

AN INFORMATION-RICH ENVIRONMENT: LINKED-RECORD SYSTEMS AND DATA QUALITY IN CANADA

Leslie L. Roos, Ruth-Ann Soodeen and
Laurel Jebamani¹

ABSTRACT

Research programs building on population-based longitudinal administrative data and record linkage techniques are found in Oxford, Scotland, Rochester (the Mayo Clinic), Western Australia and Canada. Such systems can markedly expand both methodological and substantive research in health and health care.

Here we summarize published Canadian data quality studies regarding registries, hospital discharge abstract, prescription drugs, and physician claims. Suggestions for improving registries facilitate linkage and expand research into social epidemiology. New trends in case identification and health status measurement using administrative data are noted. The differing needs for data quality research in each province are highlighted.

1. INTRODUCTION

“Entry points” for science are new techniques or concepts which open a field and lead to a number of further advances.(Weiner, 1999) Two “entry points” for the study of health and health care are longitudinal administrative data from defined populations and record linkage techniques; they have been critical in developing information-rich environments.(Roos and Roos, 2001) Several Canadian provinces have developed the pieces – registries, claims files, and so on – necessary to produce an information-rich research environment building on population-based linked record systems.

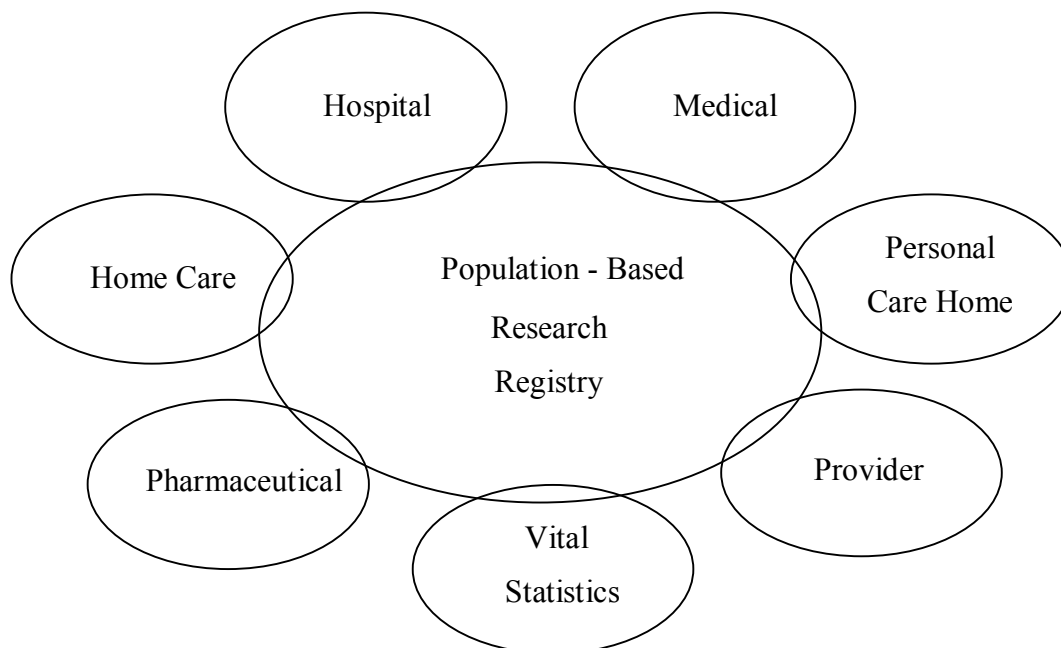
Linked-record systems depend on the quality of their components. Comparisons among files with administrative data, surveys, and clinical information to assure quality is aided by both record linkage and longitudinal data.

