LEADING ARTICLE

Drug Safety 2000 Feb; 22 (2): 97-102
0114-5916/00/0002-0097/$20.00/0

© Adis International Limited. All rights reserved.

Use of Routine Healthcare Data in Safe and Cost-Effective Drug Use

Craig J. Currie[1] and *Tom M. MacDonald*[2]

1 Global Health Outcomes, GlaxoWellcome R&D, Greenford, Middlesex, England
2 Medicines Monitoring Unit (MEMO), Department of Clinical Pharmacology, Ninewells Hospital and Medical School, University of Dundee, Dundee, Scotland

Abstract

Routine healthcare data is becoming widely available, usually as a result of administrative systems. Other related data are also often available, such as biochemistry results, mortality data, and sometimes prescribing data. These records are often linked via a common identification system or by probability matching techniques. These data sources offer many opportunities to undertake research, and where prescription data are recorded and linked, the facility to research the outcome of drug use often exists. There are now a number of research agencies around the world that use these large routine data sources to undertake drug safety and outcome studies. The purpose of this commentary is to describe some of the history behind the development of these systems, illustrate some of their uses with respect to postmarketing drug safety and to other healthcare research objectives. The review then describes the data sources necessary to develop a system that would offer an optimal system to undertake a range of studies, including population drug safety surveillance. There are both positive and negative considerations when using routine data. On the positive side, these data come from 'real life' experiences and not from the clinical trial situation. On the other hand, there are important biases to be aware of such as confounding by indication. On the whole, it is argued that large databases originating from routine healthcare procedures have an important role to play in the cost-effective prescription drug use in the postmarketing setting. These systems cannot replace other methods of drug safety evaluation but they do offer an important adjunct to spontaneous reporting systems.

The purpose of this commentary is to discuss the use of routine computerised healthcare systems for drug safety surveillance, and argue that they are important alongside the present systems used for drug safety evaluation and monitoring. The use of routine computerised mechanisms offers many advantages over traditional methods; for example, the complete population coverage that routine systems allow can assist detection of rare adverse effects by allowing hypothesis generation and hypothesis testing. Population coverage does not select for the sample subgroups that are typically necessary to evaluate safety in drug trials. Methods involving large population databases complement spontaneous reporting methods. We then list examples of these systems in secondary and primary care in the UK and elsewhere. These systems are cost effective when employed for this purpose because they can simultaneously allow for the investigation of drug safety for all prescribed medicines should the

need or hypotheses arise. The review then describes the data requirements for an optimal system for population drug safety surveillance, and for other health intelligence purposes.

1. Computerisation in Health Systems

Computers and information technology have increasingly been adopted to help administer hospitals, general practices and other agencies involved in healthcare provision. In the UK, routine hospital patient administration systems were introduced in the mid to late 1960s. Along with the necessary demographic and administrative patient details, a typical routine inpatient record includes diagnostic coding using clinical coding systems such as the International Classification of Diseases (ICD), and coding that describes any operative procedures that have been undertaken.

More recently, in the 1980s, computerisation was introduced into primary care. In the UK, the introduction of these systems in primary and secondary care was uncoordinated; this was true especially in the primary care sector where a number of commercial organisations courted general practitioners, for example, Value Added Medical Products (VAMP)[1] and AAH-Meditel.[2] The majority of these primary care systems employed the Read clinical coding system.[3,4] Routine hospital inpatient data in the UK became more organised in 1982 after the recommendations of the Steering Group on Health Service Information were adopted (commonly known as the Körner report).[5] Other health systems, for example, in the US, were notably better in terms of completeness and coding quality than those in use in the UK because more financial resources were targeted at healthcare administration in response to differing methods of healthcare remuneration.

Although healthcare computer systems have been generally designed for administrative purposes, collation of health data on a routine basis has offered opportunities for public health surveillance and research by converting this into meaningful health intelligence.

2. Routine Health Information and the Intelligence Cycle

The 'intelligence cycle' is the means by which we convert data into useful and disseminated health intelligence.

