Working More Productively: Tools for Administrative Data

*Leslie L. Roos, Ruth-Ann Soodeen, Ruth Bond,
and Charles Burchill*

Objective. This paper describes a web-based resource (<http://www.umanitoba.ca/centres/mchp/concept/>) that contains a series of tools for working with administrative data. This work in knowledge management represents an effort to document, find, and transfer concepts and techniques, both within the local research group and to a more broadly defined user community. Concepts and associated computer programs are made as “modular” as possible to facilitate easy transfer from one project to another.

Study Setting/Data Sources. Tools to work with a registry, longitudinal administrative data, and special files (survey and clinical) from the Province of Manitoba, Canada in the 1990–2003 period.

Data Collection. Literature review and analyses of web site utilization were used to generate the findings.

Principal Findings. The Internet-based Concept Dictionary and SAS macros developed in Manitoba are being used in a growing number of research centers. Nearly 32,000 hits from more than 10,200 hosts in a recent month demonstrate broad interest in the Concept Dictionary.

Conclusions. The tools, taken together, make up a knowledge repository and research production system that aid local work and have great potential internationally. Modular software provides considerable efficiency. The merging of documentation and researcher-to-researcher dissemination keeps costs manageable.

Key Words. Record linkage, information systems, Internet, administrative databases, longitudinal

Managing research without stifling creativity has been an issue for a number of years (Brooks 1995; Kidder 2000). In health services research and epidemiology, the growth of large administrative databases and the teamwork required for their use pose significant management challenges. How can we get some economies of scale in analyses of these data? Working with databases is aided by the development of well-documented concepts and modular software to allow easy transfer from one project to another. This paper forwards web-based tools, including a Concept Dictionary and a series of SAS modules, developed at the Manitoba Centre for Health Policy (MCHP) for facilitating research production. Some of the SAS modules are freestanding.

Others have been designed to fit into larger systems, in particular LINKS (record linkage software) and RATES (for analyzing population-based information). Although these tools were developed specifically for researchers analyzing Manitoba data, much of the content can be easily adapted for use with administrative data available elsewhere.

Such tools promise to be particularly useful for the complex linked-record systems of the Canadian provinces, Western Australia, Scotland, and Oxfordshire in the United Kingdom (Holman et al. 1999; Goldacre et al. 2002; Kendrick et al. 1998; Chamberlayne et al. 1998; Roos and Roos 2001). Obvious applications in the United States include federal data with individual identifiers: Veterans Affairs, Medicare, and Medicaid files, and the National Death Index (Fleming et al. 1992; Mitchell et al. 1994; Virnig and McBean 2001; Howe 1998). Linkage has enriched a number of American files—tumor registries, vital records, and hospital discharge abstracts for newborns (Cooper et al. 1999; Perkins et al. 2001; Gilbert, Nesbitt, and Danielsen 1999). The Hopkins Adjusted Clinical Group (ACG) morbidity measure has been applied in conjunction with individual identifiers to account for comorbidities in analyses of private and public sector health care in both the United States and Canada (Starfield et al. 1991; Reid et al. 2002). There is considerable scope for using these tools to amplify research opportunities, but privacy issues may restrict the number of datasets having individually identifiable information (Greenfield 1996; Starr 1997; Gould 1999; Friedman et al. 2001; United States General Accounting Office 2001).

ADMINISTRATIVE DATA

Many analyses of administrative databases rely on one or two files. Documentation and software are increasingly important as files become more numerous and interconnected. Research centers, diverse projects, and multi-investigator studies should benefit from tools to facilitate productive collaboration. The two key tools described in this paper, the Concept Dictionary Resource and a series of SAS modules, have proven especially valuable for analyzing an administrative database in Manitoba, Canada, with a

Address correspondence to Leslie L. Roos, Ph.D., Manitoba Centre for Health Policy, Department of Community Health Sciences, 4th Floor Brodie Centre, Room 408-727 McDermot Avenue, University of Manitoba, Winnipeg, Manitoba, Canada, R3E 3P5. Ruth-Ann Soodeen, M.Sc., Ruth Bond, M.A., and Charles Burchill, M.Sc., are with the Manitoba Centre for Health Policy as well.

population-based research registry playing a central role (the inner ring of Figure 1). The files—primarily acquired from Manitoba Health—are longitudinal and linkable as needed for each study. The data are available to Manitoba researchers after ethical and privacy reviews.

The research registry developed by the Manitoba Centre for Health Policy has integrated information from various sources in a convenient format, providing certain variables for any date since 1970. Specifically, the Manitoba registry contains for each resident: an encrypted number, demographic characteristics, place of residence (a six-digit postal code), and family composition. While patient identifiers such as name and address have been removed from the registry maintained by Manitoba Health, the capacity to link records together to form individual histories of health care use has been preserved. Time-sensitive data elements (place of residence, family composition) are updated using “snapshot” registries provided by Manitoba Health every six months. Vital Statistics files are incorporated into the registry annually (Roos and Nicol 1999; Roos and Roos 2001).

Population-based registries of this type can provide information on all residents in an area, as well as their dates of arrival and departure (e.g., births, deaths, and moves). Thus, individuals can be followed through time and space. Each substantive file from Manitoba Health (included in the inner ring of Figure 1) can be checked against the registry for accuracy of the identifiers and particular information (for example, date of in-hospital death). Linkages among and within databases follow a clear set of rules. Such Canadian databases can be appropriately aggregated at the level of the individual, the physician, the hospital, or the population (Roos, Walld et al. 1996; Roos and Nicol 1999).

