

Developing and Maintaining a Population Research Registry to Support Primary Healthcare Research

Création et maintien d'un registre démographique
pour la recherche sur les soins de santé primaires



by ANNE-MARIE BROEMELING, PHD
Director, Information Support and Research
Interior Health Authority
Kelowna, BC

Centre for Health Services and Policy Research
University of British Columbia
Vancouver, BC

KERRY KERLUKE, BSC
Programmer/Analyst, Centre for Health Services and Policy Research
University of British Columbia
Vancouver, BC

CHARLYN BLACK, MD, SCD
Centre for Health Services and Policy Research
University of British Columbia
Vancouver, BC

Anne-Marie Broemeling et al.

SANDRA PETERSON, MSC

*Research Analyst, Centre for Health Services and Policy Research
University of British Columbia
Vancouver, BC*

ALLYSON MACDONALD, BSC

*Documentation Specialist, Centre for Health Services and Policy Research
University of British Columbia
Vancouver, BC*

RACHAEL MCKENDRY, MA

*Project Manager, Centre for Health Services and Policy Research
University of British Columbia
Vancouver, BC*

Abstract

What did we do? This paper describes the creation of a population research registry as part of an information system to support primary healthcare (PHC) research in British Columbia. The population registry includes all residents of the province who were either eligible to use or actually used healthcare services, together with demographic, geographic, health status, registration and service use data. The PHC population research registry is built using administrative data inputs, and data are anonymized to comply with privacy and confidentiality standards.

What did we learn? The registry provides data to undertake research into PHC needs and service utilization. It facilitates both population-based research as well as research on population subgroups. Combined with anonymous physician and utilization data, the information system can be used to study service utilization rates for population-based analyses. Over the longer term, the information will contribute to our understanding of PHC qualities and outcomes.

What are the implications? Continued completeness of the population research registry depends upon full administrative source data. Planning to ensure complete data capture is critical both for the research registry and our ability to undertake population-based PHC research.

Résumé

Ce que nous avons fait : Cet article décrit la création d'un registre de recherche démo-

graphique qui fait partie d'un système d'information pour la recherche sur les soins de santé primaires (SSP) en Colombie-Britannique. Le registre démographique recense tous les résidents de la province qui ont utilisé les services de santé ou qui y sont admissibles et comprend des données démographiques, géographiques, de même que des données sur l'état de santé, l'immatriculation et l'utilisation des services de santé. Le registre de recherche démographique pour les SSP est constitué à partir de données administratives, lesquelles conservent l'anonymat afin de respecter les normes de confidentialité.

Ce que nous avons appris : Le registre présente des données qui permettent de faire de la recherche sur l'utilisation et les besoins en matière de SSP. Il permet de faire à la fois des recherches démographiques et des recherches sur des sous-groupes de la population. Combiné aux données anonymes sur les médecins et sur l'utilisation, le système d'information peut être employé pour étudier les taux d'utilisation afin d'effectuer des analyses de nature démographique. À long terme, cette information aidera à mieux connaître la qualité et les résultats liés aux SSP.

Répercussions : Le caractère exhaustif du registre dépend d'un accès complet aux sources de données administratives. Il est essentiel de planifier une saisie complète des données tant pour le registre que pour la capacité d'entreprendre des recherches démographiques sur les SSP.

POPULATION RESEARCH REGISTRIES PROVIDE NECESSARY FOUNDATIONAL information for primary healthcare (PHC) research and, more generally, for health services research. Data on population socio-demographic, morbidity, mortality, geographic, service registration and other characteristics provide critical information to describe and study need, demand for and use of services by the population overall and by population subgroups. Research registries using linked health data have been established in a number of jurisdictions to support population-based research initiatives (Roos and Nicol 1999; Roos et al. 2003). Data from these registries enable policy analyses, planning and management of healthcare resources (Roos et al. 2004) and information to support system-level performance measurement and evaluation of services.

This special issue of *Healthcare Policy* documents the development of an information system for PHC research in British Columbia. In this paper, we describe the development of a population research registry as part of this PHC information system. The population research registry includes almost all individuals living in British Columbia, together with anonymized core data for each individual. Strict protocols have been established to ensure security and to protect privacy and confidentiality of personal health information.