Many healthcare systems now record enormous amounts of health related data, for example, from patient administration systems, biochemistry results, pharmaceutical prescribing, or from other routine sources, as well as from *ad hoc* research initiatives. These data often describe the health experiences of entire populations, certainly in relation to hospital care. Nevertheless, they are rarely used for anything other than their original and limited purpose, and only very unusually are they analysed or examined in some collective way such as in some form of record linkage study. Enormous potential research yield is being missed by limiting the use of routine healthcare data. Secondary analyses of these sources are relatively inexpensive since such data already exist for their original administrative purposes.

The potential uses of routine health service data are very varied, for example, from low level intelligence on patient length of stay and other 'process' parameters, to higher level intelligence related to resource use,[6] quality of care,[7] and epidemiology.[8] Interestingly, the use of routine data sources potentially reduces or eliminates some sources of bias found in biomedical research, for example, procedure or patient selection bias, recall bias and volunteer bias.[9] However, studies using these data suffer from their own sources of biases, such as confounding by indication.[10] If, or when, prescribing data are linked to these various sources of data, a wealth of opportunities arise to study drug use in relation to outcome.

3. Drug Safety Surveillance

Drug safety surveillance can be broadly separated into 2 categories, pre- and postmarketing surveillance. Premarketing safety evaluation relates to laboratory assessment of drug toxicity and phases I to III clinical trials in patients. Use of large databases for drug safety evaluation and surveil-

© Adis International Limited. All rights reserved.

lance relates to mechanisms of postmarketing drug surveillance. Excluding methods that use these large databases – which do have some recognised limitations[11] – postmarketing drug surveillance data have come mainly from spontaneous reporting systems such as the 'Yellow Card' system in the UK, and from *ad hoc* company-based studies. However, Waller and colleagues[12] reviewed 31 such studies and identified serious limitations.

The wide array of drugs we use, as well as newly introduced compounds, are often very effective. The down-side is that pharmacotherapy can result in adverse effects. There is therefore a balance of risk in taking a drug between its positive therapeutic effects and its negative side effects. These adverse effects can evade detection in clinical trials, particularly if they occur with low frequency or involve a long time lag effect, or if the adverse event is of an unusual nature. Increasing the risk of road traffic accidents with benzodiazepines would be an example.

There is no doubt that the toxicological studies presently employed to avoid dangerous drugs getting into the marketplace usually work; nevertheless, it does sometimes happen that drugs with dangerous adverse effects are marketed. For example, the insulin sensitising agent troglitazone was withdrawn in the UK as a precautionary measure after it was found to be hepatotoxic in a very small number of cases.[13] This example illustrates that the present mechanism of drug safety evaluation is fallible, and needs supporting with postmarketing surveillance for spontaneous adverse effects, such as in the UK 'Yellow Card' system.

In summary, although we accept that the present methods of drug safety evaluation and surveillance are largely effective, as shown by the minimal number of occasions where a dangerous adverse effect is discovered after approval, there is still room for improvement. This improvement could be achieved by establishing routine mechanisms that bring together data that describe a range of patients' health experiences and include reference to the drugs that these patients are taking, whether being prescribed or encashed. Methods are available to draw together data from these disparate sources to identify the records that pertain to a single individual. Once established, these systems then offer a valuable resource for the study of a range of public health issues in populations.

4. Examples of Drug Surveillance Systems That Use Large Databases

The use of record-linkage techniques for the postmarketing surveillance of drugs was pioneered in the US and Canada, partly as a result of the more commercial approach to provision of healthcare. For example, Medicaid is the US joint federal-state healthcare insurance system, in existence since 1965 to provide medical care for the poor, needy and disadvantaged. Most widely used for postmarketing surveillance studies have been Tennessee data for 400 000 patients, data for 1 million patients in New Jersey, and data for 8 million patients from 11 states which have been combined in the Computerized On-Line Medical Pharmaceutical Analysis and Surveillance System (COMPASS).[14,15] Linkage in Medicaid is deterministic and therefore does not relay on probability matching techniques, with registered patients allocated a unique Medicaid number. Computerised data generated by healthcare contacts, including detailed information on dispensed drugs and inpatient and outpatient hospitalisation data, are collected for payment purposes and indexed by the Medicaid number. The main advantage of the Medicaid databases is that they are frequently large. However, because Medicaid caters for disadvantaged citizens, the registered population is not representative of the general US population and the generalisability of studies could be questioned. Also, people tend to move in and out of Medicaid eligibility, causing considerable loss to follow-up. There are no data on drug indication or lifestyle factors such as smoking which are often potentially confounding.