The outer ring of studies in Figure 1 represents special files—both clinical and survey data—used for various projects (Bernstein et al. 2001; Robinson et al. 1997; Murray et al. 2002; Mustard et al. 1999). Increasingly, primary data are being collected with record linkage in mind. Such files—and the ways in which they may be interrelated—lend both great potential and considerable complexity to the research enterprise.

WEB-BASED RESEARCH RESOURCES

The tools highlighted in this paper are accessible from the Research Resources page on MCHP’s web site (<http://www.umanitoba.ca/centres/mchp/concept>). In keeping with the dynamic nature of both research and the Internet,

Figure 1: The Manitoba Population Health Research Data Repository and Linkable Databases

the content and organization of these resources are modified regularly to maximize their relevance to various user groups.

Concept Dictionary

The Concept Dictionary was developed as a centralized repository that expands as MCHP's working knowledge requires documentation (Davenport and Prusak 1997; Burchill et al. 2000). Related resources include a glossary, medical/research definitions, a meta-index, and a protocol for project management. The dictionary operationally defines key ideas or terms used by health researchers, describing "in detail the development of new variables or creation of variables based on existing data" (Burchill et al. 2000). Most of the concepts derive from discussions with programmers and researchers about what they feel that their colleagues, both locally and externally, need to know for research continuity. As of this writing, 218 concepts have been documented, indexed, and linked.

The concepts contain answers to frequently asked questions, programming tips and cautions, write-ups of alternate formulations, and discussions of associated problems; analytic strategies are described as appropriate. When relevant, SAS programs containing the code to implement the concept are also included. The dictionary supports the software and works toward standard documentation. Content has been designed to provide ongoing operational assistance to researchers and programmers. Each description also includes hyperlinks to related documents and external web sites; for example, while the comorbidity concept focuses on the Charlson Index (used extensively in Manitoba research), a link to alternative software from the Agency for Healthcare Research and Quality (AHRQ) is also included. Similarly, the Adjusted Clinical Group (ACG) concept includes a link to AHRQ's Clinical Classification System. Finally, contact information for individuals who can provide more details about the particular research concept is provided. The Concept Dictionary makes as much information publicly available as possible via an external web site: (<http://www.umanitoba.ca/centres/mchp/concept/>).

A glossary, providing a shorter description of important terms (keyed to specific reports), supplements the dictionary; just over a thousand glossary terms and abbreviations were recently noted. The dictionary and glossary enhance research production in helping move previously developed concepts to new projects. The modular design has proven efficient in relaying information to analysts and students; concepts can be used as hyperlinks in other research aids (such as protocols) and in teaching materials (both site-specific courses and the Epidemiology Supercourse: <http://www.pitt.edu/~super1/>) (LaPorte et al. 2002). The Medical/Research Definitions section defines the codes for surgical procedures, tests, and diagnoses commonly analyzed at MCHP. The codes include those from classification systems such as ICD-9-CM, from physician tariff codes, and from common groupers like Diagnosis Related Grouper (DRG), Refined Diagnosis Related Grouper (RDRG), and Case Mix Grouper (CMG). Finally, a meta-index organizes concepts according to the Medical Sub-Heading (MeSH) system of the National Library of Medicine. Standardizing various concepts enhances collaboration among researchers within a single group and across centers using similar or identical technology (Kohler 1994). With information provided on an external web site, documentation and dissemination overlap to lower marginal costs.

Critical to effective use of the data is accessible documentation of how the research topics have been operationalized. The Concept Dictionary has aided Manitoba research on physician and hospital bed supply, access to care,

intensity of use, and other important topics (Roos, Black et al. 1996; Roos, Fransoo, Bogdanovic, Carriere et al. 1999; Roos, Fransoo, Bogdanovic, Friesen, and MacWilliam 1999; Brownell, Roos, and Roos 2001; Watson et al. 2003). The top half of Table 1 presents a sample of entries in the Concept Dictionary and related glossary terms. Two topics addressed extensively in the Concept Dictionary are costing and geographic units.

Costing. The lower portion of Table 1 shows the concepts and glossary entries associated with ongoing cost analyses. Since patients are not charged for hospital stays in Canada, tables of average costs per weighted case (with an annual inflation adjustment) were produced to permit estimating the dollars per day for inpatient hospital care and for day surgery procedures. The development of these case weights (based on Revised DRGs or their Canadian equivalents) has been described in detail elsewhere (Jacobs and Roos 1999; Jacobs et al. 2000). Total cost of care for an inpatient or day surgery patient includes all physician services received during the hospital stay. Hospital costs can be produced for each individual for any time period by adding up the dollars per services across the number of services.

