

# Data Quality in an Information-Rich Environment: Canada as an Example

Leslie L. Roos,<sup>1</sup> Sumit Gupta,<sup>2</sup> Ruth-Ann Soodeen,<sup>1</sup> and Laurel Jebamani<sup>1</sup>

---

## RÉSUMÉ

Cette étude vise à évaluer la qualité des données administratives disponibles dans les provinces canadiennes, tout en mettant l'accent sur les renseignements nécessaires pour créer des systèmes intégrés. Nous comparons explicitement diverses approches en matière de mesure de la qualité, en indiquant dans quel cas le couplage des dossiers peut ou non se substituer à la méthode plus onéreuse de la seconde saisie des dossiers. Quarante-neuf études originales visant à évaluer les données administratives canadiennes (registres, résumés d'hospitalisation, demandes des médecins et médicaments sur ordonnance) sont résumées de manière structurée. Les registres, les résumés d'hospitalisation et les dossiers des médecins semblent généralement de qualité satisfaisante, bien qu'il reste beaucoup de travail à accomplir. La qualité des données n'a pas fait l'objet de variations systématiques entre les provinces. Des données primaires doivent être recueillies afin de vérifier les lieux de résidence et effectuer un suivi longitudinal dans les registres provinciaux. Les vérifications initiales des données pharmaceutiques se sont révélées prometteuses et doivent être poursuivies. Étant donné que les études fondées sur le couplage des dossiers étaient « prudentes » dans leurs conclusions en matière de fiabilité, la réduction du nombre de secondes saisies qui prennent beaucoup de temps semblerait faisable dans bien des cas. Enfin, il pourrait être possible d'étendre la portée des données administratives de manière à étudier l'état de santé, ainsi que les soins de santé pour certaines conditions chroniques. Le potentiel de recherche des milieux riches en informations qui sont en train d'être créés permet de souligner l'importance de la qualité des données.

**Les auteurs tiennent à remercier le centre de recherche de l'hôpital St. Boniface General pour l'aide fournie. Les auteurs remercient également Jo-Anne Baribeau et Phyllis Jivan pour la préparation du manuscrit. Une partie de cette étude a été présentée, en 2001, au symposium de Statistique Canada intitulé « 18th International Symposium on Methodological Issues and Workshops », tenu à Ottawa, du 16 au 19 octobre 2001.**

## ABSTRACT

This review evaluates the quality of available administrative data in the Canadian provinces, emphasizing the information needed to create integrated systems. We explicitly compare approaches to quality measurement, indicating where record linkage can and cannot substitute for more expensive record re-abstraction. Forty-nine original studies evaluating Canadian administrative data (registries, hospital abstracts, physician claims, and prescription drugs) are summarized in a structured manner. Registries, hospital abstracts, and physician files appear to be generally of satisfactory quality, though much work remains to be done. Data quality did not vary systematically among provinces. Primary data collection to check place of residence and longitudinal follow-up in provincial registries is needed. Promising initial checks of pharmaceutical data should be expanded. Because record linkage studies were "conservative" in reporting reliability, the reduction of time-consuming record re-abstraction appears feasible in many cases. Finally, expanding the scope of administrative data to study health, as well as health care, seems possible for some chronic conditions. The research potential of the information-rich environments being created highlights the importance of data quality.

---

1 Department of Community Health Sciences, University of Manitoba

2 Faculty of Medicine, University of Toronto

\* The authors gratefully acknowledge the help of the St. Boniface General Hospital Research Centre. The authors also thank Jo-Anne Baribeau and Phyllis Jivan for manuscript preparation. A portion of this work was presented at the Statistics Canada 2001 Symposium, 18th International Symposium on Methodological Issues and Workshops, held in Ottawa, October 16–19, 2001. This work was partially funded by the Institute for Health Economics, Edmonton, AB.

Manuscript received: / manuscrit reçu : 01/03/03

Manuscript accepted: / manuscrit accepté : 16/06/04

**Mots clés:** qualité des données; données administratives; utilisation des soins de santé; registre; résumés d'hospitalisation; demandes des médecins; vieillissement

**Keywords:** data quality; administrative data; health care utilisation; registry; hospital abstracts; physician claims; aging

Requests for offprints should be sent to: / Les demandes de tirés-à-part doivent être adressées à :

Leslie L. Roos, Ph.D.

Manitoba Centre for Health Policy  
Department of Community Health Sciences  
Faculty of Medicine, University of Manitoba  
4th Floor Brodie Centre, Room 408, 727 McDermot Avenue  
Winnipeg, MB R3E 3P5  
(Leslie\_Roos@cpe.umanitoba.ca)

## Introduction

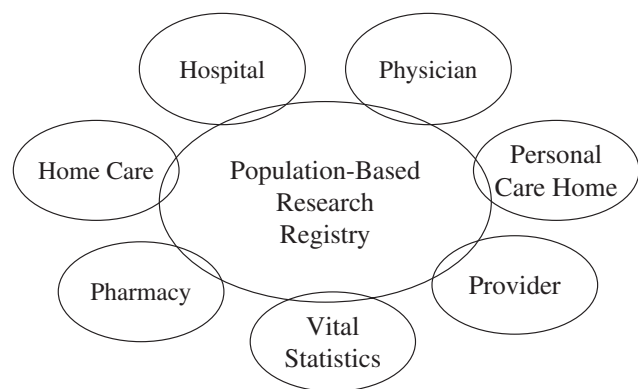
Population-based administrative data are currently used to facilitate innovative analyses of health and health care in a number of countries (Holman, Bass, Rouse, & Hobbs, 1999; Roos & Shapiro, 1999; Sheps et al., 2000). Canada's single-payer financing and several provincial initiatives have facilitated the creation of the pieces – registries, claims files, and so on – necessary to produce world-class information systems (Naylor, 1999).

Population registries can provide not only standard demographic data, but also location of residence, length of residence in a given location, marital status, and family size. Incorporating substantive files around such a central registry (Figure 1) can generate information for all people enrolled, with full protection of individual privacy and confidentiality. Negotiated relationships with other organizations providing clinical, educational, socio-economic, and family services data further magnify the research possibilities (Roos, Menec, & Currie, 2004).

Quality concerns are critical for health care assessment and for increased research productivity. The

Canadian Institute for Health Information (CIHI) and *Maclean's* have been using administrative data to rank health care across different jurisdictions over the last several years (Hawaleshka, 2002). Such efforts are valuable, but variation in reporting among provinces (and among institutions within a province) present problems here. In 2000, the Conference of Deputy Ministers defined a series of health indicators "to be regularly reported to Canadians"; many of these measures depend on information from registries and Vital Statistics (mortality and follow-up), from hospital discharge abstracts (diagnoses and procedures), and from physician claims (surveillance and some immunization reporting systems). Auditors in each province expressed concern about data quality; in Manitoba, the auditor was "unable to form an opinion on the accuracy of the data or on the adequacy of disclosure" vis-à-vis 21 of 56 health indicators used in the provincial report (Manitoba Health, 2002).

Assessing the quality of large administrative data sets is not easy. One reviewer has argued that validity can be assessed only by determining "if the diagnostic records themselves were correct, through an extensive review of the chart. Even then, it would not be the same as seeing the patient." Given such concerns, this paper refers to reliability – the agreement between data sources – throughout. We concentrate upon the quality of specific provincial databases, the databases that have accounted for the overwhelming majority of Canadian papers relying on administrative data. In the 1994–2002 period alone, 146 and 138 such publications using Manitoba and Ontario data, respectively, were identified (Roos, Menec, & Currie, 2004). This review evaluates the reliability of four important Canadian data sets: registries, hospital discharge abstracts, physician visits, and prescription drugs. Different approaches to measurement are compared, noting important strategic issues and making judgments about cost-effectiveness.