The BC population research registry can be used to identify and study groups in the province and to support analysis of treatment prevalence and other population-based rates. Not only does the registry provide data on the population, but these can be combined with data from other sources to analyze service utilization, some qualities of PHC and selected outcomes for the population. The population research registry, along with a physician information system (see Watson et al. 2009, page 77 of this special issue of *Healthcare Policy*), forms the foundation for a PHC information system to support research on population, providers and utilization of healthcare services.

What Did We Do?

As a foundation for PHC research, we created a population research registry containing anonymized data for individual residents of British Columbia. The registry includes individuals who were eligible to receive medical services as well as residents who were not registered but who used healthcare services during the year. The population research registry provides a near-complete file of all individuals living in the province each year, together with key data for each person. It was created to provide core, population-based data while ensuring anonymity of information and complying with privacy and confidentiality requirements.

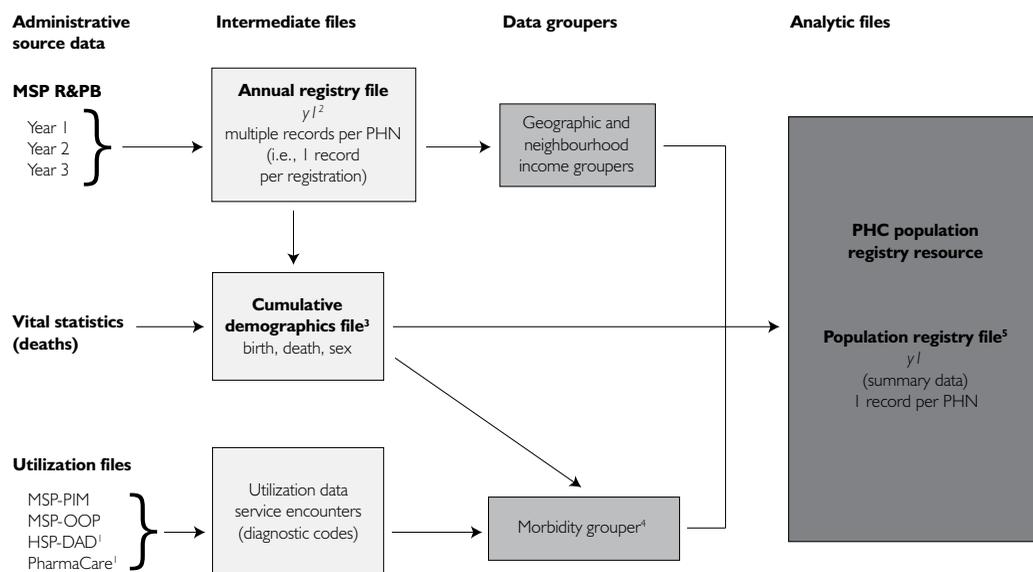
The project builds on previous work at the Centre for Health Services and Policy Research (CHSPR) to establish a population-based linked health database: the British Columbia Linked Health Database (BCLHD) (Chamberlayne et al. 1998). Using the BCLHD, our work incorporates additional information for PHC research, including derived morbidity data for individuals in the registry. While our work focuses on PHC as part of a comprehensive information system to support PHC research, it is also applicable to other health services research.

We used a three-stage process to create the anonymized population research registry, starting with (a) individual-level data from a number of existing administrative sources, (b) the creation of intermediate data files to link administrative source data and (c) the addition of derived geographic, income and morbidity data using data groupers to create summary annual population registry files for analysis. Figure 1 describes the process used to develop the registry.

Individual residents registered to receive provincial medical services were identified using Medical Services Plan Registration and Premium Billing (MSP R&PB) data. This annual registry file identifies all residents eligible to receive services and summarizes three years of medical plan data. Since the MSP R&PB files are compiled at a specific time (e.g., July 1), multiple files were required to ensure data completeness and to incorporate all registration changes. For example, the annual registry file for 2002/03 was derived from MSP R&PB files for 2002/03, 2003/04 and 2004/05 in order to ensure complete identification of registered residents (see Figure 1).