Table 1: Concepts and Glossary Terms Relevant to Several Research Topics

<i>Research Topic</i>	<i>Related Concepts</i>	<i>Sample Glossary Terms</i>
Supply	Physician Service Areas Regional Health Authorities Bed Supply Hospital Types	Physician Supply Hospital Service Areas Physician Service Delivery Areas
Access to care	Distance to Hospital Regions of Manitoba	Access to Home Care Access to Hospital Services Access to Pharmaceutical Care Physician Services Region of Residence
Intensity of use	Intensity of Home Care Use Resource Intensity Weights	Intensity of Resource Use (IRU) Interqual Criteria
Costs per capita		
Dollars per service (visit, day of care)	Costing Hospital Care, Cost List	Costs of Care Costs per Weighted Case (CWC) Standard Cost List
Services per user	Hospital Claims Pharmaceutical Claims Physician Claims	Claims Hospital Abstracts Hospital Separation Service Codes Service Types - Physician Visits

Such cost tables have proven very useful (Shanahan et al. 1999; Jacobs et al. 2001). One paper examined Medical Savings Accounts, an effort “to reduce health care costs by transferring responsibility for expenditures to patients, while providing them with state-supported base amounts to cover some of the costs” (Forget, Deber, and Roos 2002, p. 143). This research summed physician bills and hospital costs (both covered by the Canada Health Act) to produce an estimated annual cost of care for each individual in the Manitoba population. The highest-using 1 percent of the Manitoba population accounted for 26 percent of all spending on hospital and physician care, whereas the lowest-using 50 percent accounted for 4 percent. Such skewed costs have highlighted the difficulty with Medical Savings Accounts: few individuals are in the “moderate cost” category where they could readily save health care costs by cutting back on system use.

Geographic Units. Geographic information provided by the Concept Dictionary has supported the calculation of physician and hospital bed supply, the construction of control groups, the estimation of socioeconomic categories, and the feedback of data to provincial Regional Health Authorities. Data on individual place of residence are generally available at the postal code or municipality levels and can be organized in several ways. In Manitoba, we obtain information at the lowest possible level of aggregation to facilitate development of denominators and to allow flexibility in “building” small areas up to larger areas. Canadian research centers typically provide a crosswalk between these residential codes and Statistics Canada census enumeration areas (the source of education, income, and employment information). Format statements are used—moving up from postal codes to define Regional Health Authorities, districts within these authorities, and hospital (or physician) service areas. This approach—in conjunction with maps (many of which are described in the Concept Dictionary)—has been particularly helpful for Regional Health Authorities (Roberts et al. 2002; Black et al. 1999). Similarly, flexibility in constructing areas has proven useful in drawing geographic controls for cohort studies (Kryger, Walld, and Manfreda 2002; Smith et al. 2002).

Protocol. In response to feedback from researchers, a protocol for working with administrative data was added to the knowledge repository (<http://www.umanitoba.ca/centres/mchp/protocol>). This protocol operationalizes issues of “bringing data together” which have been identified elsewhere (Buckeridge et al. 2002). Such intellectual middleware provides a methodo-

logical checklist or template, taking advantage of web capabilities to aid project management. Content of the protocol is organized under eight headings:

1. MCHP Background
2. Proposal Preparation
3. Approval Process
4. Data Preparation
5. Project Management
6. Data Analysis
7. Dissemination of Results
8. Project Completion

Efforts are being made to consistently apply the guidelines to all studies using administrative data at MCHP. The protocol recommends that documentation on study-specific resources be revisited for each project in case of changes to concepts, methodological criteria, or analytical resources.

Of considerable general interest, the Data Preparation section includes checklists for defining the study population and suggestions for variable construction. Checklists regarding data cleaning and eligibility have been developed for the most frequently used databases (hospital abstracts and physician claims). Information on the various files in the inner ring of Figure 1 is conveniently accessible from the Data Preparation section.

The study titled "Changes in Health and Health Care Use of Manitobans: 1985–1998" (noted as "14-Year trends" on the menu) provides an example of the protocol in use (Lix et al. 2002). The Study Population entry is divided into four categories: data sources, study period, unit of analysis, and eligible population. The entries under "Data Preparation: record eligibility" are presented in Table 2. Additional details are provided under "Data Preparation: data cleaning." Exclusions are highlighted under the appropriate headings. Finally, the "Analysis" entry notes that rates were age- and sex-adjusted using the direct method, with eleven age groups and 1998 as the standard population.

The protocol addresses major issues such as how to count different types of occurrences. For example, the Data Analysis section of the protocol discusses counting events given: multiple data fields; multiple events, same record; and multiple records, same diagnosis. The "multiple records, same diagnosis" notation is made because some diagnoses/procedures get coded twice even though they actually occur only once. For example, transfer from a rural hospital to a teaching hospital for angiography may lead to both hospitals coding the procedure (Roos, Walld et al. 1996).

Table 2: Data Preparation: Record Eligibility

Event Definition

Events were broadly defined (i.e., discharges, physician visits); it was not necessary to deal with ICD-9-CM definitions, for example, other than for analyses for several high-profile procedures.

Counting Events

Again, events were broadly defined, and issues of multiple data fields, etc., were not applicable for most analyses (except for the high-profile procedures).

Counting Days

In-year LOS (length of stay), but only within a given fiscal year, was counted. (Note: a preferred method is partitioning across years, which means going to adjacent data years to pick up stays overlapping two or more fiscal years). Day surgery was counted as one day. (Note: Again, this is not a usual way of defining hospital days. While day surgery is typically counted as an admission, it is not typically counted as one hospital day).

Type of Facility

For hospital abstracts, only acute care hospitals (Manitoba) were kept (i.e., excluded were personal care homes, chronic/rehabilitation facilities, and nursing stations).