**Figure 1: An ideal administrative database**

## Method

### Data Collection

We searched for original papers and reports (up to the end of 2002) dealing with the quality of Canadian databases. The National Library of Medicine's *PubMed* was searched using the keywords *administrative data quality* and *Canadian administrative data*. Research centres in Manitoba and Ontario also provided lists of published papers. Senior authors of relevant papers were queried about studies that might have been missed. Forty-five papers relevant to data quality were located using this methodology. Because the British Columbia database provides anonymized data sets to researchers in Canada and elsewhere, project by project, a separate effort was made to find all papers published using administrative data from British Columbia in the 1994–2002 period. Thirty-one such studies were identified, starting from a list of investigators provided by the data holder (the Centre for Health Services and Policy Research at the University of British Columbia); papers by listed researchers and their co-investigators were located using *PubMed*. Four of these papers provided some additional information on data quality (Anderson & Kerluke, 1996; Dimich-Ward et al., 1996; Hertzman et al., 1997; Rankin et al., 1999).

Combining lists from research centres with searches using *PubMed* keywords is the most practical way to locate studies on health and health care, but important research can be missed. Many projects undertaken by government agencies are never published as academic papers. Most conference proceedings, book chapters, and certain journals (including the *Canadian Journal on Aging*) are not covered in *PubMed* (Roos, Menec, & Currie, 2004). Nonetheless, its broad coverage of peer-reviewed papers and their associated keywords makes *PubMed* indispensable for literature reviews. Some papers are no doubt missed, but considerable effort has been expended in data collection.

### Measuring Agreement

Administrative data are generally evaluated on the completeness of the data set (Are all residents included in the registry? Are all inpatients included in the hospital data set?) and the quality of the specific information recorded.

A gold standard is typically "a definitive diagnosis attained by biopsy, surgery, autopsy, long-term follow-up or other acknowledged standard" (Sackett, Haynes, Guyatt, & Tugwell, 1991). The best practical technology (hospital charts) is generally used to

measure diagnostic accuracy. For example, when computerized hospital discharge abstracts are compared with physician claims, the number of records with concordant information is divided by the total number of hospital records to calculate per cent agreement. Other investigators have treated hospital and physician data more equally, using each type of record as the standard and comparing the results (Iron, Goel, & Williams, 1995; Roos, Sharp, & Wajda, 1989).

The kappa measure adjusts for rare conditions by presenting the proportion of potential agreement between data sources beyond chance. A kappa of .0–.2 has been noted as "slight" agreement, .2–.4 = "fair," .4–.6 = "moderate," .6–.8 = "substantial," and .8–1.0 = "almost perfect" (Sackett et al., 1991). When cases with particular procedures or diagnoses are first identified for re-abstraction or linkage, kappa cannot be calculated because "true negative" cases – cases where both records are negative – are not available. Although per cent agreement can be calculated in such circumstances, the comparability of the data is unclear (Edouard & Rawson, 1996; Rawson & Malcolm, 1995a, 1995b; Rawson, Malcolm, & D'Arcy, 1997; Roos, Roos, Cageorge, & Nicol, 1982; Roos, Walld, Wajda, Bond, & Hartford, 1996). On the other hand, when relatively rare conditions are studied using a population or population-based sample, the high proportion of true negative cases (cases without the condition) will lead to an artificially high per cent agreement. This emphasizes the need to calculate kappa when possible, but only occasionally was sufficient information available to permit producing this statistic when the calculation was not done originally.

### Approaches

The two major approaches used to evaluate data quality are record linkage and re-abstraction.

Record linkage joins two or more separate sources of information to specify individuals present on both files (Acheson, 1967; Fellegi & Sunter, 1969; Howe, 1998; Newcombe, 1988; Wajda, Roos, Layefsky, & Singleton, 1991). When both files are supposed to contain the same individuals, the overall match gives an indication of completeness. For matched individuals, the degree of agreement between items on both records provides a measure of reliability. Although neither source may "faithfully represent an individual's state of health," the economies involved have popularized this technique (Potvin & Champagne, 1986).

Re-abstraction focuses on how reliably information (generally from hospital charts or physician records)

moves into computerized form (Potvin & Champagne, 1986). Costs associated with re-abstraction vary with the information needed, charges assessed by hospital record rooms, and so forth, but are almost invariably higher than those generated by record linkage.

## Results

### Registries

A registry based on a defined population provides flexibility in tracking both users and non-users of care, while accurate identifiers permit linkages with substantive files for reliability checks and ongoing research (Table 1) (Roos & Nicol, 1999). Only relatively small numbers of military personnel, the RCMP, and prison inmates are not included in provincial registries. Checks against a registry (and demographic data generally) are typically not subject to the "rare diagnosis" problem. For example, since birthdate is to be entered on each record, per cent agreement can be a useful statistic. Discrepancies in mortality information between registries and Vital Statistics of 1 or 2 per cent have been found in recent years (Hatcher & Hervas, 2001; Roos & Nicol, 1999; Wajda et al., 1991). British Columbia and Saskatchewan comparisons matching unique personal identifiers have reported between 96 and 99+ per cent agreement on individuals on the health registry and on substantive files (Chamberlayne et al., 1998; Risch & Howe, 1994). Capabilities to match by name proved useful in several complicated checks performed in British Columbia (Dimich-Ward et al., 1996; Hertzman et al., 1997).

In Manitoba, all available information (snapshot registries produced every 6 months, retroactive cancellations, deaths in hospitals and personal care homes) has been used to build a research registry. Checks have shown:

- Research registry counts by subgroup exceeded Winnipeg census figures, with differences diminishing over time and varying from 4.1 per cent (males in 1986) to 2.1 per cent (females in 1996).
- Differences were greatest for the age group with the greatest mobility: those aged 15 to 24.

The reliability of critical registry variables – place of residence (in-province or out-of-province) and migration (both out-of-province and in-province) – needs to be explored. Ministry mailings that update registries can help. Manitoba primary data collection found approximately 98 per cent of deaths and out-of-province moves (migration) to be recorded within 1 year after the event: British Columbia data suggest somewhat longer reporting lags in that province

(Hertzman et al., 1997; Roos & Nicol, 1999; Roos, Sharp, & Cohen, 1991; Wajda & Roos, 1987).

Other information on registry quality has focused less on time lags. Comparisons involving the Canadian Mortality Data Base have suggested relatively few moves out of British Columbia; for example, between four and five per cent of the deaths of sawmill workers who once lived in BC have been reported as occurring elsewhere in Canada (Hertzman et al., 1997). Analyses of emigration patterns of cancer cases in Alberta have provided estimates of error in both the health insurance plan registry and the cancer registry (Hatcher & Hervas, 2001).

Finally, reporting may require linking files organized at different times. One study relied on an alternative postal code (from "place of residence" a year before or after the index date) in about 7 per cent of the cases to provide high-quality matches (Houle et al., 1999; Roos & Nicol, 1999). About 12 per cent of Manitobans appear to move among postal codes annually, with up to half of these codes subject to change after late reporting.

### Hospital Discharges

With few explicit checks being done, researchers have generally assumed that hospital databases are essentially complete in recording inpatient stays. In one example using record linkage, maternal delivery records accounted for more than 99 per cent of Manitoba births from 1989 to 1991 (Mustard, Harman, Hall, & Derksen, 1995). In some circumstances, counts may be too large. Because procedures involving short-term transfers (typically done just in teaching hospitals) may be recorded in more than one hospital, double-counting of such procedures needs to be checked (Roos et al., 1996).

