Getting the most from routinely collected data

Jenny Hargreaves

This issue of the *Health Information Management Journal* has as its theme ‘Health Information Management in epidemiological research’, and all four papers illustrate differing ways in which Health Information Management can contribute to epidemiological endeavours. More than that, though, the papers also convey valuable messages about the importance of appropriate use of routinely collected data, and of understanding and improving the quality of the data and the infrastructures that underpin the data collections. These messages are all important for epidemiological research and other analytical uses of Australia’s routinely collected health data.

Michelle Bramley’s paper (Bramley 2005) provides a comprehensive guide to the evaluation of health classifications, which are core components of health data collection infrastructures. The paper includes useful background information comparing classifications and terminologies, and describes the important attributes of classifications, including their mono-hierarchical structure, comprehensiveness, aggregation of multiple concepts within rubrics, and meaningful coding systems. It also details the features that good classifications should have, covering the administrative, structural, content and usability domains by which classifications can be evaluated. Practical examples of evaluating health classifications include the Anatomical, Therapeutic, Chemical classification with Defined Daily Doses (ATC/DDD) drugs classification that has been submitted for inclusion in the Australian Family of Health Classifications, and is being evaluated against the criteria for inclusion in the Family (AIHW 2005a).

To guide evaluations of classifications, the paper provides useful examples to illustrate the nature of classifications and the characteristics of good classifications. This comprehensive array also provides valuable information for those who use classifications in data analysis work, including epidemiology. In addition to the explanation of mono-hierarchical structures and aggregation of multiple concepts within rubrics, the paper explains the existence of ‘residual’ categories: multiaxial structures; revision processes and mappings between versions; appropriate use of category descriptors; and the importance of definitions, indexes and guidelines that accompany classifications. The paper also urges readers to consider the suitability of classifications to the use proposed for them (for example, in terms of granularity), and their comparability with other classifications used in related domains or internationally. All these are important messages for data analysts.

Andrew Miller’s paper (Miller 2005) is an example of an endeavour that is likely to become more common: optimising interactions between data collections established for clinical purposes (increasingly as electronic health records) and data collected for administrative or epidemiological purposes, with an aim of improving the quality of data for research and monitoring activities.

The paper argues that the epidemiological data collected in relation to radiation oncology are useful to the extent that they are quantified and systematised, but are really only useable if they are also of good quality, accurate and complete. The paper also argues that data compiled in separate clinical systems, such as those used by radiation oncologists, are highly accurate and complete, but suffer from being less quantified and analyst-friendly.

Dr Miller therefore argues that the clinical data systems should be established so that all the data are regarded as and collected as if they were research data, so that data flows can be established between the clinical systems and hospital information systems. He argues that the quality of the clinical data will be enhanced through data ‘ownership’; that is, if the generators of the information also have responsibility for data creation, collection and integrity.

The paper by Quoc Nguyen and Beth Reid (Nguyen and Reid 2005) on sources of data on fungaemia provides a helpful illustration of the importance of understanding the strengths and limitations of routinely collected data, and the way in which work to detail the limitations could be used to improve routinely collected hospital morbidity data. The technique used was to compare two separate data sources. This is a classical method used to evaluate surveillance systems (Centers for Disease Control and Prevention 2001), and considered to be a very useful method when the two data sources are compiled independently, as was the case for this paper. The second data source (considered to be the gold standard for the study) was the pathology data system at the two study hospitals. In a useful addition to the comparison of the data, the relevant medical records were reviewed to assess whether and where the fungaemia was reported.

Comparison of the two data sources showed some miscoding and missed coding, with fewer than half of the pathology database cases including an appropriate fungaemia code, despite evidence of an infection being found in 97% of the relevant medical records. Coding was found to have been affected by the location and nature of the information about the fungaemia in the medical record. The authors indicated a need for improved coding, documentation and coding guidance in this area, and possibly more specific ICD-10-AM codes for recording fungaemias and other fungal infections.

This is an excellent example of a detailed local assessment of the quality of hospital morbidity data with the potential to lead to data quality improvements, at the local and national levels, for an important topic relevant to the safety and quality of hospital care. There has been a range of previous studies of the quality of coded data in Australia (including those cited in the paper) but many have been based on re-extracting data from the medical records (rather than a comparison with an external data source) and they have not usually included information on whether or where the information to be coded was included in the medical record. This paper includes that additional information and therefore could be used to inform data quality improvement, through appropriate changes to guidance on documentation or coding.
The limitations of the codes available for fungal infections described in the paper could also be considered, perhaps through creation of more specific codes. Alternatively, data collection systems could be redesigned to link aetiology and manifestation codes, perhaps allowing them to be paired more meaningfully for purposes such as fungaemia surveillance, and unambiguous interpretation of the codes in downstream applications (such as the Australian Institute of Health and Welfare’s [AIHW’s] National Hospital Morbidity Database).