Type of Care

For hospital abstracts, nonsurgical outpatient records were removed. For physician claims, nonambulatory visits were removed. For personal care home data, levels (of care) 1–4 were kept for analysis. For pharmaceutical data, records with drugs that were not part of the master formulary were removed.

Counting days can also be complicated. Links have been included for the following concepts:

- In-Year Days—for hospital inpatients, with SAS code
- Conservable Bed Days—for day surgery
- Expected Length of Stay—for personal care homes
- ICU Days—to calculate LOS (length of stay) for intensive care units

Having more than one programmer, one database, or one research site highlights differences in datasets and gaps in documentation. Without a protocol, complicated studies are likely to suffer from the four or five repeated runs at different sites, as experienced in one earlier collaborative project (Romano et al. 1994). When research centers produce anonymized datasets for outside investigators (often analyzing linked data in the outer ring of Figure 1), protocols are especially useful. Without such advice, Canadian researchers have been prone to make inconsistent decisions in, for example, separating inpatient from outpatient hospitalizations and acute from extended care stays. Changing Ministry of Health styles of coding transfers and planned readmissions generate intricacies in longitudinal analyses that call for concept dictionaries and protocols. Use of the protocol has increased since it was first added to the external web site in 2001; the count for March 2003 was 717 hits from 137 hosts.

SAS Macros

The Concept Dictionary describes the SAS macros developed for repeated use (SAS Institute 1999). The macros build on SAS's data management and analytic capabilities, integrating seamlessly with our larger SAS-based system. These macros are distributed to external users by request (rather than by automatic download) to keep track of outside interest for funding purposes.

Table 3 classifies some of the most popular macros according to their functions. The macros noted under "Analysis" provide agreement statistics (agreetab, written by Cancer Care Manitoba), aid in working with individual longitudinal data (build, combine, history), and calculate mortality-related outcomes (life, pyll, premort). Those under "Measures" use different datasets to generate measures of comorbidity (charlsn) and of continuity of care (concare) for individuals, specify episodes of hospital care (episode), and characterize census enumeration areas by wealth (quint) and by scores on a Socio-Economic Factor Index (sefi). The Socio-Economic Factor Index was developed from a principal component analysis of Canadian Census data. The

Table 3: Selected SAS Macros Developed by the Manitoba Centre for Health Policy

<i>Macro Category and Name</i>	<i>Summary Description</i>
Analysis	
_agreetab	Creates two-by-two tables with several agreement statistics*
_build	Summarizes output from the _combine macro
_combine	Tracks individual records for two related events
_history	Creates individual histories from multiple record files
_life	Calculates life expectancy
_pyll	Calculates potential years of life lost by death
_premort	Calculates premature mortality rates (premature mortality defined as death before age 75)
Measurement	
_charlsn	Generates Charlson Comorbidity Index scores from hospital discharge abstracts
_concare	Generates continuity of care measures from physician claims
_episode	Marks episodes of hospital care
_quint	Generates income quintiles by geographical area from census data on average household income
_sefi	Generates Socio-Economic Factor Index scores by geographical area from census data on education, unemployment, single parent households, etc.

*Developed by CancerCare Manitoba.

components were labor force participation of women, age dependency ratio (the population > 65 divided by the population 15–64), percent single parent households, percent female single parent households, and two aggregated factors of “unemployment” and “education” (Martens et al. 2002).

Links. LINKS is a set of SAS macros designed to perform record linkage, the bringing together of information from two independent source records believed to relate to the same individual or family (Acheson 1967). To conduct such linkage, a set of identifiers on each individual record in file A are run against a similar set of identifiers on each record in file B. The successfully “linked” records can be subsequently treated as a single record for one individual or family. The technical issues of record linkage (particularly the challenges of working with measurement error and marginal information) have been explored in depth elsewhere (Fellegi and Sunter 1969; Newcombe 1988). Record linkage has received considerable attention in epidemiology and health services research, facilitating studies of many types in several different countries (Howe 1998; Whiteman et al. 2000).

Record linkage has been critical for developing the previously noted systems in Canada, Western Australia, Scotland, and Oxfordshire. In Manitoba, linkage with external data may involve just mortality follow-up (e.g., the alcoholism survey) or more extensive linkage using both survey and clinical information (e.g., the sleep disorders research program) (Figure 1) (Murray et al. 2002). The sleep researchers have used a number of features of the database: longitudinal tracking of utilization and mortality among cohorts and controls, place of residence information to help generate appropriate controls, and costing of physician, hospital, and pharmaceutical utilization (Kryger, Walld, and Manfreda 2002; Smith et al. 2002).

Complicated software is often used in the United States, Australia, and Canada to link a comparatively small number of individuals with known risk factors (such as radiation exposure) with very large mortality databases (Howe 1998). LINKS is a straightforward, relatively simple program incorporating the features most common to record linkage. Two macros within LINKS (TESTPWR and TESTPKT) provide insight into the structure of the files being linked. Other macros aid in the overall linkage strategy, usually a combination of deterministic matching (identifier must agree on both files) and probabilistic matching (based on the degree to which agreement or disagreement on a given identifier argues for or against a linkage). Internal “housekeeping” (calculating the necessary weights to be given to agreement or disagreement) and resolving ties is automated (Roos

and Wajda 1991; Wajda et al. 1991). LINKS can use SAS features such as SOUNDEX, a name-matching algorithm that associates numbers with different groups of consonants to produce a code robust to variations in names that sound alike. Linkage and analysis in one computer run is often possible; minimizing the number of existing linked files has helped satisfy the confidentiality-related concerns of Manitoba oversight committees.