Patients in the US who are not eligible for Medicaid can enrol in a Health Maintenance Organisation (HMO), which provides comprehensive medical coverage for paid-up members. In 1991, there were 550 HMO plans in the US, covering 38 mil-

© Adis International Limited. All rights reserved.

lion people. Of these, the Group Health Co-operative of Puget Sound in Seattle (GHC), has been most widely used for postmarketing surveillance.[16] In 1992, GHC had approximately 370 000 members, who had accumulated 2 million outpatient visits, 30 000 hospital admissions, more than 3 million prescriptions and 2.5 million laboratory tests. Person-specific computerised information on prescriptions from a formulary that are dispensed in GHC pharmacies (over 90%), including over-the-counter medication, and hospital admissions to GHC hospitals, are collected for costing. Other datasets include inpatient drugs, a cancer registry, outpatient visits and laboratory and radiographic data. So, although the database is relatively small, the varied nature of the datasets is a major asset.

An example of a government or provincial healthcare plan in North America is Saskatchewan, which funds a comprehensive range of health services almost universally to the residents of Saskatchewan, Canada (population of 1 million). Computerised data on these services are collected centrally, with 10 separate databases available for research in 1993, with the proviso that rigorous guidelines are followed and confidentiality is maintained.[17] Record linkage is deterministic, dependent on a unique 9-digit number that patients are allocated when they register with Saskatchewan Health. A computerised prescription plan was established for Saskatchewan Health in 1975, with data available from pharmacy claims for dispensed drugs; but only those from the Saskatchewan drug formulary (comprising 2306 drug products in 1995). Examples of outcome databases include hospital discharges with limited routine validation of discharge codes (about 200 000 per year), physician services data, a cancer registration database, birth and death certificates, alcohol and drug abuse treatment, institutionalised care for the elderly and disabled, home care services and children's dental care. So, in contrast to Medicaid and HMOs, the Saskatchewan databases cover a large representative population, with centralised validation of linkage. However, information on exposure is limited by the use of a drug formulary, and there are no data on over-the-counter or inpatient medications. The indication for the drug is also not recorded. The databases are sensitive to administrative changes that can affect their suitability for research, a notable example being the collection of family-specific rather than patient-specific exposure data between July 1987 and January 1989. Access to the data is also strictly controlled.

An example of a record-linkage database that is unusual in that it developed within a primary care setting in the UK is the General Practice Research Database (GPRD).[18,19] GPRD data are derived from a GP computer system that records medical information for individual patients. Information routinely recorded includes an identification number for each patient, demographics, a record of all prescriptions written, and coded GP and hospital diagnoses. In theory, GPRD has coverage of several million patients over 6 or more years, although not all their data are of research quality. The exposure data are for prescribed rather than dispensed medication, with an estimate of 90% of recorded prescriptions actually dispensed. Information on hospital admissions may be incomplete, with inpatient diagnoses recorded for 87% of a sample of 1191 patients.[14]

The Medicines Monitoring Unit (MEMO) collects data for the population of Tayside, Scotland (approximately 400 000 people) which are record-linked and used for a variety of research purposes.[20] These data cover all prescriptions dispensed in community pharmacies, all inpatient hospital discharges (acute, maternity, psychiatric, neonatal), cancer registrations, a diabetes register, biochemistry data and road traffic accidents. The main strengths of MEMO are that it covers a representative population accurately defined, there are no formulary restrictions, there are detailed dispensed prescribing data, numerous outcome datasets and easy access to original medical records. However, its main weaknesses are limited size, no data on inpatient or over-the-counter medication, no information on indication, and limited information on confounding factors, although these criticisms are invariably true about the other sources.

