The role of record linkage in post-marketing drug surveillance

I. K. CROMBIE

Medicines Evaluation and Monitoring Group, Ninewells Hospital, Dundee

Introduction

Record linkage is a particular type of cohort study of post-marketing drug surveillance (PMDS). Recent concern over adverse drug reactions has led to the development of several cohort methods of PMDS and Rawlins (1984) has reviewed four separate completed studies. Cohort studies involve identifying patients taking a particular drug, and subsequently monitoring these patients to detect any adverse events occurring among them. Identification of the drug takers is comparatively straightforward because of the strict legal and financial controls governing the issue and dispensing of prescriptions. It is the subsequent follow-up of these patients which is difficult and time consuming since it involves collecting and processing data on the events which have occurred to every person. The particular advantage offered by record linkage is that it greatly facilitates this follow-up.

Record linkage

The term record linkage refers to a method of bringing together information about an individual from different records, although it is most often used to describe the linkage of records about every individual in a large population

(Skegg, 1980). Several groups have investigated different systems for record linkage in PMDS and the nature and size of the populations under study are shown in Table 1.

The central feature of all these studies is their use of a computerised medical database. With the exception of the Oxford Group, these medical data have been collected, coded and computerised for financial or administrative purposes. Thus, they can be conveniently and cheaply used for the follow-up of patients in PMDS studies.

The Tayside record linkage system

The essential features and advantages of record linkage as well as its limitations can be illustrated by considering a study which we have recently completed in Tayside on the drug cimetidine. The advantage of Tayside for the conduct of PMDS studies is that every person registered with a GP has been allocated a unique number, the so called Community Health Number. This number provides the method of identification and access to separate computer files which contain patient biographical details, details of all hospital discharges, death notifications and also child development records (Angus *et al.*, 1978).

Table 1 Nature and size of record linkage population

<i>Nature of Population</i>	<i>Size</i>	<i>Reference</i>
Members of the Kaiser-Permanente Health Centre, San Francisco	120 000	Friedman <i>et al.</i> (1971)
Population of Finland	4 600 000	Indänpään-Heikkilä (1977)
Members of Group Health Cooperative, Seattle	250 000	Jick <i>et al.</i> (1979)
Patients of 20 General Practitioners, Oxford, England	43 000	Skegg & Doll (1981)
Residents of Tayside, Scotland	400 000	Crombie <i>et al.</i> (1984)
Medicaid recipients in the States of Michigan and Minnesota, USA	1 440 000	Jones <i>et al.</i> (1984)

Details of the methods and results of the feasibility study which we conducted using the drug cimetidine have been published elsewhere (Crombie *et al.*, 1984). Briefly, the prescriptions for cimetidine which were dispensed in Tayside over a 9 month period were obtained from the Prescription Pricing Division. The Community Health Numbers of those who had been prescribed cimetidine were identified manually by matching the name and address of the prescription to a complete alphabetic listing of Tayside residents. The computerized hospital discharge records were then searched to retrieve the morbidity data on the cimetidine recipients. To assist in the interpretation of the follow-up data on the cimetidine takers, we selected an equal number of community based controls matched to the drug takers by age, sex and General Practitioner. The feasibility study showed that Community Health Numbers could be allocated to a high proportion (76%) of prescriptions and that data on the hospital discharges of patients identified could be retrieved.

Details of the discharge diagnoses among the cimetidine takers and an equal number of controls are shown in Table 2. The drug takers experienced large excesses of many diseases, most notably of the digestive system, but also of many other systems including the musculoskeletal and the respiratory systems. Remarkably, similar results were obtained in a separate PMDS study of cimetidine (Colin-Jones *et al.*, 1981) suggesting that the pattern of disease excess is a true reflection of the morbidity experience of the cimetidine takers. Fuller consideration of these morbidity data indicates that the disease excesses may arise through non-causal drug-disease associations (Crombie *et al.*, 1984). Some examples of the associations and an outline of possible methods to overcome the problems which they pose are given in the next section.

Non-causal drug-disease associations

The problem of non-causal drug disease associations was raised by Skegg & Doll (1981) in the Oxford Record Linkage Study. These authors identified several categories of non-causal drug-disease associations in their study and examples from the cimetidine data are shown in Table 3. The excess of peptic ulcer could arise because it is a disease for which cimetidine is prescribed and the drug could also be prescribed for the indigestion occurring with hiatus hernia, again accounting for the excess. It is well known that non-steroidal anti-inflammatory drugs cause gastric irritation for which cimetidine might be prescribed. The excess of respiratory disease could occur if, as has been suggested (Bouchier, 1981), smoking is also a risk factor for peptic ulcer.