Four Saskatchewan studies combined record linkage and re-abstraction approaches, showing generally similar results vis-à-vis demographic and clinical information; the linkage-based studies occasionally produced less concordance (for example, in diagnoses accompanying hysterectomy) (Table 2) (Edouard & Rawson, 1996; Rawson & Malcolm, 1995a; Rawson et al., 1997; Rawson & Malcolm, 1995b).

Five stand-alone comparisons linked hospital and physician data, using large numbers of records, procedures, and diagnoses (table available from the senior author) (Iron et al., 1995; Roos et al., 1989). Agreement on the procedure undertaken – cholecystectomy, hysterectomy, coronary artery bypass surgery, coronary angiography, hip replacement, knee replacement, and so on – was generally well over 90 per cent (Iron et al., 1995; Rawson & Malcolm,

Table 1: Registry checks

Author Linkage-Based	Province	Records Examined	Years Examined	Cross-Sectional vs. Longitudinal	Identifiers Used	Agreement
Bryant & Brasher, 1995	Alberta	11,000 breast augmentation patients	1973–1990	Cross-sectional	Name, birth date, Alberta Health Care number, birth name	95%
Dimich-Ward et al., 1996	British Columbia	19,675 births (from 9,512 fathers)	1952–1988	Longitudinal	Unique birth registration number, birth date, gender, gestational age, birth weight	14 cases disagreed on both birth date and gender of child
Hertzman et al., 1997	British Columbia	26,487 sawmill workers	1950–1990	Longitudinal	Name, birth date, social insurance number (check with Statistics Canada)	Vital status not verifiable for 14% of cohort
Chamberlayne et al., 1998	British Columbia	700,000 hospital separations (discharges plus deaths); 50 million physician claims; 20,000 deaths; 40,000 births; 16,000 long-term care services recipients; 4.5 million elderly prescription drug recipients	1985–1995, 1985–1997, 1995–1997	Cross-sectional	Name, gender, birth date, postal code, Personal Health Number (PHN), Medical Services Plan (MSP) identity number	96%–99.8%
Cohen, 1993	Manitoba	30,000 records in a cervical cancer study	1982	Cross-sectional	Manitoba family registration number	95% (cases) 99.8% (controls)
Cohen et al., 1996	Manitoba	10,107 women enrolled in a breast screening study	1992	Cross-sectional	Registration number, birth date, initial, postal code	97%
Hatcher & Hervas, 2001	Alberta	83,446 cases from Cancer Registry	1985–1993	Longitudinal	Name, gender, birth date, PHN, Alberta Health Care Number, location (Alberta or left blank)	99%
Muhajarine et al., 1997	Manitoba	2,275 participants in the Heart Health Survey	1987–1990	Cross-sectional	Year of birth, gender, registration number	98%

Continued



**Table 1: Continued**

Risch & Howe, 1994	Saskatchewan	all women in Saskatchewan	1976–1990	Successive cross-sectional	Name, date of birth, registration beneficiary number	>99.99%
Robinson et al., 1997	Manitoba	2,792 participants in the Heart Health Survey	1990	Cross-sectional	Encrypted PHN	98%
Roos & Nicol, 1999	Manitoba	1,961 individuals with malignancies of the lung, colon, and breast (from Cancer Registry)	1990	Cross-sectional	Year and month of birth, gender, registration number, initial	98%
Roos & Wajda, 1991	Manitoba	8,000 deaths taken from vital statistics	1973, 1986, 1987	Cross-sectional	Registration number, initial, gender, year of birth	92%–99%
Roos et al., 1985	Manitoba	51,075 records from the Health Services Commission	1970–1982	Cross-sectional	Registration number, initial, gender, year of birth	91%
Houle et al., 1999	Manitoba	261,861 individuals in Canadian census	1986	Cross-sectional	Year of birth, gender, postal code, marital status, year of birth of spouse and children*	74%
Primary Data Collection						
Roos & Nicol, 1999	Manitoba	4,387 individuals hospitalized for acute myocardial infarction, hip fracture, and certain cancer conditions	1992–1995	Longitudinal	Registration number, manual lookup at Manitoba Health	99.8–100%

\*This is the only study that linked records without a family registration or personal identification number.

**Table 2: Hospital discharge abstracts, comparing linkage-based and re-abstracting approaches**

Author	Abstracts Examined (I = linked, r = re-abstracted)	Variable Examined	Agreement (Linkage with Physician Claims)	Agreement (Re-abstracting)
Edouard & Rawson, 1996	Hysterectomy (1,537 -l; 226 -r)	Diagnosis, procedure	56% (diagnosis) – 98% (type of hysterectomy)	85% (diagnosis) – 97% (type of hysterectomy)
Rawson & Malcolm, 1995a	Cholecystectomy (1,957 -l; 218 -r)	Diagnosis, procedure	6–71% (diagnosis) – 90–99% (procedure)	From 87% (diagnosis) to 100% (procedure)
Rawson & Malcolm, 1995b	Ischaemic heart disease and chronic obstructive pulmonary disease (9,537 -l, acute myocardial infarction, 224 -r; chronic airways obstruction, 225 -r)	Diagnosis	85–94% (ischaemic heart disease); 80–95% (chronic obstructive pulmonary disease)	97% (acute myocardial infarction); 94% (chronic airways obstruction);
Rawson et al., 1997	Depressive disorder and schizophrenia (1,474 -l; 281 -r)	Diagnosis	67–93% (depressive disorder); 62–89% (schizophrenia)	58–94% (depressive disorder); 77–94% (schizophrenia)

The same data on hysterectomy were used in the Edouard & Rawson (1996) and the Rawson & Malcolm (1995a) studies. Only the cholecystectomy data are presented under the latter study.

1995a; Roos et al., 1982; Roos et al., 1996). Vascular surgery, biliary tract operations other than cholecystectomy, and radical versus modified mastectomy showed lower agreement. Specificity of a code was sometimes an issue. For example, satisfactory agreement between Ontario hospital abstracts and physician claims could be produced only by aggregating various breast surgery procedures into an "other" category (Iron et al., 1995).

Analysis of diagnoses among individuals hospitalized for a particular reason (such as for a specific surgical procedure) has strengths and weaknesses. Such a convenient strategy produces a diagnostic mix unrepresentative of the hospitalized population as a whole. On the other hand, because the diagnoses studied (i.e., obstetrical diagnoses accompanying a hysterectomy) are generally not rare in the population studied, inflation of the per cent agreement measure is limited. One Manitoba study of 11 surgical procedures showed 75 per cent success in finding the one diagnosis on the physician claim (at the three-digit level of ICD-9-CM) for surgery among the up to 16 diagnoses on the discharge abstract (Roos et al., 1989). Diagnoses other than the primary or "most responsible" diagnosis show less accuracy. Chart reviews (from regional, inpatient, and outpatient services) of recipients of mental health services showed 76 per cent agreement on primary ICD-9-CM diagnosis at the five-digit level and 83 per cent on three-digit level diagnosis (Robinson & Tataryn, 1997). Co-morbid diagnoses agreed 55 per cent and 57 per cent at the five- and three-digit levels, respectively.

Twelve stand-alone re-abstraction studies (table available from the senior author) showed required demographic items to be quite accurate (Hawker, Coyte, Wright, Paul, & Bombardier, 1997; Rawson et al., 1997). Such items are supposed to be included on all charts, so the issue of "rare conditions" is not of concern. Slightly lower agreement levels (84% for date of birth and 86% for marital status) were found between charts and computerized records of mental health service recipients (Robinson & Tataryn, 1997).