Developing and Maintaining a Population Research Registry to Support Primary Healthcare Research

FIGURE 1. Developing the PHC population research registry



Notes

¹ The HSP-DAD and the PharmaCare data include some out-of-province (OOP) data.

² Snapshots of years 1, 2 and 3 from MSP R&PB are combined to develop a cumulative annual registry file that represents year 1.

³ The cumulative demographics file includes one record for each person ever registered with MSP in BC from 1986 forward.

⁴ We used The Johns Hopkins ACG® Case-Mix System PC-Unix Version 5.01 licensed software, adapted for BC coding, to create the morbidity measures. Each person is assigned only one ACG but multiple ADGs and EDCs.

⁵ The population registry file includes one record for each person either registered or a service user in that year, i.e., year y1. It includes morbidity measures for each person.

Abbreviations

MSP	Medical Services Plan
R&PB	Medical Services Plan Registration and Premium Billing File
PHN	Personal Health Number
MSP-PIM	Medical Services Plan Payment Information Masterfile
MSP-OOP	Medical Services Plan Out-of-Province Billing File
HSP-DAD	Hospital Separations File (Discharge Abstract Database)
ACG	Adjusted Clinical Groups
ADG	Aggregated Diagnosis Groups
EDC	Expanded Diagnosis Clusters

In addition to individual residents registered with the Medical Services Plan (MSP), the population research registry identifies individuals who were not registered but who used services. To identify non-registered service users, the registry includes utilization data for fee-for-service medical payments (MSP PIM), out-of-province physician services (MSP-OOP), hospital discharges for residents of British Columbia both in and out of province (HSP-DAD) and seniors' pharmaceutical use (PharmaCare). Utilization files were compared to the annual registry to identify non-registered individuals using these services. These non-registered service users accounted for an average of 0.5% to 1.5% of the total registry population per year.

Demographic information for each resident was identified, including age and sex (cumulative demographics file). The demographics file also identified individual resi-

dents who died during the year. Detailed checks were implemented to avoid double-counting of individuals, recognizing that people may re-register or change MSP premium payment sources over time.

Intermediate registry, demographic and utilization files were combined with data groupers as a means of categorizing individuals. These tools cluster individual-level data into groupings by home geographic area, socio-economic status and morbidity.

Geographic groupers used postal code data to categorize residents into home areas. The BC Stats Translation Master File (TMF) and Statistics Canada Postal Code Conversion File (PCCF) were used along with individual resident postal code data to group BC residents into home geographic areas. The BC Stats TMF was used to derive local health area, health service delivery area and health authority variables for individual residents. The Statistics Canada PCCF was used to derive federal geographic area measures. Data for home residential location are critical for defining geographic populations to calculate utilization, treatment prevalence and other population-based rates.

Home geographic area data were also used to derive proxy measures of socio-economic status. Individuals were grouped into neighbourhoods, and this information was combined with census neighbourhood income data to derive neighbourhood income quintiles and deciles as a proxy for individual socio-economic status.

Morbidity data for all individuals in the population research registry were derived using age and sex data from the cumulative demographics file and diagnosis data from payment (e.g., MSP-PIM) and services files (e.g., HSP-DAD). Morbidity categories were identified using the Johns Hopkins case-mix group tools: Adjusted Clinical Groups (ACGs), Aggregated Diagnosis Groups (ADGs) and Expanded Diagnosis Clusters (EDCs) (Johns Hopkins 2003). The ACG and EDC tools have been validated in Canada and have been used to estimate population health status, to develop risk adjustment measures and to study chronic health conditions (Reid et al. 1999; Reid et al. 2002; Broemeling et al. 2005). We used the registry data and case-mix tools to derive a summary morbidity measure and to identify the presence of specific conditions (e.g., diabetes, congestive heart failure), types of conditions (e.g., acute, chronic), service use categories (e.g., non-users) and co-morbidity categories based on the number of types of conditions experienced (none, low, medium, high, very high co-morbidity) for each individual in the population research registry. More detail on these morbidity measures is reported elsewhere (Broemeling et al. 2005).

As a result of linking administrative source files to create intermediate files – combined derived geographic, income and morbidity information – the population research registry provides integrated data in annual registry files for almost all BC residents. Annual files were created enabling longitudinal analyses. The 2003/04 population research registry contains anonymized individual-level data for 4.2 million British Columbians. Provincial population estimates for British Columbia for 2003/04 also

report 4.2 million residents (British Columbia 2007).