These and possibly other categories of non-causal drug-disease associations will be a common feature of cohort studies of PMDS and constitute one of the major obstacles to the detection of adverse drug reactions. This was not a limitation of the cimetidine study since it was intended to assess the feasibility of record linkage for PMDS, to identify methodological problems, and to suggest ways to overcome these problems. It was not intended to assess the safety or otherwise of cimetidine and it would be premature to do so. Cohort studies of PMDS are at an early stage of development and there is a need for critical assessment and discussion before we can be confident of their ability to detect previously unsuspected adverse reactions. Three possible approaches to the problem of non-causal drug-disease associations are considered in the next section.

Solutions for problems

The similar drug control group

The ideal would be to obtain as a control group,

Table 2 Number of discharge diagnoses in major disease groupings

<i>Disease group (ICD^a No.)</i>	<i>Number of discharge diagnoses</i>		<i>Ratio</i>
	<i>Cimetidine</i>	<i>Control^b</i>	
Digestive system (520-579)	515	52	9.9
Signs, symptoms and ill-defined conditions (780-799)	195	62	3.1
Musculoskeletal system (710-739)	72	38	1.9
Neoplasms (140-239)	163	87	1.9
Respiratory system (460-519)	93	53	1.8
Circulatory system (390-459)	199	116	1.7

^a9th Revision of the ICD.

^bEqual number of controls matched by age, sex and GP

Table 3 Categories of non-causal drug disease associations

Type	Example with cimetidine takers	Number of discharge diagnoses
Drug is a treatment for the disease	Peptic ulcer	207/2 ^a
Drug is a treatment for a complication of the disease	Hiatus hernia	15/0
Drug is treatment for unwanted effects of other drugs	Musculoskeletal disease and NSAID	72/38
Confounding	Respiratory disease via smoking	93/53

^aRatio of discharges among cimetidine takers/controls

a set of patients who resembled the drug takers in all ways except exposure to the drug under scrutiny. Inman (1981) has suggested using as controls those patients taking a drug which is chemically or pharmacologically similar and is marketed for the same indications as the index drug. This is an attractive approach suitable for record linkage and one which should be adopted where possible. But it may not solve all problems because of the difficulty of finding a drug which is marketed for and prescribed to an identical set of patients. For example, a newly marketed drug might be prescribed more frequently to patients with newly diagnosed disease or who did not respond to or were intolerant of previous therapy. In contrast, a doctor may be reluctant to prescribe a new drug, with which he has little experience, to pregnant women or young mothers, or to patients who are seriously ill with other diseases.

The careful analytical approach

Friedman (1972) has proposed that drug-disease associations be assessed and those of potential interest be investigated further. He has outlined several criteria for the assessment which include the strength and statistical significance of the association as well as expert clinical and pharmacological appraisal.

Pre-prescription morbidity analysis

A third approach to the problem is to investigate the morbidity of the drug takers prior to the prescription of the drug. The rationale for this is that adverse reactions will only occur after the drug has been prescribed, so that excesses of diseases seen in both the pre- and the post-prescription period will have another explanation. This analysis is possible with record linkage because the relevant data are already computer-

ised and can be easily accessed. Table 4 shows the average number of discharges per year for the periods 2-5 years pre-prescription, 1 year pre-prescription and 1 year post-prescription for three diseases which showed excesses in the post-prescription period. Peptic ulcer is a disease in which some patients show a history of recurrence so that we would expect to see an excess among cimetidine takers dating back over several years. In contrast, gastrointestinal haemorrhage shows a very large excess in the first year prior to drug prescription, suggesting that the bleeding may have led to cimetidine being prescribed. For musculoskeletal disease the excess among cimetidine takers dates back over many years, which is consistent with the non-causal explanation given in Table 3.

Clearly in the conduct of PMDS studies it would be advantageous to utilise all three approaches.

The consequences of spontaneous morbidity

A second problem revealed by the feasibility study is the large amount of spontaneous (i.e. not drug induced) morbidity which was observed. Thus, 401 of the 3802 controls had at least one hospital discharge in the 12 month follow-up period, and there were a total of 654 discharges among the controls. This morbidity could have an important limiting effect on the ability of cohort PMDS studies to detect adverse reactions. The problem is that we are not simply trying to detect a few cases of a unique disease (adverse reaction), but rather to detect the additional cases (induced by the drug) of a disease which also occurs among the controls. The consequences of this have been pointed out previously (Wardell *et al.*, 1979; Lewis, 1981), but may have been overlooked by some authors (Inman, 1981; Rawlins, 1984). The problem is illustrated in