The importance of a diagnosis affects reliability. Among 236 patients hospitalized with a hip fracture noted on the abstract, all but one were found to have this chart diagnosis: agreements of 95 per cent for asthma, 76 per cent for non-asthma respiratory diagnoses, and 93 per cent for non-respiratory diagnoses (Delfino, Backlake, & Hanley, 1993; Rawson & Malcolm, 1995b). Diagnoses had lower concordance among hospitalized psychiatric patients: 94 per cent agreement for schizophrenia but only 58 per cent for depressive disorder (Rawson et al., 1997).

Co-morbidities sometimes show less agreement; one early study found agreement ranging from 88 per cent (.56 kappa) for chronic obstructive pulmonary disease to 99 per cent (.78 kappa) for tumours (Houle et al., 1999; Humphries et al., 2000; Malenka, McLerran, Roos, Fisher, & Wennberg, 1994; Virnig & McBean, 2001).

Using a sample of all hospitalized patients avoids bias associated with a specific type of patient. A recent Alberta paper using such a sample showed encouraging improvements in agreement between hospital discharge abstracts and chart review for the 17 co-morbidities making up the Charlson index, a frequently used measure specifying co-morbidities known to be important in predicting outcomes (Quan, Parsons, & Ghali, 2002). Several Canadian studies have shown substantial agreement (kappa scores in the .60–.80 range) between diagnoses noted on the computerized hospital discharge abstract and those identified by chart re-abstraction (Austin, Daly, & Tu, 2002; Quan et al., 2002). Such agreement has been noted for acute myocardial infarction, diabetes, and several other conditions.

Finally, the central problem with identifying co-morbidities from computerized discharge abstracts is a lack of sensitivity; more co-morbidities are noted from detailed chart abstraction than in the computerized abstracts. However, the more important diagnoses seem to be recorded; the power of computerized diagnoses in predicting mortality is almost identical to that of the (greater number of) diagnoses generated by re-abstraction (Quan et al., 2002; Stukenborg, Wagner, & Connors, 2001; Tu et al., 2001).

#### *Physician Visits*

Outside of fee-for-service payment, completeness of claims may be a problem. More monitoring of "evaluation claims" generated by capitation – based payments or lump sum payments to clinics needed, although Papanicolaou testing was appropriately recorded at one Manitoba clinic (Cohen, 1993). Primary care reform – a popular topic in Canada – may increase the need for such monitoring, if movement away from fee-for-service payment is substantial.

Procedures and service dates on physician claims appear to be accurately reported (Table 3). One Quebec paper combined diagnostic and procedure codes in physician claims to assess injury ascertainment using the emergency room clinical record as the standard for comparison (Tamblyn, Reid, Mayo, McLeod, & Churchill-Smith, 2000). Analysing physician claims using a "test-retest" methodology has



**Table 3: Reliability of physician claims**

Author	Province	Records Examined	Years Examined	Variable Examined	Data for Comparison	Agreement
Blanchard et al., 1997	Manitoba	Claims for 434 children (0–14 years of age)	1985–1993	Diagnosis of diabetes	Diabetes Education Resource Database	Ascertainment rate of 95% for incident cases
Muhajarine et al., 1997	Manitoba	Claims for 2,275 survey respondents	1987–1990	Diagnosis of hypertension	Heart Health Survey, clinical measures	.56 kappa (82% agreement) for survey; .60 kappa (85%) for clinical measures
Pinfold et al., 2000	Ontario	735 women with node-negative breast cancer	1991	Procedure	Medical chart	95% (most definitive procedure); 98% (breast surgery type)
Roberts et al., 1994	Manitoba	253 claims in the Manitoba Immunisation Monitoring System	1980–1994	Childhood immunizations, service dates	Medical chart	98% (service dates); > 98% (immunizations)
Roos et al., 1979	Manitoba	Claims (4,757 pairs) for respiratory problems	1972–1974	Diagnoses	Two claims within a 14-day period (test-retest)	69–70% diagnostic agreement, 2 observations, 2 MDs; 93% agreement, 2 observations, same MD
Roos et al., 1979	Manitoba	Claims for 101 tonsillectomy/adenoidectomy patients	1973	Diagnoses, procedures	Claims Forms	Better than 99% in check for transcription error
Tamblyn et al., 2000	Quebec	Claims for 1,181 elderly treated in local emergency room	1993–1994	Diagnosis of injury type and associated procedure	Clinical record of type and date of injury	Variation in sensitivity from 26% (thorax fractures) to 97% (hip fracture); 81% overall

found respiratory diagnoses to be of fairly high reliability (Roos, Nicol, Johnson, & Roos, 1979).

### Prescriptions

Only Saskatchewan and Manitoba have prescription databases for almost the entire population (Miller, Blatman, & Einarson, 1996). However, Saskatchewan does not track medication use by First Nations residents (approximately 7.7% of the population) (Statistics Canada, 2001; Thiessen, Wallace, Blackburn, Wilson, & Bergman, 1990). Almost complete program data are available for British Columbia and Quebec (both provinces have prescription coverage for residents aged 65 or older), while New Brunswick provides coverage for elders not in nursing homes (Anderson, Kerluke, Pulcins, Hertzman, & Barer, 1993; Davidson, Molloy, Somers, & Bedard, 1994; Tamblyn, Lavoie, Petrella, & Monette, 1995).

Relatively little information is available on completeness and quality of the prescription databases. Primary data collection has been important in assessing the completeness of these databases. In Manitoba, given mandatory program participation, 93 per cent of events found in a manual audit had been computerized. Optional participation or First Nations membership lowered agreement with the manual audit to 90 per cent and 80 per cent, respectively (Kozyrskyj & Mustard, 1998). One British Columbia study noted that 93 per cent of their pharmaceutical claims had information on number and speciality of the prescriber "that could be linked to both the patient registry and unique physician identifiers" (Anderson & Kerluke, 1996).

### Case Identification

Using information on contact with the health care system to study health as well as utilization markedly extends the reach of administrative data. A particular condition, such as acute myocardial infarction, may sometimes be reliably identified using just hospital data (Austin et al., 2002; Cox, Melady, Chen, & Naylor, 1997; Levy, Tamblyn, Fitchett, McLeod, & Hanley, 1999).

Accumulating physician claims or hospital abstracts over two or more years provides better case ascertainment for chronic conditions (such as asthma, diabetes, hypertension, and inflammatory bowel disease) involving extended contact with the health care system (Bernstein, Blanchard, Rawsthorne, & Wajda, 1999; Brownell & Yogendran, 2001; Kozyrskyj, Mustard, Cheang, & Simons, 2001; Muhajarine et al., 1997; Robinson et al., 1997). Algorithms (the acceptance of one or more diagnoses on claims abstracts as indicating the condition) can be adjusted with a goal

of high kappa scores and prevalence estimates that show substantial agreement between administrative data and such data sources as surveys. Comparisons of hospital and physician files with the population-based Manitoba Heart Health Survey for 6 conditions varying in frequency found agreement ranging from 79 per cent (.40 kappa) for elevated cholesterol to 84 per cent (.59 kappa) for hypertension, and 96 per cent (.72 kappa) for diabetes (Muhajarine et al., 1997; Young, Roos, & Hammarstrand, 1991). To provide perspective, a number of conditions have shown kappas comparing diagnoses of two physicians to be about .60 (Koran, 1975; Sackett et al., 1991).

Sensitivity and specificity can often be "traded off" using administrative data and record linkage. Thus, two sets of algorithms (to minimize false positives and false negatives) were developed to assess depression using physician speciality, diagnoses from physician claims, and pharmaceutical data (West et al., 2000).