Rates. The RATES macro is designed to calculate rate of events per one thousand in several distinct subpopulations, adjusted for differences in those subpopulations. RATES reads a population file (the counts of individuals in a particular group or area in the province) and a data file (typically one of the files in the inner ring of Figure 1). Age distribution and sex are two of the most important differences between subgroups that influence event rates, but the macro is capable of handling others. RATES is particularly flexible for geographical analyses; the format statements noted earlier allow easy division of Manitoba into different areas. RATES computes the crude, the directly standardized, and the indirectly standardized rates for events. A confidence interval is calculated for each rate at a user-defined level. Comparisons of area rates between two time periods are made using the T^2 statistic (Carriere and Roos 1994). Rates are typically run against a standard population; the macro tests if rates for each area are the same and which rates differ significantly from the baseline comparison rate. Population data in the denominator often represent a “snapshot estimate” of the Manitoba population as of December 31 taken from the Manitoba Health registry, while the numerator generally consists of utilization data spread over a period of time. Adjustments for this are noted under the Population Denominator concept.

The Data Analysis section of the protocol (see “counting occurrences”) includes documentation of the RATES macro in a pdf file, as well as detailed discussions of related issues such as age, age grouping, and defining place of residence. The importance of such documentation was recently emphasized; one MCHP report calculating rates led to an embarrassing error when residence was not defined in the same way in both numerator and denominator. In Manitoba, this means both datasets (for numerator and for denominator) should use either postal code of residence or municipal code of residence.

DISCUSSION

The tools described here are important for both the timeliness and transparency of Manitoba research. Because the Manitoba hospital abstracts

and physician claims have been well-documented with costing information available, the analysis of Medical Savings Accounts could be quickly presented (and used) by two Canadian federally mandated commissions on health care (Standing Senate Committee on Social Affairs, Science, and Technology 2002; Romanow 2002). A “fast-tracked” academic study of this proposed financing arrangement was published four months after the first computer runs (Forget, Deber, and Roos 2002).

Transparency is significant in clearly specifying how each variable is measured and in outlining each step in the research process. This transparency is not only important for moving modules from one research project to another, but also critical for assessment and replication. Transparency has become an issue in Canada; a series of health indicators reports (produced under a federal/provincial/territorial agreement) has been criticized by provincial auditors because of a lack of documentation on quality assurance (Manitoba Health 2002).

Feedback from users (particularly students) has led to several suggestions about the protocol and project management. Students from outside Manitoba expressed an interest in a protocol with fewer links to “internal” files, which were not generally available. In response, a template or guide for project management is being developed. To the extent possible, students also wish to see actual project web sites to learn more about the decisions made in various circumstances.

The tools are not without limitations. Many of the SAS modules noted in Table 3 are designed for use specifically with files having individual identifiers. However, the Charlson Index, the mortality-related analytical modules, and the RATES macro are suitable for files both with and without individual identifiers. The applicability of the kinds of tools presented here go beyond administrative data. Complex models based on secondary analyses of surveys are becoming more popular (Veugelers, Yip, and Kephart 2001; Mackenbach 1998). The maintenance costs and the startup time associated with these surveys would be reduced by using tools like the Concept Dictionary.

Besides financial commitments (perhaps two full-time equivalent staff members over several years), the development of these tools has been aided by an effort to maintain a “moral economy” based on sharing (Kohler 1994). Co-operation has been facilitated by the ongoing usefulness of the tools in maintaining productivity. In one sense, a continuously updated Internet text for working with administrative data has been created. The monthly “hit rate” on the publicly available Concept Dictionary has climbed over the five years since its inauguration (almost 32,000 hits from more than 10,200 hosts in January 2003).

The choice of concepts and SAS modules for inclusion has been driven by needs of a single academic health care research center of significant size (65 staff members at the Manitoba Centre for Health Policy). The size of the group and the amount of data available have impressed upon researchers the desirability of documentation, communication, and cooperation. The pressure for producing timely deliverables and the opportunities for investigator-initiated research have heightened the need for management. The approach outlined here has been critical to such efforts.

The user community has recently become more broadly defined; Canadian funding agencies have noted the need for more standardization by researchers across the country. Some additional work on concepts related to costs is being undertaken by the Institute for Health Economics in Alberta, while three other Canadian research centers are developing concept dictionaries modeled after MCHP's Dictionary. Cooperative development of SAS modules with Monash University in Australia is also underway. These external stakeholders will provide input on our priorities and implementation efforts. Although MCHP's "from the bottom up" approach assures local relevance, it can also lead to lags in getting important new information into the directory. Thus, the macro for the Charlson Comorbidity Index has been distributed to 34 external researchers by request (Charlson et al. 1987; Romano, Roos, and Jollis 1993). At the same time, a better index by Elixhauser et al. (1998) has recently been incorporated into the dictionary.