Unique scrambled identifiers protect privacy and ensure anonymity. Privacy and confidentiality requirements are strictly protected. Population research registry information is available for approved research projects, but all requests to use the data must be reviewed and approved by data stewards responsible for the information. Strict rules are applied for accessing information so that no individuals can be identified or their confidentiality compromised. In addition to privacy and data access review, researchers must apply for research ethics review and approval, report only aggregated results and submit research reports for review prior to release to ensure that confidentiality standards are strictly maintained.

What Did We Learn?

The population research registry, along with physician information and utilization data, provides the foundation for studying the interactions between the population and providers, including the specific types of services used by individuals and delivered by providers.

In British Columbia, the creation of a population research registry was complicated by the collection of premiums for medical services plan registration. The province is one of a few in Canada collecting premiums for medical services coverage. The majority of individuals residing in British Columbia are registered with the provincial medical services plan, and they either pay their premiums themselves or have them paid on their behalf by employers or provincial or federal programs. However, premiums create a disincentive for some individuals to enrol with the medical services plan, resulting in incomplete data on provincial residents from the province's medical services plan registration file (MSP R&PB). Additional data were required to identify individuals in the population who were not registered with MSP. For the population research registry, we used service utilization data to identify individuals who used healthcare services but who were not registered with the provincial medical plan.

While the population research registry is as complete as possible using administrative data, it does not include individuals who have never registered with the provincial medical plan or used provincial healthcare services. At this point, no data are available to identify individuals living in the province who have never registered with the Medical Services Plan and who have not used publicly supported healthcare services, and we are thus unable to provide information on this population subgroup.

The population research registry enables system-level research and performance reporting for primary healthcare.

The population research registry enables system-level research and performance reporting for PHC. The registry provides core information on population characteristics to describe the population of the province at a point in time or over time. It also allows us to identify and study the population according to age, sex, morbidity, chronic health conditions or geographic categories. The registry provides the denominator for analyses of population-based rates such as treatment prevalence rates for specific chronic health conditions. Annual registry files support temporal analyses of the population.

The population research registry differs from other data sources such as the Census, population forecasts, Vital Statistics and disease registries. As an example, disease registries have been developed to identify individuals with specific conditions, usually chronic health conditions. Registries such as the National Diabetes Surveillance System (Health Canada 2003) and provincial diabetes registries (British Columbia 2006) identify the subset of the population diagnosed with diabetes. In some provinces, these disease registries are used to assist physicians in identifying specific patients for chronic disease management.

On the other hand, the population research registry captures almost all residents of the province, irrespective of whether they have specific health conditions, along with information on whether they have been diagnosed as having specific conditions. This anonymized registry provides the foundation for undertaking population-based research, while disease registries focus only on individuals with specific conditions.

The population research registry also provides important information (e.g., age, sex, morbidity and registration status) that can be used to identify and study subgroups in the population. For example, the registry provides the foundation for identifying such groups as “high users” – the top 5% of medical service users (Reid et al. 2003). Grouping data from the population registry with service utilization data allows us to identify other high user categories, such as high users of hospital services or users of ambulatory care sensitive (ACS) condition hospitalizations, which should be minimized with primary care services (Billings 1996; MCHP 2005).