Linkage of pharmaceutical data and physician claims can help more generally with case identification (Table 4). Case identification can be complicated; for example, attention-deficit hyperactivity disorder (ADHD) was frequently diagnosed without a prescription claim for stimulant medication (Brownell & Yogendran, 2001). Moreover, if laboratory test results (rather than clinical signs and symptoms) are critical, high concordance values may mislead for some conditions. Aplastic anemia and agranulocytosis were definitely confirmed in only about 45 per cent of Saskatchewan cases, despite over 90 per cent agreement between charts and computerized hospital abstracts (Rawson & D'Arcy, 1998). Such findings raise concerns about the extent to which diagnostic validity can be assessed.

Finally, a contact-based measure of individual health status, the Adjusted Clinical Group (ACG) case-mix adjustment system, has built on diagnostic information from hospital abstracts and physician claims. Fewer than two per cent of the diagnostic codes from British Columbia and Manitoba were left unassigned (Reid, MacWilliam, Verhulst, Roos, & Atkinson, 2001). The high correlation between the ACG measure (aggregated to small areas) and premature mortality suggests high construct validity (Reid, Roos, MacWilliam, Frohlich, & Black, 2002).

### Discussion

#### *National Efforts*

The Canadian Institute for Health Information (CIHI) works toward improving the quality of the hospital

**Table 4: Case identification using physician claims and other sources**

Author	Manitoba	Records Examined	Years		Variable Examined	Data for Comparison	Agreement
			Examined	Examined			
Blanchard et al., 1996	Manitoba	Claims and discharge abstracts for 1,267 (1986) and 2,595 (1991) adults	1984–1992	Diagnosis of diabetes	Diabetes Education Resource Database	Ascertainment rates 96–98%	
Bernstein et al., 1999	Manitoba	Claims and discharge abstracts for 2,725 survey respondents; 448 for chart review		Diagnosis of Crohn's disease or ulcerative colitis	Survey, medical charts	Sensitivity/specificity of .89/.90–.91 (Crohn's disease) and .74–.88/.91–.94 (ulcerative colitis)	
Brownell & Yogendran, 2001	Manitoba	Claims and/or pharmaceutical data for 4,787 children (0–19 years of age)	1994–1995	Diagnosis of attention-deficit hyperactivity disorder (ADHD), record of prescribed stimulant medication	Claims, pharmaceutical data	87% of prescriptions for stimulant medication linkable to previous claims for ADHD or conduct disorders (80% in prior year)	
Robinson et al., 1997	Manitoba	Claims and discharge abstracts for 2,792 survey respondents	1987–1990	Diagnosis of diabetes, hypertension, high cholesterol, stroke, or myocardial infarction	Heart Health Survey	From .38 kappa / 87% (other heart disease) to .72 kappa / 96% (diabetes)	
West et al., 2000	Saskatchewan	Claims and pharmaceutical data for 587 adults	1994–1995	Primary diagnosis of depression, record of new antidepressant use	Medical chart	.54 kappa / 77% (depression) .50 kappa / 76% (depression with anxiety)	
Young et al., 1991	Manitoba	Claims and discharge abstracts for 1,227 survey respondents	1971, 1983	Diagnosis of diabetes	Manitoba Longitudinal Study on Aging	Sensitivity of 84%; specificity of 96%	

discharge abstracts. One re-abstraction study sampled 18 hospitals from a database, including 550 hospitals generating 75 per cent of Canadian hospital discharges; facilities in Quebec and some in Manitoba did not submit data (Richards, Brown, & Homan, 2001). Not surprisingly, “diagnoses with more complex treatment protocols and those that are less easily defined such as pneumonia, injuries and ambulatory care conditions showed a higher degree of discrepancies” (p. 5). Some disagreements between coder and re-abstractor on the “principal” procedure performed, the diagnosis “most responsible” for the hospitalization, and the patient’s residential postal code were also noted. With Canada now using the ICD-10 coding system, both CIHI and at least one investigator funded by the Canadian Institutes for Health Research are evaluating the effectiveness of implementation of the new standards. Substantial work will be needed to ensure high standards!

Accessibility and peer review have been a problem with the national efforts. Although Web-site posting provides one channel for dissemination, little work on the quality of the Canadian Hospital Discharge Database has been published in a format accessible through *PubMed* or the Institute for Scientific Information. Concerted work on the part of those working with CIHI data could remedy this obstacle. The federally sponsored Canadian Health Services Research Foundation successfully expanded its “Mythbusters” review series from a “grey literature” newsletter into a regular journal format (Canadian Health Services Research Foundation, 2003). Many academics would agree with Davidoff’s (2000, p. 58) conclusion that “a scientific discovery does not exist until it is safely in print.” Dissemination through channels that include academic publication would maximize visibility of valuable work.

#### *Work in Five Provinces*

Table 5 summarizes data quality work in the five provinces in which most of the studies using administrative data have been conducted. The provincial population registries seem quite complete, but only three provinces (Alberta, British Columbia, and Manitoba) have reported systematic checks on registry quality. Research on the reliability of both registry follow-up and residential information would increase the range of substantive studies. Given registries’ importance, the lack of detailed systematic attention to their quality (except in Manitoba) represents a major problem that may hold back research. For example, the new field of social epidemiology is likely to use improved registry data to help add a longitudinal dimension and facilitate linkage with community characteristics (Berkman & Kawachi,

2000; Veugelers, Yip, & Kephart, 2001; Roos, Magoon, Gupta, Chateau, & Veugelers, 2004).

The lack of attention to registries is not the result of chance. Registries are not as immediately recognizable as critical for research on health and health care as are substantive files. Registries point toward long-term studies, toward epidemiology rather than toward much of health services research. Where registry access is more constrained than access to other files, motivation to explore these issues is decreased. Increased database research should be facilitated by less-expensive checks on reliability: linkage-based methods appear practical in many cases. When necessary, problems highlighted by linkage could be further explored using primary data collection or re-abstraction. Inexpensive test-retest methods, applicable to several types of substantive files, might be used more extensively (Roos et al., 1979). Although a number of checks on diagnoses and procedures have been made, many diagnoses have not been studied. Promising analyses of the quality of pharmaceutical data need to be expanded, both substantively and across more provinces.

#### **The Future**

Over time, the growth of electronic medical records and the availability of detailed clinical information should reduce the need to compare hospital and physician information for consistency between the written chart and the computerized record (Tamblyn, 2002). While greater standardization of coding should improve data quality, differences among physicians in diagnosing disease will continue. Well-organized systems will facilitate checking sources and providers against each other and evaluating information from a multi-method perspective; work continues on developing appropriate algorithms (Huzel, Roos, Anthonisen, & Manfreda, 2003; Kozyrskyj et al., 2001). Case identification using administrative data – combined with sensitivity testing vis-à-vis the algorithms used – can generate important comparative information. Canadian Regional Health Authorities are particularly interested in such findings (Kozyrskyj & Hildes-Ripstein, 2002; Roberts, Fransoo, Black, Roos, & Martens, 2002). With care, longitudinal studies of cases identified from administrative data can produce valuable insights (Roos, Carrière, & Friesen, 1998).

Administrative data are critical for the study of risk-adjusted outcomes, even though hospital discharge abstracts (in both Canada and the United States) underestimate the number of associated co-morbidities (Humphries et al., 2000; Levy et al.,

**Table 5: Data quality checks in five provinces – An overview**

<b>Province</b>	<b>Registry</b>	<b>Hospital</b>	<b>Physician</b>	<b>Prescription</b>
British Columbia	Need checks on follow-up and residential identifiers	Checks on diagnoses and procedures weak	Unavailable	Unavailable
Saskatchewan	Not available; need publication of specific checks performed	Checks made on diagnoses and procedures	Checks made on diagnoses and procedures	Some substantive checks
Manitoba	Need additional checks on residential identifiers	Checks made on diagnoses and procedures	Checks made on diagnoses and procedures	Checks made on claims files; some substantive checks
Ontario	Unavailable	Checks made on diagnoses and procedures	Checks made on procedures	Unavailable
Quebec	Unavailable	Checks made on diagnoses and procedures	Checks made on diagnoses and procedures	Checks made on claims filed

This table summarizes data quality checks published to the end of 2002.