Tool development is important in other fields, permitting researchers to ask new questions or approach old questions more efficiently (Dyson 1998). Tools facilitating particular research production systems can have long-term impacts. Thus, in the biological sciences innovative work has continued over almost a century, building on extensive knowledge of certain laboratory animals (e.g., *Drosophila*) for which documented, well-communicated production systems are in place (Kohler 1994; Weiner 1999). As new problems emerge, both long-term productivity and the ability to shift into new areas should be enhanced by an in-depth understanding of file content and the research production system. Administrative data, with their potential for better understanding health and health care, will continue to generate demands for such tools (Roos and Roos 2001).

ACKNOWLEDGMENTS

This research was supported by the Canadian Population Health Initiative, the Office of Health and the Information Highway (Health Canada), and the

Manitoba Centre for Health Policy. The results and conclusions are those of the authors and no official endorsement by Manitoba Health was intended or should be inferred. The authors thank Jo-Anne Baribeau and Phyllis Jivan for manuscript preparation. A portion of this work was presented at the Symposium on Health Data Linkage held in Sydney, Australia, March 20–21, 2002.

REFERENCES

- Acheson, E. D. 1967. *Medical Record Linkage*. London: Oxford University Press.
- Bernstein, C. N., A. Kraut, J. F. Blanchard, P. Rawsthorne, N. Yu, and R. Walld. 2001. "The Relationship between Inflammatory Bowel Disease and Socioeconomic Variables." *American Journal of Gastroenterology* 96 (7): 2117–25.
- Black, C. D., N. P. Roos, R. Fransoo, and P. Martens. 1999. *Comparative Indicators of Population Health and Health Care Use for Manitoba's Regional Health Authorities: A POPULIS Project*. Winnipeg: Manitoba Centre for Health Policy and Evaluation.
- Brooks, F. P. Jr. 1995. *The Mythical Man-Month: Essays on Software Engineering*. Boston: Addison-Wesley.
- Brownell, M., N. P. Roos, and L. L. Roos. 2001. "Monitoring Health Reform: A Report Card Approach." *Social Science and Medicine* 52 (5): 657–70.
- Buckeridge, D. L., R. Mason, A. Robertson, J. Frank, R. Glazier, L. Purdon, C. G. Amrhein, N. Chaudhuri, E. Fuller-Thomson, P. Gozdyra, D. Hulchanski, B. Moldofsky, M. Thompson and R. Wright. 2002. "Making Health Data Maps: A Case Study of a Community/University Research Collaboration." *Social Science and Medicine* 55 (7): 1189–206.
- Burchill, C., L. L. Roos, P. Fergusson, L. Jebamani, K. Turner, and S. Dueck. 2000. "Organizing the Present, Looking to the Future: An Online Knowledge Repository to Facilitate Collaboration." *Journal of Medical Internet Research* 2 (2): e10.
- Carriere, K. C., and L. L. Roos. 1994. "Comparing Standardized Rates of Events." *American Journal of Epidemiology* 140 (5): 472–82.
- Chamberlayne, R., B. Green, M. L. Barer, C. Hertzman, W. J. Lawrence, and S. B. Sheps. 1998. "Creating a Population-Based Linked Health Database: A New Resource for Health Services Research." *Canadian Journal of Public Health* 89 (4): 270–3.
- Charlson, M. E., P. Pompei, K. L. Ales, and C. R. McKenzie. 1987. "A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation." *Journal of Chronic Diseases* 40 (5): 373–83.
- Cooper, G. S., Z. Yuan, K. C. Stange, L. K. Dennis, S. B. Amini, and A. A. Rimm. 1999. "The Sensitivity of Medicare Claims Data for Case Ascertainment of Six Common Cancers." *Medical Care* 37 (5): 436–44.
- Davenport, T. H., and L. Prusak. 1997. *Working Knowledge: How Organizations Manage What They Know*. Boston: Harvard Business School Press.
- Dyson, F. 1998. *Imagined Worlds*. Cambridge: Harvard University Press.

