accuracy of the patient identifier in a family practice data system

RUSSELL M. BOYLE, MA
Instructor, Department of Biostatistics, Medical College of Virginia, Richmond, Virginia, USA

SUMMARY. The accuracy of a general practice data system has been measured and has improved considerably over three years. It is difficult to identify all the factors contributing to this change, but an overall effort to emphasize the importance of recording patient identifiers correctly has been effective. No directly comparable error rates have been reported elsewhere; consequently, relative accuracy cannot be known. Administrators and managers of data systems are urged to determine levels of error and disseminate this information.

Introduction

GENERAL practice is unquestionably a 'data intensive' field. That is, a plethora of observations, measurements and facts can be generated for every patient. A logical method of managing these data has been the general practice data system, defined as the organized collection, storage (in some computer-readable medium) and analysis of data concerning any aspect of this area of medicine.

Systems of this type have numerous applications in patient care, education and research (Green *et al.*, 1978; Hollison *et al.*, 1978; Dickie *et al.*, 1976; Froom *et al.*, 1977). However, doctors and researchers often assume that all data, particularly those presented in impressive reports, are correct. The question of accuracy (the true value of a measurement) remains.

Aims

The purpose of this paper is to present one measure of the accuracy of the Virginia Family Practice Data System, a large-scale, continuous morbidity-recording system described earlier in several publications (Wood *et al.*, 1975; Marsland *et al.*, 1976; Boyle and Gehrig, 1977). The levels of accuracy uncovered are then compared with those in the literature. The importance of determining levels of accuracy and reporting them uniformly is stressed.

Methods

Patients are identified in the Virginia Family Practice Data System with an 11-character modified Hogben Code: the first three letters of last name, the first letter of first name and a six-digit numerical equivalent of date of birth and sex. Thus David Harris, born 4 July 1945, would be designated as HARD0704451. In the first stage of this study, patient records for the period 1 October to 31 December 1976 were assigned random numbers. These numbers were sorted sequentially and a 10 per cent sample was selected for each participating practice. Lists of these identifiers were subsequently verified at the practices; all codes which could not be matched with patients through age/sex registers, master files or patient charts were determined to be errors. The resulting Hogben Code error rate is simply the incorrect identifiers divided by the number reviewed. This same process was repeated for data recorded between 1 April and 30 June 1979.

Results

Table 1 presents the results. Thirteen practices participated in the study, but only those which were involved in both stages are shown for comparative purposes.

During the first period, there was a total Hogben Code error rate of 11·3 per cent. This indicates that at that time more than one in 10 patient identifiers in the Virginia Family Practice Data System represented a non-existent patient. A distinction can be made between teaching practices (practices 1 to 4 in Table 1) and non-teaching practices (practices 15 and 19). The former have larger patient populations, more doctors and (unfortunately) more stages through which the data must pass. Further, recording is mandatory. Non-teaching practices are considerably smaller, and the recorders know their patient populations better. Recording is voluntary at these practices, so that the more research-oriented doctors tend to participate.

Table 1. Virginia family practice data system: Hogben Code validation study.

Practice	1 October-31 December 1976			1 April-30 June 1979		
	Incorrect Hogben Codes	Hogben Codes reviewed	Hogben Code error rate (percentage)	Incorrect Hogben Codes	Hogben Codes reviewed	Hogben Code error rate (percentage)
1	57	751	*7.6*	40	619	*6.5*
2	29	200	*14.5*	65	669	*9.7*
3	61	384	*15.9*	38	336	*11.3*
4	53	300	*17.7*	73	457	*16.0*
15	6	143	*4.2*	8	114	*7.0*
19	8	108	*7.4*	4	98	*4.1*
Teaching	200	1635	*12.2*	216	2081	*10.4*
Community	14	251	*5.6*	12	212	*5.7*
Total	214	1886	*11.3*	228	2293	*9.9*

In the follow-up phase, there was an improved figure of 9·9 per cent. Lower error rates were found in each teaching practice, while non-teaching practices remained essentially unchanged.

Discussion

Clearly, the 11·3 per cent error rate is unacceptable. Among the causes are:

1. Inconsistent recording of the Hogben Code: using nicknames rather than first names, using middle names; errors in recording birth dates. (Marital name changes also create "new" patients.)
2. Transcription errors.
3. Keypunch errors during data entry. (A sample review showed that such errors could be roughly divided equally between transcription and keypunch.)

The second figure of 9·9 per cent was encouraging. On a general level, improvements may be attributed to continual emphasis (by working with recording secretaries and doctors) on the critical nature of the Hogben Code. Also noteworthy was the progress of a newly formed data and communication committee, which every month brought together representatives from all the teaching practices to discuss research projects and problems in recording. Looking at individual practices, the significant improvement in Practice 2, for example, was partially due to an innovation which required a fixed Hogben Code to be recorded for each patient on the actual patient chart. Ideally, this identifier is used consistently and transcends nicknames and name changes.

The consequences of error rates of the size revealed by the study should be obvious. For instance, patient counts are artificially inflated: routine and special reports show physicians seeing larger numbers of patients than is actually the case. This makes the use of patient consultations as the denominator in workload and morbidity rates inaccurate. Linkage of patient records over time, potentially one of the most useful aspects of a longitudinal data system, is also more difficult. Finally, data burdened with inaccuracies of 10 per cent are inappropriate for virtually any type of statistical analysis (Rockhold and Kilpatrick, unpublished).