This presentation will review the most important data sets outlined in Figure 1: registries, hospital discharge abstracts, physician visits, and prescription drugs. We explore two approaches to establishing data quality:

- 1) The first approach links two or more sources of already available information and compares their extent of agreement on key items. Often, both sources are taken from administrative data (most commonly, hospital abstracts and physician claims). The process of linkage also builds knowledge about how many individual records link appropriately; this is important for assessing overall system quality.
- 2) The second approach, usually based on reabstraction of hospital charts or physician records, expends considerable resources to measure the accuracy of the information used.(Potvin and Champagne, 1986) In addition to assessing transcription errors, physical records may be scanned looking for more test results or chart entries.

¹ Manitoba Centre for Health Policy
Department of Community Health Sciences, Faculty of Medicine
University of Manitoba, 4th Floor John Buhler Research Centre
408 – 727 McDermot Avenue
Winnipeg, Manitoba R3E 3P5 Canada

Fig. 1 An Ideal Administrative Database



2. REGISTRY

As suggested in Figure 1, a registry based on a defined population is central to a linked-record system; many studies are designed to use a registry to help track individuals across files. Registry completeness is particularly critical for the study of defined populations.

Almost all quality analyses of provincial registries have been cross-sectional, using checks through linkage with another source (census data, Vital Statistics, or surveys). Eighteen such checks have been reported in the published literature. Databases are usually compared as to the number of 'individuals' on both files. Reported levels of agreement have been high. However, such comparisons are only part of the story. Specifying individual mortality and place of residence (in-province or out-of-province) is critical for many kinds of research. Accurate data on place of residence are important for assessing health status and health care utilization for different populations and for longer-term cohort studies. Such checks on mortality and residence have characteristically not been done; only one study (conducted in Manitoba) has been reported using primary data collection to analyze registry quality.

3. HOSPITALS AND PHYSICIANS

Linkage-based studies of the quality of hospital abstracts and physician have been carried out in several provinces, often using large numbers of records, procedures, and diagnoses. High rates of agreement have been reported for demographic information and for procedures; diagnostic agreement is typically somewhat lower. The linkage approach is versatile: hospital data quality studies have involved linkages to mental health data, Vital Statistics, and financial data, in addition to physician claims. Nine linkage-based studies based on hospital discharge abstracts were reviewed. Physician claims have also been linked with pharmaceutical data, disease registries, surveys, and laboratory records. One problem surfaced several times in this work. When two files

use somewhat different coding schemes, the level of aggregation of the data – the degree to which a code is highly specific – becomes an issue in these data quality studies.

Four Saskatchewan hospital studies show generally similar results when linkage-based and reabstraction data can be compared within the same paper. (Rawson and Malcolm, 1995a; Rawson and Malcolm, 1995b; Edouard and Rawson, 1996; Rawson et al., 1997) Another nine hospital reabstraction studies were also analyzed. Eight studies using physician claims were about equally divided between use of the linkage and the reabstraction methodologies. Results regarding the quality of information on demographics, procedures, and diagnoses were generally similar to those found in studies using hospital discharge abstracts.

The possibilities for disagreement in reabstraction studies are limited; formal reliability relates to these differences in transcription and coding. The comparisons made in linkage studies (factual reliability) show more possibilities for disagreement. Transcription errors might occur in the coding of both discharge abstracts and physician claims; short-term change in patient condition as well as physician memory could affect the precise diagnoses coded in either data set. Linkage-based comparisons seem to provide conservative estimates of the reliability of the data, since change in patient condition could be a legitimate reason for differences between two types of files.

This analysis suggests a practical strategy for researchers: check the quality of hospital and physician data using linkage-based methods whenever possible. If necessary, problems highlighted by linkage could be further explored using reabstraction. Such a strategy would markedly reduce reabstraction and other primary data collection.

4. PHARMACEUTICAL DATA

Five studies of pharmaceutical data indicate generally high levels of completeness and accuracy, although Canadian pharmaceutical coverage typically extends over only a select portion of the population. Only Saskatchewan and Manitoba have prescription databases relevant for almost the entire population. (Miller et al., 1996)

The provincial pharmaceutical databases do have certain limitations. Calculating dosage level may not be possible. Drugs dispensed in hospitals, nursing homes, or outpatient clinics have not been recorded in provincial databases. Although diagnostic information has been lacking in prescription claims databases, such data linked with physician claims (which typically provide at least one diagnosis) are beginning to be used for case identification.