- Elixhauser, A., C. Steiner, R. Harris, and R. M. Coffey. 1998. "Comorbidity Measures for Use with Administrative Data." *Medical Care* 36 (1): 8–27.
- Fellegi, I. P., and A. B. Sunter. 1969. "A Theory for Record Linkage." *Journal of the American Statistical Association* 64 (328): 1183–210.
- Fleming, C., E. S. Fisher, C. H. Chang, T. A. Bubolz, and D. J. Malenka. 1992. "Studying Outcomes and Hospital Utilization in the Elderly: The Advantages of a Merged Data Base for Medicare and Veterans Affairs Hospitals." *Medical Care* 30 (5): 377–91.
- Forget, E. L., R. B. Deber, and L. L. Roos. 2002. "Medical Savings Accounts: Will They Reduce Costs?" *Canadian Medical Association Journal* 167 (2): 143–7.
- Friedman, D. J., M. Anderka, J. W. Krieger, G. Land, and D. Solet. 2001. "Accessing Population Health Information through Interactive Systems: Lessons Learned and Future Directions." *Public Health Reports* 116 (2): 132–41.
- Gilbert, W. M., T. S. Nesbitt, and B. Danielsen. 1999. "Childbearing beyond Age 40: Pregnancy Outcome in 24,032 Cases." *Obstetrics and Gynecology* 93 (1): 9–14.
- Goldacre, M. J., M. Griffith, L. E. Gill, and A. Mackintosh. 2002. "In-Hospital Deaths as Fraction of all Deaths within 30 Days of Hospital Admission for Surgery: Analysis of Routine Statistics." *British Medical Journal* 324 (7345): 1069–70.
- Gould, J. B. 1999. "Vital Records for Quality Improvement." *Pediatrics* 103 (1, supplement E): 278–90.
- Greenfield, L. 1996. "Without Universal Coverage, Health Care Use Data Do Not Provide Population Health." *Milbank Quarterly* 74 (1): 33–6.
- Holman, C. D. J., A. J. Bass, I. L. Rouse, and M. S. T. Hobbs. 1999. "Population-Based Linkage of Health Records in Western Australia: Development of the Health Services Research Linked Database." *Australian New Zealand Journal of Public Health* 23 (5): 453–9.
- Howe, G. R. 1998. "Use of Computerized Record Linkage in Cohort Studies." *Epidemiologic Reviews* 20 (1): 112–21.
- Jacobs, P., L. Assiff, J. Bachynsky, J. -F. Baladi, C. K. Botz, M. Brown and Working Group members. 2000. *A National List of Provincial Costs for Health Care: Canada 1997/98*. Edmonton: The Institute of Health Economics. Accessed on June 13, 2003. Available at (<http://www.ihe.ca/costlist.cfm>).
- Jacobs, P., J. F. Blanchard, R. C. James, and N. Depew. 2001. "Excess Costs of Diabetes in the Aboriginal Population of Manitoba, Canada." *Canadian Journal of Public Health* 91 (4): 298–301.
- Jacobs, P., and N. P. Roos. 1999. "Standard Cost Lists for Health Care in Canada: Issues in Validity and Inter-Provincial Consolidation." *Pharmacoeconomics* 15 (6): 551–60.
- Kendrick, S. W., M. M. Douglas, D. Gardner, and D. Hucker. 1998. "Best-Link Matching of Scottish Health Data Sets." *Methods of Information in Medicine* 37 (1): 64–8.
- Kidder, T. 2000. *The Soul of a New Machine*. London: Little Brown.
- Kohler, R. E. 1994. *Lords of the Fly*. Chicago: University of Chicago Press.
- Kryger, M. H., R. Walld, and J. Manfreda. 2002. "Diagnoses Received by Narcolepsy Patients in the Year Prior to Diagnosis by a Sleep Specialist." *Sleep* 25 (1): 36–41.

- LaPorte, R. E., A. Sekikawa, E. Sa, F. Linkov, and M. Lovalekar. 2002. "Info-points: Whisking Research into the Classroom." *British Medical Journal* 324 (7329): 99.
- Lix, L., C. Newburn-Cook, N. P. Roos, and S. Derksen. 2002. "Trends in Health and Health Care Utilization in Manitoba." *Healthcare Management Forum* 15 (4, supplement): 35-38.
- Mackenbach, J. P. 1998. "Multilevel Ecoepidemiology and Parsimony." *Journal of Epidemiology and Community Health* 52 (10): 614-5.
- Manitoba Health. 2002. *Manitoba's Health Indicators Report*. Winnipeg: Manitoba Health.
- Martens, P., N. Frohlich, M. Brownell, K. C. Carriere, S. Derksen, L. MacWilliam, and T. Mayer. 2002. "Embedding Child Health within a Framework of Regional Health: Population Health Status and Sociodemographic Indicators." *Canadian Journal of Public Health* 93 (2, supplement): S15-20.
- Mitchell, J. B., T. A. Bubolz, J. E. Paul, C. L. Pashos, J. J. Escarce, L. H. Muhlbaier, J. M. Wiesman, W. W. Young, R. S. Epstein, and J. C. Javitt. 1994. "Using Medicare Claims for Outcomes Research." *Medical Care* 32 (7): 38-51.
- Murray, R. P., J. E. Connett, S. L. Tyas, R. Bond, O. Ekuma, C. K. Silversides, and G. E. Barnes. 2002. "Alcohol Volume, Drinking Pattern and Cardiovascular Morbidity and Mortality: Is There a U-Shaped Function?" *American Journal of Epidemiology* 155 (3): 242-8.
- Mustard, C. A., S. Derksen, J. -M. Berthelot, and M. C. Wolfson. 1999. "Assessing Ecologic Proxies for Household Income: A Comparison of Household and Neighbourhood-Level Income Measures in the Study of Population Health Status." *Health and Place* 5 (2): 157-71.
- Newcombe, H. B. 1988. *Handbook of Record Linkage*. New York: Oxford University Press.
- Perkins, C. I., W. E. Wright, M. Allen, S. J. Samuels, and P. S. Romano. 2001. "Breast Cancer Stage at Diagnosis in Relation to Duration of Medicaid Enrollment." *Medical Care* 39 (11): 1224-33.
- Reid, R. J., N. P. Roos, L. MacWilliam, N. Frohlich, and C. D. Black. 2002. "Assessing Population Health Care Need Using a Claims-Based ACG Morbidity Measure: A Validation Analysis in the Province of Manitoba." *Health Services Research* 37 (5): 1345-64.
- Roberts, J. D., R. Fransoo, C. D. Black, L. L. Roos, and P. Martens. 2002. "Research Meets Reality: Administrative Data to Guide Planning for Canadian Regional Health Authorities." *Healthcare Management Forum* 15 (4): 13-21.
- Robinson, J. R., T. K. Young, L. L. Roos, and D. E. Gelskey. 1997. "Estimating the Burden of Disease: Comparing Administrative Data and Self-Reports." *Medical Care* 35 (9): 932-47.
- Romano, P. S., L. L. Roos, and J. G. Jollis. 1993. "Adapting a Clinical Comorbidity Index for Use with ICD-9-CM Administrative Data: Differing Perspectives." *Journal of Clinical Epidemiology* 46 (10): 1075-9.
- Romano, P. S., L. L. Roos, H. S. Luft, J. G. Jollis, K. M. Doliszny and the Ischemic Heart Disease Patient Outcomes Research Team. 1994. "A Comparison of Admin-