This article can be linked with the article by Miller, in that it describes a data collection system that is separate from the routinely collected morbidity data, but is regarded as the gold standard. Perhaps in this type of situation, similar to Miller’s suggestion in relation to radiation oncology, there should be investigation of an automated mechanism for pathology results to be added to coded hospital morbidity records, on confirmation that a positive test result represented a clinically significant diagnosis that met the definition of an additional diagnosis.

Merilyn Riley’s paper on birth defects presents a range of interesting statistics on birth defects from the Victorian Birth Defects Register (Riley 2005). However, a major focus is information about the sources and quality of the data, analysis techniques used and appropriate interpretation of the statistics. The paper provides detailed information on the definitions used for the VBDR, the multiple sources of data used, and the methods used to ensure the quality of the data. Data from the multiple sources are linked using demographic data and names of the mother and baby, as possible, and would help to ensure that the multiple sources do not result in over-counts of defects.

The discussion canvasses a range of issues relevant to appropriate analysis and interpretation of the data. Included are impacts of variation in the types of birth defects included, variation in the maximum age of children for whom birth defects can be reported, variation in case ascertainment over time, whether induced abortions before 20 weeks’ gestation are included, whether cases or individual birth defects are being counted, and the time periods chosen for trend analyses. Examples of different statistics resulting from different decisions made in relation to these points provide excellent illustrations of considerations relevant to use of routinely collected data for epidemiological purposes.

Important messages from this paper are that the data should be analysed to suit the question being asked and, to ensure appropriate interstate comparisons, data collation and analysis methods used in other states and territories will also need to be understood. This paper therefore provides plenty of food for thought for the current work to develop a National Minimum Data Set for Congenital Anomalies, being undertaken by the AIHW’s National Perinatal Statistics Unit (AIHW NPSU 2004). This Data Set will aim to improve the usefulness and comparability of data on birth anomalies at the national level.

My own particular interest in this type of work relates to use of the National Hospital Morbidity Database held at the AIHW. This data source is used for a range of epidemiological purposes by the AIHW (for example, in relation to chronic kidney disease [AIHW 2005b], and others (for example, in relation to the use of hospital beds by older people [Gray, Yeo & Duckett 2004]).

Because of the database’s comprehensive coverage of essentially all hospitals in Australia, the generally well established infrastructures for compiling it, and the potential for use of the database in describing the impact of disease, the epidemiology of interventions and aspects of quality of care, it is important that as much as possible is known about the quality of the data. At the national level, there have been some studies of the quality of the data, using techniques that are feasible on the larger scale. For example, the comparability of aspects of additional diagnosis information has been assessed by interstate comparison of the distribution of AR-DRGs with and without complications and comorbidities (Coory & Cornes 2005); the quality of Indigenous status data has been gauged using a mixture of techniques such as patient reinterview study results and analysis of missing data and ratios of separations for Indigenous patients and other patients (AIHW 2005c); and the usefulness of the data for estimating the number of induced abortions in Australia has been examined by comparison of the data with state-based induced abortion registers (AIHW NPSU 2005).

The vast quantity of data contained in the database, covering the wide range of admitted patient activity in Australia, means that detailed knowledge of the quality of all aspects of the data does not exist, and there will be no one source of such information. Efforts at the national level to understand and improve the quality of the data (such as through classification development and other national data development activities) need to be supplemented and informed by investigations and activities at the local level. Hence contributions at the local level like Nguyen and Reid’s paper, and papers such as those by Miller, Riley and Bramley, that are relevant to infrastructure developments for health information, are very welcome.

Finally, I want to add my congratulations to Associate Professor Johanna Westbrook for her election as an International Fellow of the American College of Medical Informatics.

Although many of Professor Westbrook’s recent publications have been concerned with clinicians’ use of online information retrieval systems and other research related to decision support, she has also made valuable contributions over a considerable period to Health Information Management relevant to the assessment of the quality of routinely collected data, and appropriate use of the data in epidemiological analyses. These contributions have included assessment of coding quality for burns (Alechna, Westbrook & Roberts 1998) and appropriate interpretation of complications data in hospital separations data (Westbrook, Rushworth & Rob 1994). They have also included use of Pharmaceutical Benefits Scheme data to assess effects of prescribing restrictions for antilister drugs (Westbrook, Duggan & McIntosh 2001) and analysis of hospital morbidity data to assess subpopulation use of diagnostic upper gastrointestinal tract endoscopy (Westbrook 2002), and to investigate...
the effects of the introduction of paediatric ear, nose and throat (ENT) surgery guidelines on rates of ENT surgery among young people (Rob et al. 2004).

I look forward to Professor Westbrook’s continuing leadership in the use of Australia’s routinely collected data for epidemiological analyses.

References


Jenny Hargreaves BSc (Hons), GradDipPopulation Health
Head, Hospitals and Mental Health Services Unit
Australian Institute of Health and Welfare
GPO Box 570
Canberra, ACT 2601
AUSTRALIA
Phone: +61 2 6244 1121
Facsimile: +61 2 6244 1121
Email: jenny.hargreaves@aihw.gov.au