1999; Malenka et al., 1994; Quan et al., 2002; Virnig & McBean, 2001). Algorithms selecting appropriate co-morbidities both explain considerable variation in outcomes and work across several sites (Ghali et al., 2000; Tu et al., 2001). Although imperfect, coding improvements suggest possibilities for reducing expensive primary data collection (Ghali, Quan, & Brant, 2002; Quan et al., 2002). As discussed earlier, recent modelling efforts show few differences between adjustment using administrative data and that relying on more costly data collection.

The information-rich environments highlighted in Figure 1 offer great economies of effort, the ability to consider entire populations, and a large number of cases. As noted with genetic research, population-oriented work can tolerate a higher degree of error than can that more tightly focused on the individual (Kohler, 1994). Efficient approaches to generating data have proved revolutionary in certain life sciences (Brown, 2000; Weiner, 1999). This paper represents one step toward constructing a methodology built around appropriately organizing the available administrative data.

Such work should enable small jurisdictions with limited funding to creatively develop information infrastructures and improve research productivity. Although heterogeneity in diffusion of health care databases is seen in Australia and the United States, interest has been growing among Australian states (particularly Western Australia and South Australia) (Armstrong & Kricker, 1999; Holman et al., 1999; Starr, 1997). Successful integrated systems provide geographic and socio-economic information at the lowest possible level of aggregation to facilitate development of denominators and allow flexibility in "building" small areas up to different levels. In Canada, common areas for within-province analysis include regional health authorities, districts within the regional health authorities, and physician service areas (Roos & Shapiro, 1999).

Provincial and national cost data estimated at the micro-level are being incorporated in ongoing research (Jacobs, Blanchard, James, & Depew, 2001, Jacobs & Roos, 1999). In Manitoba, for example, home care data are being evaluated; efforts to monitor adult, as well as childhood, immunizations have also been initiated. Major efforts at linking administrative and clinical data are underway in Quebec. Such information-rich environments are shaping research in health policy, clinical epidemiology, and health services (Roos & Shapiro, 1999; Naylor & Slaughter, 1999; Canadian Institute for Health Information, 2000). The issues highlighted here will remain topical in the foreseeable future.

## References

- Acheson, E.D. (1967). *Medical record linkage*. London: Oxford University Press.
- Anderson, G., & Kerluke, K. (1996). Distribution of prescription drug exposures in the elderly: Description and implications. *Journal of Clinical Epidemiology*, 49, 929–935.
- Anderson, G.M., Kerluke, K.J., Pulcins, I.R., Hertzman, C., & Barer, M.L. (1993). Trends and determinants of prescription drug expenditures in the elderly: Data from the British Columbia Pharmacare Program. *Inquiry*, 30, 199–207.
- Armstrong, B.K., & Kricker, A. (1999). Record linkage: A vision renewed. *Australian New Zealand Journal of Public Health*, 23, 451–452.
- Austin, P.C., Daly, P.A., & Tu, J.V. (2002). A multicenter study of the coding accuracy of hospital discharge administrative data for patients admitted to cardiac care units in Ontario. *American Heart Journal*, 144, 290–296.
- Berkman, L.F., & Kawachi, I. (Eds.). (2000). *Social epidemiology*. New York: Oxford University Press.
- Bernstein, C.N., Blanchard, J.F., Rawsthorne, P., & Wajda, A. (1999). Epidemiology of Crohn's disease and ulcerative colitis in a central Canadian province: A population-based study. *American Journal of Epidemiology*, 149, 916–924.
- Blanchard, J.F., Ludwig, S., Wajda, A., Dean, H., Anderson, K., Kendall, O., & Depew, N. (1996). Incidence and prevalence of diabetes in Manitoba, 1986–1991. *Diabetes Care*, 19, 807–811.
- Blanchard, J.F., Dean, H., Anderson, K., Wajda, A., Ludwig, S., & Depew, N. (1997). Incidence and prevalence of diabetes mellitus in children 0–14 years in Manitoba, Canada 1985–1993. *Diabetes Care*, 20, 512–515.
- Bryan, H., & Brasher, P. (1995). Breast implants and breast cancer: Re-analysis of a linkage study. *New England Journal of Medicine*, 332, 1535–1539.
- Brown, K. (2000). The human genome business today. *Scientific American*, 283, 50–55.
- Brownell, M., & Yogendran, M. (2001). Attention-deficit hyperactivity disorder in Manitoba children: Medical diagnosis and psychostimulant treatment rates. *Canadian Journal of Psychiatry*, 46, 264–272.
- Canadian Health Services Research Foundation. (2003). User fees would stop waste and ensure better use of the health care system. *Journal of Health Services Research Policy*, 8, 105–107.
- Canadian Institute for Health Information. (2000). *Health care in Canada 2000: A first annual report*. Ottawa: Author.