Table 4 Pre-prescription morbidity analysis

	<i>Pre-prescription</i>		<i>Post-prescription</i>
	<i>2-5 years</i>	<i>1 year</i>	<i>1 year</i>
Peptic ulcer	64/6 ^a	195/3	207/2
Gastrointestinal haemorrhage	0/0	82/1	21/1
Musculoskeletal disease	42/22	74/38	72/38

^a Average number of discharges per year: cimetidine takers/controls

Table 5 Cohort size to detect adverse reactions of frequency 1/1000^a

<i>Background frequency of disease</i>	<i>Example of disease occurring with that frequency</i>	<i>Size of cohort of drug takers^b</i>
3/10 000	Cancer of the large bowel	14 000
1/1000	Stroke	26 000
1/100	Heart disease	177 000

^a At the 5% significance level (one-tailed) with power of 90%, using the equation given by Schlesselman (1974).

^b An equal number of control patients will also be required.

Table 5 which shows the size of cohort study which would be needed to detect an adverse reaction occurring among one patient per thousand, given different background frequencies of the disease.

If the disease occurred with a very low frequency among the controls say that of cancer of the large bowel, then a cohort of 14 000 drug takers and a similar number of controls would be required to detect an adverse reaction (for a significance level of 5%, and a power of 90%). If the disease occurred among the controls at the same frequency as stroke, a cohort of 26 000 would have to be recruited. However, if the disease occurred as commonly among the controls as heart disease then a cohort of nearly 180 000 would be required. Note that an equal number of suitable controls would also have to be recruited. These statistical calculations show that in practice adverse reactions of a frequency of one in a thousand will seldom be detected.

The advantages and limitations of record linkage

The advantages of record linkage derive from its use of a computerised medical database which provides convenient and cheap follow-up data, so that studies using this method may cost less than one tenth of alternative methods (Crombie *et al.*, 1984). Further, because the follow-up

period can be extended for as long as desired, record linkage may be the only method suitable for the detection of adverse reactions such as cancer which take many years to develop. Because it does not require the cooperation of physicians or other professional groups, the method will enable many drugs to be investigated. In addition, record linkage enables morbidity in the pre-prescription period to be investigated to identify non-causal drug-disease associations.

Some record linkage systems will also enable the investigation of potential drug induced congenital abnormalities. This is possible if there is a link between computerised records of mother and baby, and the way such a study could be conducted in Tayside has already been described (Crombie, 1984). Jick *et al.* (1981) have reported a study on the possible link between vaginal spermicides and congenital disorders using record linkage.

It is an interesting irony that the source of the advantages of record linkage, the computerised database, is also the cause of most of its limitations. The medical data are collected, coded and computerised for financial or administrative purposes. Thus the accuracy and precision of coding of disease may not be of the standard required for medical research. Further, the diseases are coded to the International Classification of Diseases, which may not be the most appropriate code for the detection of potential adverse

Table 6 ICD codes to which jaundice might be allocated

<i>Disease group</i>	<i>Description</i>	<i>ICD code^a</i>
Skin	Jaundice (yellow)	782.4
Infectious disease	Jaundice, febrile (acute)	70.1
Blood	Jaundice, haematogenous	283.9
Metabolic	Jaundice idiopathic (chronic)	277.4
Liver	Jaundice, malignant	570
Biliary tract	Jaundice, obstructive	576.8
Gallbladder	Cholestasis	575.8

^a9th revision of the ICD (WHO, 1977)

reactions. The problem can be appreciated if we consider what might have happened in a PMDS study to detect the cholestatic jaundice caused by benoxaprofen. We would be dealing with a previously unsuspected adverse reaction, so that the cases would be diagnosed as some kind of jaundice of unknown cause. Table 6 shows that depending on the qualifying terms which accompany it, jaundice could be coded to any of the six widely separated diagnostic categories and, if the adverse reaction were more precisely described as cholestasis, it would be allocated to a seventh category. This problem of coding may be particularly acute for previously unrecognised adverse reactions, since these will tend to be symptomatic descriptions and the ICD is not intended for the classification of symptoms. It may prove necessary to develop a symptomatic classification in which the ICD codes for particular symptoms are grouped together.

A further limitation of record linkage system is the size of population which they embrace. Table 1 shows that although the Finland and Medicaid schemes cover over 1 000 000 people the others are restricted to a few hundred thousand. This will make it difficult for these smaller groups to recruit cohorts of drug takers of the order of 10 000, and these schemes may be restricted to drugs which are frequently prescribed, i.e. to drugs prescribed to between 1–5% of the population.

Finally, with the exception of the Medicaid system, all these record linkage systems are limited to hospital in-patients events and so will only be able to detect serious reactions.