It is interesting to examine the error rates presented above with regard to others reported in the literature. No directly comparable studies of North American data systems were found, so the following figures concern accuracy of general practitioners' records in the NHS. Table 2 represents a summary of measured "practice list inflation" in general practitioners' surgeries (Fraser, 1978). Thus, Morris (1964) found that one master list contained 16 per cent more patients than were actually registered with that particular practice, while Warren (1976) calculated a six per cent inflation. (It is not clear whether these practice lists were manual or automated.) In 60 practices participating in the 1971 National Morbidity Study, the percentage of non-matching records between individual practices and central NHS files ranged from zero per cent (eight practices) to seven per cent (one practice), with an overall error rate of approximately two per cent (Lambert and Birch, 1977).

Those responsible for operating the Virginia Family Practice Data System recognize the problem of patient identifier accuracy and have developed a partial solution. An automated age/sex file, consisting of all unique Hogben Codes ever recorded at each practice, has recently been created as a patient register. Incoming records are checked against this file by practice: the recording process remains the same with the exception of an "N" coded in the record for each new patient. The four possible outcomes are listed in Table 3. This has not yet been implemented at all practices, so no results are available. Given the close co-operation of the recording secretaries in correcting rejected patient records along with consistently identifying new patients, this should substantially reduce the problem of inaccurate patient identifiers.

Table 2. Practice list inflation: general practitioners' records in the National Health Service.

Year of publication	Author	Percentage inflation	Number of practices	Location
1964	Morris	*16*	1	London
1966	Shepard et al.	*8.5*	12	London
1969	Rees	*4.1*	1	England and Wales
1970	Morrell et al.	*20*	1	London
1976	Warren	*6*	1	Paddock Wood

Source: Fraser (1978).

Table 3. Automated age/sex file.

Patient identifier status	Age/sex register	Action
Old	Matched	Accept
New (with "N")	Not matched	Accept
Old	Not matched	Reject
New (with "N")	Matched	Reject

There are two irrefutable facts concerning data systems in general practice: people make mistakes, and mistakes are inevitable. With this in mind, a delineation of the three M's of data management is appropriate:

1. *Minimize*—Every effort should be made to reduce the amount of error.

2. *Measure*—The accuracy, validity and reliability of the data should be established.

3. *Make known*—Error rates should be reported to all internal users of the data system, at meetings of professional organizations and through publication in national and international journals. Ideally, measures should be uniform so that direct comparisons can be made: rates of inaccurate patient identifiers, missing or incorrect data elements in patient records and agreement between automated records and patient charts are examples. In this way, the quality of the data is known and improper applications can be avoided.

References

Boyle, R. & Gehrig Jr., J. J. (1977). Toward comparable family practice data. *Virginia Medical,* **104,** 553-558.

Dickie, G. L., Newell, J. P. & Bass, M. J. (1976). An information system for family practice. Part 4: encounter data and their uses. *Journal of Family Practice,* **3,** 639-644.

Fraser, R. C. (1978). The reliability and validity of the age–sex register as a population denominator in general practice. *Journal of the Royal College of General Practitioners,* **28,** 283-286.

Froom, J. *et al.* (1977). An integrated medical record and data system for primary care. *Journal of Family Practice,* **4,** 951-953, 1149-1151; **5,** 113-120, 265-270, 427-432, 627-630, 845-849, 1007-1015.

Green, L. A., Simmons, R. L., Reed, F. M., Warren, P. S. & Morrison, J. D. (1978). A family medicine information system: the beginning of a network for practising and resident family physicians. *Journal of Family Practice,* **7,** 567-576.

Hollison Jr., R. V., Vazquez, A. M. & Warner, D. H. (1978). A medical information system for ambulatory care, research and curriculum in an army family practice residency: 51, 113 patient problems. *Journal of Family Practice,* **7,** 787-795.

Lambert, P. M. & Birch, D. (1977). Errors in the morbidity study. (Letter). *Journal of the Royal College of General Practitioners,* **27,** 185.

Marsland, D. W., Wood, M. & Mayo, F. (1976). A data bank for patient care, curriculum and research in family practice: 526, 196 patient problems. *Journal of Family Practice,* **3,** 25-28.

Morrell, D. C., Gage, H. G. & Robinson, N. A. (1970). Patterns of demand in general practice, *Journal of the Royal College of General Practitioners,* **19,** 331-342.

Morris, J. N. (1964). *Uses of Epidemiology.* 2nd edn. Edinburgh and London: Livingstone.

Rees, M. S. (1969). *Journal of the Royal Statistical Society,* Series A, **132,** 526-542.

Rockhold, F. W. & Kilpatrick Jr., S. J. The denominator problem in family practice. (Unpublished manuscript).

Shepherd, M., Cooper, B., Brown, A. C. & Kalton, G. (1966) *Psychiatric Illness in General Practice.* London: Oxford University Press.

Warren, M. D. (1976). Identifying handicapped people in general practice. *Journal of the Royal College of General Practitioners,* **26,** 828-836.

Wood, M., Mayo, F. & Marsland, D. (1975). A systems approach to patient care, curriculum and research in family practice. *Journal of Medical Education,* **50,** 1106-1112.

Address for reprints

Russell M. Boyle, Department of Biostatistics, Box 32, MCV Station, Richmond, Virginia, 23298, USA.

Treating chronic bronchitis

Twenty consecutive elderly patients with chronic bronchitis admitted to a geriatric ward on maintenance salbutamol therapy were studied before and after their bronchitis therapy was discontinued for 24 hours. The patients were monitored over two days and forced expiratory volume in litres, forced vital capacity, peak flow, arterial oxygen and carbon dioxide pressures were measured. There was no significant change in any of the parameters measured when the salbutamol treatment was discontinued for 24 hours.

Source: Molloy, W. & Hyland, M. (1980). Maintenance salbutamol in chronic bronchitis in the elderly. *Age and Ageing,* **9,** 272-274.