5. EXTENDING THE REACH

Extending the reach of linked-record systems involves using utilization data to identify particular conditions relating to health, not just health care. (Roos and Roos, 2001; Virnig and McBean, 2001) Sometimes, a particular condition may be reliably identified using just hospital data; with the proper algorithms, acute myocardial infarction seems to be such a condition. (Cox et al., 1997; Levy et al., 1999) In four studies directed towards studying other conditions, a construct validity approach using more than one data source (physician claims, hospital abstracts) or more than one point in time was used to best approximate the prevalence identified from surveys or physical examinations. More specifically, linking surveys and laboratory measurements with administrative data has provided independent case ascertainment for chronic conditions (such as diabetes and hypertension) which typically involve contact with the health care system over a period of time. (Robinson et al., 1997; Muhajarine et al., 1997; Hebert et al., 1999) Broader measures of health status are also being developed. For example, a contact-based measure of individual health status, the Adjusted Clinical Group (ACG) case-mix system, has been able to build on diagnostic information available in two provinces. The high correlation between the ACG measure (aggregated to small areas) and premature mortality, the best small area measure of health status, suggests considerable construct validity. (Reid et al., 2001a and 2001b)

6. WHAT HAS BEEN DONE, WHAT IS NEEDED

Canada shows considerable heterogeneity by province in development of its health care databases (Table 1). A similar pattern of diffusion of linked-record systems is seen in one or more states in Australia. Such patterns have resulted from dedicated “entrepreneurship” by one or more individuals over a period of years, as well as from varying degrees of cooperation between academia and government.

Registries are not as immediately recognizable as critical for research on health and health care as are substantive files dealing with hospitals, physicians, and so forth. Registries point toward long-term studies, towards epidemiology rather than towards some types of health services research. The lack of research on the quality of follow-up and place of residence information is a critical problem for Canadian researchers.

New studies will also build on the considerable progress made in using multiple files for case identification and health status measurement. On the other hand, the strength of agreement among claims, self-reports, and clinical measurements has been found to vary significantly by age, occupation, obesity, and health status. Attention may need to be paid to such differences.

In the future, the growing field of social epidemiology is likely to build on improvements in registry data. (Berkman and Kawachi, 2000) Place of residence is an important indicator of social status which can be critical for such studies. Provincial and national cost data have been estimated at the micro-level and are being incorporated in ongoing research. Home care data are being evaluated and organized. Efforts to monitor adult, as well as childhood, immunizations have been initiated in Manitoba. Major efforts at linking administrative and clinical data are underway in Quebec.

A necessary concern for measurement must be traded off against the great economies of effort, the ability to consider entire populations, and the large number of cases available with administrative data. Because of the multiplier effects operative as more data sets become available for analysis, understanding the quality of these data sets is of great importance. Although population-oriented research can tolerate a higher degree of error than can work more tightly focused on the individual, the information-rich environments being created have certain limitations. Awareness of these can help us maintain and improve a provincial and national research capability which should facilitate insights across a number of fields.

Table 1
Data Quality Checks in Five Provinces – An Overview

	Registry	Hospital	Prescription	Physician
British Columbia	need checks on follow-up and residential identifiers	checks on diagnoses and procedures weak	not available	not available
Saskatchewan	not available; need publication of checks performed	checks made on diagnoses and procedures	some substantive checks	checks made on diagnoses and procedures
Manitoba	need checks on residential identifiers	checks made on diagnoses and procedures	checks made on claims filed; some substantive checks	checks made on diagnoses and procedures
Ontario	not available	checks made on diagnoses and procedures	not available	checks made on procedures
Quebec	not available	checks made on diagnoses and procedures	checks made on claims filed	checks made on diagnoses and procedures

This overview reflects published studies as of June 2001

ACKNOWLEDGEMENTS

This research was supported by the Institute of Health Economics, Edmonton, Canada, by the Manitoba Centre for Health Policy and Evaluation, and by a National Health Research and Development Program Career Scientist Award to Dr. L.L. Roos. The authors gratefully acknowledge the help of the St. Boniface General Hospital Research Centre.