The role of record linkage

This paper has reviewed some of the outstanding problems facing cohort studies of PMDS. These

problems will require further investigation and, because of its ease and cheapness, would be best studied using record linkage. For example, by selecting a set of similar drugs, such as the β -adrenoceptor blockers or the non-steroidal anti-inflammatories, it would be possible to investigate the similarities and differences in the disease profiles of patients taking the individual drugs. This would assess the ability of different approaches to identify non-causal drug-disease associations. Further, it might indicate what constitutes an appropriate control group or whether more than one type of control (other drug, hospital attendee, community-based) would be required.

It is clear that record linkage is most suited to the detection of serious adverse reactions, particularly those which are clearly defined clinically. The problems of the complexity of the morbidity of the drug takers and the occurrence of spontaneous morbidity among the controls will restrict the ability of record linkage and other cohort methods of PMDS to detect rare adverse reactions. In most circumstances, adverse reactions with frequencies of one per thousand or less will fail to be detected. Record linkage will be of particular importance for the detection of drug induced congenital abnormalities and for adverse reaction which take several years to develop.

The development of the Tayside record linkage project is due to the vision and enthusiasm of the late Professor James Crooks CBE. This is one of the many areas of drug epidemiology in which he has made important contributions and it is particularly appropriate to acknowledge his central role in record linkage at this Memorial Symposium.

References

- Angus, J., Cameron, L., Finlayson, C. F., Hopkins, R. T., Martin, K. B. & Scott, J. H. (1978). *The Tayside master patient index*. Dundee: Tayside Health Board.
- Bouchier, I. A. D. (1981). Peptic ulcer: the changing scene. *Scot. med. J.*, **26**, 297-302.
- Colin-Jones, D., Langman, M. J. S., Lawson, D. H. & Vessey, M. P. (1981). Postmarketing surveillance. In *Cimetidine in the 80s*, ed Baron, J. H., pp. 270-274. Edinburgh: Churchill Livingstone.
- Crombie, I. K. (1984). The Dundee Record Linkage Study. In *Monitoring for adverse drug reactions*, eds Goldberg, A., Grahame-Smith, D. & Walker, S. Lancaster: MTP Press.
- Crombie, I. K., Brown, S. V. & Hamley, J. G. (1984). Postmarketing drug surveillance by record linkage in Tayside. *J. Epidemiol. Community Health*, **38**, 226-231.
- Friedman, G. D. (1972). Screening criteria for drug monitoring. *J. Chronic Dis.*, **25**, 11-20.
- Friedman, G. D., Collen, M. F., Harris, L. E., Van Brunt, E. E. & David, L. S. (1971). Experience in monitoring drug reactions in out-patients. *J. Am. med. Ass.*, **217**, 567-572.
- Idänpään-Heikkilä, J. (1977). Population monitoring: Medical record linkage for drug safety surveillance. In *Drug Monitoring*, eds Gross, F. H. & Inman, W. H. W., pp. 17-26. London: Academic Press.
- Inman, W. H. W. (1981). Postmarketing surveillance of adverse drug reactions in general practice. *Br. med. J.*, **282**, 1131-1132, **282**, 1216-1217.
- Jick, H., Walker, A. M. & Rothman, K. J. (1981). Vaginal spermicides and congenital disorders. *J. Am. med. Ass.*, **245**, 1329-1332.
- Jick, H., Watkins, R. N. & Hunter, J. R. (1979). Replacement estrogens and endometrial cancer. *New Engl. J. Med.*, **300**, 218-222.
- Jones, J. K., van der Carr, S. W., Rosa, F., Morse, L. & LeRoy, A. (1984). Medicaid drug-event data: an emerging tool for evaluation of drug risk. *Acta med. Scand. (Suppl)*, **683**, 127-134.
- Lewis, J. A. (1981). Postmarketing surveillance: how many patients? *Trends Pharmac. Sci.*, **2**, 93-94.
- Rawlins, M. D. (1984). Postmarketing surveillance of adverse reactions to drugs. *Br. med. J.*, **288**, 879-880.
- Schlesselman, J. J. (1974). Sample size requirements in cohort and case-control studies of disease. *Am. J. Epidemiol.*, **99**, 381-384.
- Skegg, D. C. G. (1980). Medical record linkage. In *Monitoring for drug safety*, ed Inman, W. H. W., pp. 337-348. Lancaster: MTP Press.
- Skegg, D. C. G. & Doll, R. (1981). Record linkage for drug monitoring. *J. Epidemiol. Community Health*, **35**, 25-31.
- Wardell, W. M., Tsianco, M. C., Anavekar, S. N. & Davis, H. T. (1979). Postmarketing surveillance of new drugs. I. Review of objectives and methodology. *J. clin. Pharmac.*, **19**, 85-94.
- WHO (1977). *International classification of disease*, Volume 2. London: HMSO.