- Chamberlayne, R., Green, B., Barer, M.L., Hertzman, C., Lawrence, W.J., & Sheps, S.B. (1998). Creating a population-based linked health database: A new resource for health services research. *Canadian Journal of Public Health, 89*, 270–273.
- Cohen, M.M. (1993). Using administrative data for case-control studies: The case of the Papanicolaou smear. *Annals of Epidemiology, 3*, 93–98.
- Cohen, M.M., Kaufert, P.A., MacWilliam, L., & Tate, R.B. (1996). Using an alternative data source to examine randomization in the Canadian National Breast Screening Study. *Journal of Clinical Epidemiology, 49*, 1039–1044.
- Cox, J.L., Melady, M.P., Chen, E., & Naylor, C.D. (1997). Towards improved coding of acute myocardial infarction in hospital discharge abstracts: A pilot project. *Canadian Journal of Cardiology, 13*, 351–358.
- Davidoff, F. (2000). Suppose there were no printers. *Annals of Internal Medicine, 133*, 57–58.
- Davidson, W., Molloy, W., Somers, G., & Bedard, M. (1994). Relation between physician characteristics and prescribing for elderly people in New Brunswick. *Canadian Medical Association Journal, 150*, 917–921.
- Delfino, R. J., Backlake, M.R., & Hanley, J.A. (1993). Reliability of hospital data for population-based studies of air pollution. *Archives of Environmental Health, 48*, 140–146.
- Dimich-Ward, H., Hertzman, C., Teschke, K., Hershler, R., Marion, S.A., Ostry, A., & Kelly, S. (1996). Reproductive effects of paternal exposure to chlorophenolate wood preservatives in the sawmill industry. *Scandinavian Journal of Work Environment and Health, 22*, 267–273.
- Edouard, L., & Rawson, N.S.B. (1996). Reliability of the recording of hysterectomy in the Saskatchewan health care system. *British Journal of Obstetrics and Gynaecology, 103*, 891–897.
- Fellegi, I.P., & Sunter, A.B. (1969). A theory for record linkage. *Journal of the American Statistical Association, (328)*, 1183–1210.
- Ghali, W.A., Rothwell, D.M., Quan, H., Brant, R., Tu, J.V., & Steering Committee of the Cardiac Care Network of Ontario. (2000). A Canadian comparison of data sources for coronary artery bypass surgery outcome “report cards.” *American Heart Journal, 140*, 402–408.
- Ghali, W.A., Quan, H., & Brant, R. (2002). Risk adjustment using administrative data: Impact of a diagnosis-type indicator. *Journal of General Internal Medicine, 16*, 519–524.
- Hatcher, J., & Hervas, M. (2001). Emigration patterns of cancer cases in Alberta, Canada. *Chronic Diseases in Canada, 22*, 12–17.
- Hawaleshka, D. (2002). The Maclean’s Health Report: Measuring health care: North and West Vancouver, Edmonton, Victoria, Kelowna, B.C.—Western centres top the list, *Maclean’s*, June 17, 23–31.
- Hawker, G.A., Coyte, P.C., Wright, J.G., Paul, J.E., & Bombardier, C. (1997). Accuracy of administrative data for assessing outcomes after knee replacement surgery. *Journal of Clinical Epidemiology, 50*, 265–273.
- Hertzman, C., Teschke, K., Ostry, A., Hershler, R., Dimich-Ward, H., Kelly, S., Spinelli, J.J., Gallagher, R.P., McBride, M., & Marian, S.A. (1997). Mortality and cancer incidence among sawmill workers exposed to chlorophenolate wood preservatives. *American Journal of Public Health, 87*, 71–79.
- Holman, C.D.J., Bass, A.J., Rouse, I.L., & Hobbs, M.S.T. (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*, 453–459.
- Houle, C., Berthelot, J.-M., David, P., Wolfson, M.C., Mustard, C.A., & Roos, L.L. (1999). Matching census database and Manitoba health care files. In *Record linkage techniques – 1997: Proceedings of an international workshop and exposition*. Washington, D.C.: National Academy Press, 305–318.
- Howe, G.R. (1998). Use of computerized record linkage in cohort studies. *Epidemiologic Reviews, 20*, 112–121.
- Humphries, K.H., Rankin, J.M., Carere, R.G., Buller, C.E., Kiely, F.M., & Spinelli, J.J. (2000). Co-morbidity data in outcomes research: Are clinical data derived from administrative databases a reliable alternative to chart review? *Journal of Clinical Epidemiology, 53*, 343–349.
- Huzel, L., Roos, L.L., Anthonisen, N.R., & Manfreda, J. (2003). Diagnosing asthma: The fit between survey and administrative database. *Canadian Respiratory Journal, 9*, 407–412.
- Iron, K., Goel, V., & Williams, J.I. (1995). *Concordance of hospital discharge abstracts and physician claims for surgical procedures in Ontario*. (ICES working paper series #42.) North York: Institute for Clinical Evaluative Sciences in Ontario.
- Jacobs, P., Blanchard, J.F., James, R., & Depew, N. (2001). Excess costs of diabetes in the aboriginal population of Manitoba, Canada. *Canadian Journal of Public Health, 91*, 298–301.
- Jacobs, P., & Roos, N.P. (1999). Standard cost lists for health care in Canada: Issues in validity and interprovincial consolidation. *Pharmacoeconomics, 15*, 551–560.
- Kohler, R.E. (1994). *Lords of the fly*. Chicago: University of Chicago Press.
- Koran, L.M. (1975). The reliability of clinical methods, data and judgements (Part 1). *New England Journal of Medicine, 293*, 642–646.

- Kozyrskyj, A., & Hildes-Ripstein, G.E. (2002). Assessing health status in Manitoba children: Acute and chronic conditions. *Canadian Journal of Public Health, 93*, S63–S69.
- Kozyrskyj, A., & Mustard, C.A. (1998). Validation of an electronic, population-based prescription database. *Annals of Pharmacotherapy, 32*, 1152–1157.
- Kozyrskyj, A., Mustard, C.A., Cheang, M.S., & Simons, F.E.R. (2001). Income-based drug benefit policy: Impact on receipt of inhaled corticosteroid prescriptions by Manitoba children with asthma. *Canadian Medical Association Journal, 165*, 897–902.
- Levy, A.R., Tamblyn, R., Fitchett, D., McLeod, P.J., & Hanley, J.A. (1999). Coding accuracy of hospital discharge data for elderly survivors of myocardial infarction. *Canadian Journal of Cardiology, 15*, 1277–1282.
- Malenka, D.J., McLerran, D.F., Roos, N.P., Fisher, E.S., & Wennberg, J.E. (1994). Using administrative data to describe casemix: A comparison with the medical record. *Journal of Clinical Epidemiology, 47*, 1027–1032.
- Manitoba Health. (2002). *Manitoba's health indicators report*. Winnipeg: Author.
- Miller, E., Blatman, B., & Einarson, T.R. (1996). A survey of population-based drug databases in Canada. *Canadian Medical Association Journal, 154*, 1855–1864.
- Muhajarine, N., Mustard, C.A., Roos, L.L., Young, T.K., & Gelskey, D.E. (1997). Comparison of survey data and physician claims data for detecting hypertension. *Journal of Clinical Epidemiology, 50*, 711–718.
- Mustard, C.A., Harman, C.R., Hall, P.F., & Derksen, S. (1995). Impact of a nurses' strike on the cesarean birth rate. *American Journal of Obstetrics and Gynecology, 172*, 631–637.
- Naylor, C.D. (1999). Health care in Canada: Incrementalism under fiscal duress. *Health Affairs (Millwood), 18*, 9–26.
- Naylor, C.D., & Slaughter, P.M. (Eds.). (1999). *Cardiovascular health & services in Ontario: An ICES atlas*. Toronto: Institute for Clinical Evaluative Sciences.
- Newcombe, H.B. (1988). *Handbook of record linkage*. New York: Oxford University Press.
- Pinfold, S.P., Goel, V., & Sawka, C. (2000). Quality of hospital discharge and physician data for types of breast cancer surgery. *Medical Care, 38*, 99–107.
- Potvin, L., & Champagne, F. (1986). Utilization of administrative data files in health research. *Social Indicators Research, 18*, 409–423.
- Quan, H., Parsons, G.A., & Ghali, W.A. (2002). Validity of information on comorbidity derived from ICD-9-CCM administrative data. *Medical Care, 40*, 675–685.
- Rankin, J.M., Spinelli, J.J., Carere, R.G., Ricci, D.R., Penn, I.M., Hilton, J.D., Henderson, M.A., Hayden, R.I., & Buller, C.E. (1999). Improved clinical outcome after widespread use of coronary-artery stenting in Canada. *New England Journal of Medicine, 341*, 1957–1965.
- Rawson, N.S.B., & D'Arcy, C. (1998). Assessing the validity of diagnostic information in administrative health care utilization data: Experience in Saskatchewan. *Pharmacoepidemiology & Drug Safety, 7*, 389–398.
- Rawson, N.S.B., & Malcolm, E. (1995a). *Validity of the recording of cholecystectomy and hysterectomy in the Saskatchewan health care datafiles*. (Technical report series report #3). Saskatoon: Pharmacoepidemiology Research Unit.
- Rawson, N.S.B., & Malcolm, E. (1995b). Validity of the recording of ischaemic heart disease and chronic obstructive pulmonary disease in the Saskatchewan health care datafiles. *Statistics in Medicine, 14*, 2627–2643.
- Rawson, N.S.B., Malcolm, E., & D'Arcy, C. (1997). Reliability of the recording of schizophrenia and depressive disorder in the Saskatchewan health care datafiles. *Social Psychiatry and Psychiatric Epidemiology, 32*, 191–199.
- Reid, R.J., MacWilliam, L., Verhulst, L., Roos, N.P., & Atkinson, M. (2001). Performance of the ACG Case-Mix System in two Canadian provinces. *Medical Care, 39*, 86–99.
- Reid, R.J., Roos, N.P., MacWilliam, L., Frohlich, N., & Black, C. (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*, 1345–1364.
- Richards, J., Brown, A., & Homan, C. (2001). *The data quality study of the Canadian discharge abstract database*. Proceedings of the Statistics Canada's 2001 Symposium October 16–19, "Achieving Data Quality in a Statistical Agency: A Methodological Perspective." Ottawa: Statistics Canada.
- Risch, H.A., & Howe, G.R. (1994). Menopausal hormone usage and breast cancer in Saskatchewan: A record-linkage cohort study. *American Journal of Epidemiology, 139*, 670–683.
- Roberts, J.D., Poffenroth, L.A., Roos, L.L., Bebchuk, J.D., & Carter, A.O. (1994). Monitoring childhood immunizations: A Canadian approach. *American Journal of Public Health, 84*, 1666–1668.
- Roberts, J.D., Fransoo, R., Black, C., Roos, L.L., & Martens, P. (2002). Research meets reality: Administrative data to guide planning for Canadian Regional Health Authorities. *Healthcare Management Forum, 15*, 13–21.
- Robinson, J.R., & Tataryn, D.J. (1997). Reliability of the Manitoba Mental Health Management Information System for Research. *Canadian Journal of Psychiatry, 42*, 744–749.