- istrative Versus Clinical Data: Coronary Artery Bypass Surgery as an Example." *Journal of Clinical Epidemiology* 47 (3): 249–60.
- Romanow, R. J. 2002. *Commission on the Future of Health Care in Canada*. Ottawa, ON: Government of Canada.
- Roos, L. L., and J. P. Nicol. 1999. "A Research Registry: Uses, Development, and Accuracy." *Journal of Clinical Epidemiology* 52 (1): 39–47.
- Roos, L. L., and N. P. Roos. 2001. "Of Space and Time, of Health Care and Health." *Journal of Health Services Research Policy* 6 (2): 120–2.
- Roos, L. L., and A. Wajda. 1991. "Record Linkage Strategies: Part I. Estimating Information and Evaluating Approaches." *Methods of Information in Medicine* 30 (2): 117–23.
- Roos, L. L., R. Walld, A. Wajda, R. Bond, and K. Hartford. 1996. "Record Linkage Strategies, Outpatient Procedures, and Administrative Data." *Medical Care* 34 (6): 570–82.
- Roos, N. P., C. D. Black, N. Frohlich, C. DeCoster, M. M. Cohen, D. J. Tataryn, C. A. Mustard, L. L. Roos, F. Toll, K. C. Carriere, C. A. Burchill, L. MacWilliam, and B. Bogdanovic. 1996. "Population Health and Health Care Use: An Information System for Policy Makers." *Milbank Quarterly* 74 (1): 3–31.
- Roos, N. P., R. Fransoo, B. Bogdanovic, K. C. Carriere, N. Frohlich, D. Friesen, D. Patton, and R. Wall. 1999. "Needs-Based Planning for Generalist Physicians." *Medical Care* 37 (6, supplement): JS206–28.
- Roos, N. P., R. Fransoo, B. Bogdanovic, D. Friesen, and L. MacWilliam. 1999. "Issues in Planning for Specialist Physicians." *Medical Care* 37 (6, supplement): JS229–53.
- SAS Institute. 1999. *SAS.STAT User's Guide* (version 8). Cary, NC: SAS Institute.
- Shanahan, M., M. Loyd, N. P. Roos, and M. Brownell. 1999. "A Comparative Study of the Costliness of Manitoba Hospitals." *Medical Care* 37 (6, supplement): JS101–22.
- Smith, R., J. Ronald, K. Delaive, R. Walld, J. Manfreda, and M. H. Kryger. 2002. "What Are Obstructive Sleep Apnea Patients Being Treated for Prior to This Diagnosis?" *Chest* 121 (1): 164–72.
- Standing Senate Committee on Social Affairs, Science, and Technology. 2002. *The Health of Canadians—The Federal Role. Volume Six: Recommendations for Reform*. Ottawa, ON: The Senate of Canada.
- Starfield, B. H., J. P. Weiner, L. M. Mumford, and D. M. Steinwachs. 1991. "Ambulatory Care Groups: A Categorization of Diagnoses for Research and Management." *Health Services Research* 26 (1): 53–74.
- Starr, P. 1997. "Smart Technology, Stunted Policy: Developing Health Information Networks." *Health Affairs (Millwood)* 16 (3): 91–105.
- United States General Accounting Office. 2001. *Record Linkage and Privacy: Issues in Creating New Federal Research and Statistical Information (GAO-01-126SP)*. Washington, DC: United States General Accounting Office.
- Veugelers, P. J., A. M. Yip, and G. Kephart. 2001. "Proximate and Contextual Socioeconomic Determinants of Mortality: Multilevel Approaches in a Setting

- with Universal Health Care Coverage." *American Journal of Epidemiology* 154 (8): 725–32.
- Virnig, B. A., and M. McBean. 2001. "Administrative Data for Public Health Surveillance and Planning." *Annual Review of Public Health* 22: 213–30.
- Wajda, A., L. L. Roos, M. Layefsky, and J. A. Singleton. 1991. "Record Linkage Strategies: Part II, Portable Software and Deterministic Matching." *Methods of Information in Medicine* 30 (3): 210–4.
- Watson, D., B. Bogdanovic, P. Heppner, A. Katz, R. Reid, and N. P. Roos. 2003. "The Use of Physician Services by Older Adults: Temporal Trends 1991/92 to 2000/01." *Canadian Journal on Aging*. In press.
- Weiner, J. 1999. *Time, Love, Memory: A Great Biologist and His Quest for the Origins of Behavior*. New York: Alfred A. Knopf.
- Whiteman, D., M. Murphy, K. Hey, M. O'Donnell, and M. J. Goldacre. 2000. "Reproductive Factors, Subfertility, and Risk of Neural Tube Defects: A Case-Control Study Based on the Oxford Record Linkage Study Register." *American Journal of Epidemiology* 152 (9): 823–8.