© Adis International Limited. All rights reserved.

From this brief description of just a few of the databases that are available for research, it is clear that each has its own unique set of advantages and disadvantages. So, while no one database has a clear advantage over the others, the most important priority is that it is suited for the research questions for which it is used.

5. Other Public Health Benefits of Routine Surveillance Involving Drug Therapy

The evaluation of drug effectiveness does not stop at the drug safety stage. Appropriate prescribing practice such as correct dosage, patient compliance and awareness of drug interactions all play a role in maximising the potential benefits from pharmacotherapy. Other issues such as drug cost-effectiveness, the prescription of cheaper generic formulations and variation in prescribing practice are also important to public health organisations and purchasing authorities. The use of routine surveillance systems allows for all of these demands to be satisfied. In 2 examples from studies using the MEMO database we can illustrate the potential diversity for a range of studies surrounding the entire health experience in relation to the administration of prescription drugs. First, a recent paper seriously questions the notion of 'brittle diabetes'[21] by showing that the outcome of poor glycaemic control (ketoacidotic events) in these participants is a function of the quantity of insulin that they collect from pharmacies, and they simply do not collect enough insulin to maintain normoglycaemia.[22] Secondly, a study that examined the relationship between newly prescribed nonsteroidal anti-inflammatory drugs (NSAIDs) and upper gastrointestinal haemorrhage and perforation concluded that the increased risk with NSAIDs was only apparent for participants without a history of upper gastrointestinal events.[23]

6. What Would the Optimal System Look Like?

In an ideal world, for any given population, the health experiences that need to be recorded for comprehensive postmarketing drug safety surveillance, cost-effectiveness and other analyses are as follows:

- patient demographic details
- details of all drugs used (hospital, community and over-the-counter)
- routine computerised hospital records, including process measures (length of stay, method of admissions and discharge and so on), health plan enrolment/eligibility data, diagnostic coding and operation and procedure codes
- primary care data from general practice and community care
- biochemistry, radiology, microbiology, and other test results
- mortality records (ideally including coded cause of death and any contributing factors or diagnoses)
- patient experiences (functional health status, health related quality of life).

These various sources of data then need a mechanism to identify the records from each source that relate to the same individual. In Scotland, for example, the 'community health number' is used routinely on most relevant sources.[24] Methods do exist, however, to match records by probability where similar methods of attributing patient identity are not employed, for example the matching techniques developed for the Oxford Record Linkage Study.[25] The resources required to achieve this would not be that great, but other changes to the present drug remuneration mechanism may be necessary for most cost-efficient data capture. For example, throughout the UK, when details of prescribed drugs are computerised from prescription by the authorities, patient demographic details are excluded. Here in MEMO these details are included *post hoc*, although computer systems are now being piloted to download data directly from pharmacies.

Again, in an ideal world, these systems would all be online and in real time. There are considerations concerning patient confidentiality, but it can be argued that when used for research purposes these can be easily addressed because the identity

© Adis International Limited. All rights reserved.

of the patient, in terms of their name and address, is of little or no significance to the research investigator.

7. Conclusion

It can be argued that large databases originating from routine healthcare procedures have an important role to play in the achievement of safe and cost-effective prescription drug use in the post-marketing setting. These systems cannot replace other methods of drug safety evaluation such as spontaneous reporting mechanisms. What they can offer, though, is a flexible, long term method of collating data related to drug safety – and simultaneously many other public heath issues – in order to provide an information resource that describes the majority of the important health experiences of those in our populations. They offer an ongoing facility to test hypotheses related to questions of drug safety as and when they arise. Since they collate data from entire populations, the use of large databases also overcomes some of the important criticisms of the present methods of drug safety evaluation by drug companies, for example by eliminating sample selection bias.

Acknowledgements

Craig Currie works for GlaxoWellcome R&D. However, the objectives of this review involve no conflict of interest.