REFERENCES

- Berkman, LF. Kawachi, I. (eds). (2000), *Social Epidemiology*, New York,: Oxford University Press.
- Cox, JL. Melady, MP. Chen, E. Naylor, CD. (1997), "Towards improved coding of acute myocardial infarction in hospital discharge abstracts: A pilot project", *Can J Cardiol*, 13(4),pp. 351-358.
- Edouard, L. and Rawson NSB. (1996), "Reliability of the recording of hysterectomy in the Saskatchewan health care system", *Br J Obstet Gynecol*, 103(9), pp. 891-897.
- Hebert, PL. Geiss, LS. Tierney, EF. Engलगau, MM. Yawn, BP. McBean, AM. (1999), "Identifying persons with diabetes using medicare claims data", *Am J Med Qual*, 14(6), pp. 270-277.
- Levy, AR. Tamblyn, RM. Fitchett, D. McLeod, PJ. Hanley, JA. (1999), "Coding accuracy of hospital discharge data for elderly survivors of myocardial infarction", *Can J Cardiol*, 15(11), pp. 1277-1282.
- Miller, E. Blatman, B. Einarson, TR. (1996), "A survey of population-based drug databases in Canada", *Can Med Assoc J*, 154(12), pp. 1855-1864.
- Muhajarine, N. Mustard, CA. Roos, LL. Young, TK. Gelskey, DE. (1997), "Comparison of survey data and physician claims data for detecting hypertension", *J Clin Epidemiol*, 50(6), pp. 711-718.

- Potvin, L. and Champagne, F. (1986), "Utilization of administrative data files in health research", *Social Indicators Research*, 18, pp. 409-423.
- Rawson, NSB. and Malcolm, E. (1995), *Validity of the Recording of Cholecystectomy and Hysterectomy in the Saskatchewan Health Care Datafiles. Technical Report Series (Report #3)*. Saskatoon, SK: Pharmacoepidemiology Research Unit, Technical report series.
- Rawson, NSB. and Malcolm, E. (1995), "Validity of the recording of ischaemic heart disease and chronic obstructive pulmonary disease in the Saskatchewan health care datafiles", *Stat Med*, 14(24), pp. 2627-2643.
- Rawson, NSB. Malcolm, E. D'Arcy, C. (1997), "Reliability of the recording of schizophrenia and depressive disorder in the Saskatchewan health care datafiles," *Soc Psychiatry Psychiatr Epidemiol*, 32(4), pp. 191-199.
- Reid, RJ. MacWilliam, L. Verhulst, L. Roos, NP. Atkinson, M. (2001) "Performance of the ACG Case-mix System in two Canadian provinces", *Med Care*, 39(1), pp. 86-99.
- Reid, RJ. Roos, NP. MacWilliam, L. Frohlich, N. Black C. (2001). "Assessing population need using a claims-based ACG morbidity measure: A validation analysis in the province of Manitoba.", (Unpub).
- Robinson, JR. Young, TK. Roos, LL. Gelskey, DE. (1997), "Estimating the burden of disease: Comparing administrative data and self-reports", *Med Care*, 35(9), pp. 932-947.
- Roos, LL. and Roos, NP. (2001), "Of space and time, of health care and health", *J Health Serv Res Pol*, 6(2), pp. 120-122.
- Virnig, BA and McBean, M. (2001), "Administrative data for public health surveillance and planning", *Annu Rev Public Health*, 22, pp. 213-230.
- Weiner, J. (1999), *Time, Love, Memory: A Great Biologist and His Quest for the Origins of Behavior*, New York: Alfred A. Knopf.