- Robinson, J.R., Young, T.K., Roos, L.L., & Gelskey, D.E. (1997). Estimating the burden of disease: Comparing administrative data and self-reports. *Medical Care*, 35, 932–947.
- Roos, L.L., Magoon, J., Gupta, S., Chateau, D., & Veugelers, P.J. (2004). Socioeconomic determinants of mortality in two Canadian provinces: Multilevel modelling and neighborhood context. *Social Science and Medicine*, (7), 1435–1447.
- Roos, L.L., Menec, V., & Currie, R.J. (2004). Policy analysis in an information-rich environment. *Social Science and Medicine*, (11), 2231–2241.
- Roos, L.L., Nicol, J.P. & Wajda, A. (1985). Improving the quality of data banks through linkage. *Chronic Diseases in Canada*, 5, 81–82.
- Roos, L.L., & Nicol, J.P. (1999). A research registry: Uses, development, and accuracy. *Journal of Clinical Epidemiology*, 52, 39–47.
- Roos, L.L., Nicol, J.P., Johnson, C., & Roos, N.P. (1979). Using administrative data banks for research and evaluation: A case study. *Evaluation Quarterly*, 3, 236–255.
- Roos, L.L., Roos, N.P., Cageorge, S.M., & Nicol, J.P. (1982). How good are the data? Reliability of one health care data bank. *Medical Care*, 20, 266–276.
- Roos, L.L., Sharp, S.M., & Cohen, M.M. (1991). Comparing clinical information with claims data: Some similarities and differences. *Journal of Clinical Epidemiology*, (9), 881–888.
- Roos, L.L., Sharp, S.M., & Wajda, A. (1989). Assessing data quality: A computerized approach. *Social Science and Medicine*, 28, 175–182.
- Roos, L.L., & Wajda, A. (1991). Record linkage strategies: Part I. Estimating information and evaluating approaches. *Methods of Information in Medicine*, 30, 117–123.
- Roos, L.L., Walld, R., Wajda, A., Bond, R., & Hartford, K. (1996). Record linkage strategies, outpatient procedures, and administrative data. *Medical Care*, 34, 570–582.
- Roos, N.P., Carrière, K.C., & Friesen, D. (1998). Factors influencing the frequency of visits by hypertensive patients to primary care physicians in Winnipeg. *Canadian Medical Association Journal*, 159, 777–783.
- Roos, N.P., & Shapiro, E. (Eds.). (1999). Academics at the policy interface: Revisiting the Manitoba Centre for Health Policy and Evaluation and its population-based health information system. *Medical Care*, 37(Suppl. 6), JS1–JS308.
- Sackett, D.L., Haynes, R.B., Guyatt, G.H., & Tugwell, P. (Eds.). (1991). *Clinical epidemiology: A basic science for clinical medicine* (2nd ed.). Boston: Little, Brown and Company.
- Sheps, S.B., Reid, R.J., Barer, M.L., Krueger, H., McGrail, K.M., Green, B., Evans, R.G., & Hertzman, C. (2000). Hospital downsizing and trends in health care use among elderly people in British Columbia. *Canadian Medical Association Journal*, 163, 411–412.
- Starr, P. (1997). Smart technology, stunted policy: Developing health information networks. *Health Affairs (Millwood)*, 16, 91–105.
- Statistics Canada. (2001). *Population by aboriginal group, 1996 Census*. Retrieved December 14, 2004, from <http://www.statcan.ca/english/Pgdb/People/Population/dem039a.htm>
- Stukenborg, G.J., Wagner, D.P., & Connors, A.F., Jr. (2001). Comparison of the performance of two comorbidity measures, with and without information from prior hospitalizations. *Medical Care*, 39, 727–739.
- Tamblyn, R. (2002). The new millennium model for health care and research. In J. Downey & L. Claxton (Eds.), *Innovation: Essays by leading Canadian researchers*. 185–193. Toronto: Key Porter Books Ltd.
- Tamblyn, R., Lavoie, G., Petrella, L., & Monette, J. (1995). The use of prescription claims databases in pharmacoepidemiological research: The accuracy and comprehensiveness of the prescription claims database in Quebec. *Journal of Clinical Epidemiology*, 48, 999–1009.
- Tamblyn, R., Reid, T., Mayo, N.E., McLeod, P.J., & Churchill-Smith, M. (2000). Using medical services claims to assess injuries in the elderly: Sensitivity of diagnostic and procedure codes for injury ascertainment. *Journal of Clinical Epidemiology*, 53, 183–194.
- Thiessen, B.Q., Wallace, S.M., Blackburn, J., Wilson, T.W., & Bergman, U. (1990). Increased prescribing of antidepressants subsequent to B-blocker therapy. *Archives of Internal Medicine*, 150, 2286–2290.
- Tu, J.V., Austin, P.C., Walld, R., Roos, L.L., Agras, J., & McDonald, K.M. (2001). Development and validation of the Ontario acute myocardial infarction mortality prediction rules. *Journal of the American College of Cardiology*, 37, 992–997.
- Veugelers, P.J., Yip, A.M., & Kephart, G. (2001). Proximate and contextual socioeconomic determinants of mortality: Multilevel approaches in a setting with universal health care coverage. *American Journal of Epidemiology*, 154, 725–732.
- Virnig, B.A., & McBean, M. (2001). Administrative data for public health surveillance and planning. *Annual Review of Public Health*, 22, 213–230.
- Wajda, A., & Roos, L.L. (1987). Simplifying record linkage: Software and strategy. *Computers in Biology and Medicine*, 17, 239–248.
- Wajda, A., Roos, L.L., Layefsky, M., & Singleton, J.A. (1991). Record linkage strategies, part 2: Portable software and

deterministic matching. *Methods of Information in Medicine*, 30, 210–214.

Weiner, J. (1999). *Time, love, memory: A great biologist and his quest for the origins of behavior*. New York: Alfred A. Knopf.

West, S.L., Richter, A., Melfi, C.A., McNutt, M., Nennstiel, M.E., & Mauskopf, J.A. (2000). Assessing

the Saskatchewan database for outcomes research studies of depression and its treatment. *Journal of Clinical Epidemiology*, 53, 823–831.

Young, T.K., Roos, N.P., & Hammarstrand, K.M. (1991). Estimated burden of diabetes mellitus in Manitoba according to health insurance claims: A pilot study. *Canadian Medical Association Journal*, 144, 318–324.