References

1. Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. BMJ 1991; 302: 776-8
2. Goves JR, Davies T, Reilly T. Computerisation of primary care in Wales. BMJ 1991; 303: 93-4
3. Chisholm J. The Read classification system. BMJ 1990; 300: 1092
4. Information Management Group. NHS Centre for coding classification: general information. Leicester: NHS Management Executive, 1992
5. Körner E (Chairman). Report on the collection and use of information about hospital clinical practice activity in the National Health Service. London: Her Majesty's Stationery Office, 1982
6. Williams DRR. Hospital admissions for diabetic patients: information from hospital activity analysis. Diabet Med 1984; 2: 27-32
7. Henderson J, Goldacre MJ, Graveney MJ, et al. Use of medical record linkage to study readmission rates. BMJ 1989; 299: 709-13
8. Currie CJ, Morgan CL, Gill L, et al. The epidemiology and costs of acute hospital care for cerebrovascular disease in diabetic and non-diabetic populations. Stroke 1997; 28: 1142-6
9. Sackett DL. Bias in analytic research. J Chronic Dis 1979; 32: 51-63
10. Poses RM, Smith WR, McClish DK, et al. Controlling for confounding by indication for treatment. Are administrative data equivalent to clinical data? Med Care 1995; 33: AS36-46
11. Shapiro S. The role of automated record linkage in the postmarketing surveillance of drug safety: a critique. Clin Pharmacol Ther 1989; 46: 371-86
12. Waller PC, Wood SM, Langman MJS, et al. Review of company postmarketing surveillance studies. BMJ 1992; 304: 1470-2
13. Mitchell P. Shock as troglitazone withdrawn in the UK. Lancet 1997; 350: 1685
14. Ray WA, Griffin MR. Use of Medicaid data for pharmacoepidemiology. Am J Epidemiol 1989; 129: 837-49
15. Strom BL, Carson JL. Use of automated databases for pharmacoepidemiology research. Epidemiol Rev 1990; 12: 87-107
16. Jick H. The Boston Collaborative Drug Surveillance Program and the Puget Sound Health Maintenance Organization. Drug Info J 1985; 19: 237-42
17. Malcolm E, Downey W, Strand LM, et al. Saskatchewan Health's linkable data bases and pharmacoepidemiology. Post Mark Surveill 1993; 6: 175-264
18. Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. BMJ 1991; 302: 766-8
19. Jick H, Terris BZ, Derby LE, et al. Further validation of information recorded on a general practitioner based computerized data resource in the United Kingdom. Pharmacoepidemiol Drug Saf 1992; 1: 347-9
20. Evans JMM, McDevitt DG, MacDonald TM. The Tayside Medicines Monitoring Unit (MEMO): a record-linkage system for pharmacovigilance. Pharmaceut Med 1995; 9: 177-84
21. Gill GV, Lucas S, Kent LA. Prevalence and characteristics of brittle diabetes in Britain. Q J Med 1996; 89: 839-43
22. Morris AD, Boyle DI, McMahon AD, et al. Adherence to insulin therapy, glycaemic control and ketoacidosis in insulin-dependent diabetes. Lancet 1997; 250: 1505-10
23. McMahon AD, Evans JM, White G, et al. A cohort study (with re-sampled comparator groups) to measure the association between new NSAID prescribing and upper gastrointestinal haemorrhage and perforation. J Clin Epidemiol 1997; 50: 351-6
24. Morris AD, Boyle DIR, MacAlpine R, et al. The diabetes audit and research in Tayside Scotland (DARTS) study: electronic record linkage to create a district diabetes register. BMJ 1997; 325: 524-8
25. Gill L, Goldacre M, Simmons H, et al. Computerised linkage of medical records: methodological guidelines. J Epidemiol Community Health 1993; 47: 316-9

Correspondence and reprints: Dr *Craig J. Currie*, Global Health Outcomes, GlaxoWellcome R&D, Greenford, Middlesex UB6 OHE, England.
E-mail: cc53675@glaxowellcome.co.uk

© Adis International Limited. All rights reserved.