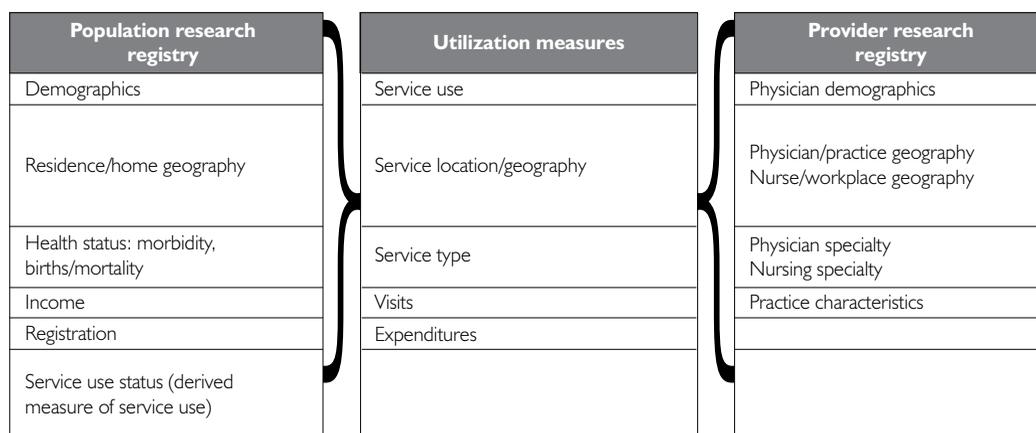
Not only does the population research registry provide data for the population, but it can be combined with anonymized provider data to enable analyses of service utilization and the interaction between BC residents and PHC providers. The population research registry complements a physician research registry (see Watson et al. 2009) to support measurement of service utilization such as the number and type of visits to PHC physicians by the population overall, by population subgroups (e.g., people with different patterns of morbidity) and over time.

Figure 2 describes the key measures that comprise the CHSPR Primary Care project, including the population research registry.

What Are the Implications?

The value of the population research registry depends upon the quality and completeness of administrative data inputs to build the registry files. In recent years, an increasing proportion of medical services has been provided through alternative funding arrangements such as salaried physician services and sessional payments to physicians. These changes in funding arrangements have had an impact on the completeness of payment files (e.g., MSP-PIM) and, therefore, on diagnostic data available to support the accuracy of morbidity profiles in the population research registry. Therefore, the increasing proportion of physician services funded through alternative funding arrangements has had an impact on the completeness of not only utilization data to measure service encounters but also on the completeness of diagnostic data to identify morbidity levels in the population.

FIGURE 2. CHSPR Primary Care Project



It is unlikely that the increasing proportion of alternative funding arrangements has had a substantive impact on the total number of people identified in the research registry. The majority of individuals were identified using medical plan registration data, and historically, only 0.5% to 1.5% of the population has been identified using service utilization data. However, the main impact of increasing alternative funding arrangements has been to reduce information available for each individual in the registry, including morbidity, types of conditions and co-morbidity measures.

Moreover, the increasing proportion of alternative funding arrangements will have a differential impact on some population groups because these arrangements vary by service type and geographic area. For example, alternative funding arrangements account for an increasing proportion of paediatric, psychiatric, emergency and oncologic services (British Columbia 2003). Specific geographic areas are also affected as a

result of a higher proportion of alternative funding services in some geographic areas (Watson, Peterson et al. 2009).

Some jurisdictions have implemented data collection tools, such as shadow billing, along with alternative funding arrangements to ensure reporting of activity and individual patient–physician interactions and to avoid information loss (Ontario Ministry of Health and Long-Term Care 2005). In the absence of such data for British Columbia, the population research registry can be used to highlight these data issues and to focus analyses on geographic areas, age groups or services with more complete information.

Secondly, increasing concern for privacy of personal health information contributes to escalating complexity of research and evaluation activities in Canada (Black et al. 2005). On the one hand, there is recognition of the value of population research registries and data linkage to meet expectations for accountability, performance reporting, planning for service needs and health services research. Indeed, in many areas, there is recognition of the value of adding new data sources to enrich existing research data. Yet, there is also growing concern about the use of personal health information, including non-consensual use required for population-based research. Furthermore, legislative and policy frameworks are not consistent across Canada, despite initiatives such as the proposed Health Information Privacy and Confidentiality Framework (Health Canada 2005).

It is a priority for future work in Canada to ensure consistent, comprehensive standards for use of health data that supports population-based research while ensuring security and protecting privacy and confidentiality of individual data. Recommendations have proposed the development of consistent best practice standards for use of health information and protection of privacy, including data stewardship, privacy toolkits and research ethics review standards (Black et al. 2005). This work is required to sustain and develop population research registries and information systems that are critical to health services research, performance reporting and accountability.

It is a priority for future work in Canada to ensure consistent, comprehensive standards for use of health data that supports population-based research while ensuring security and protecting privacy and confidentiality of individual data.

Conclusion

The BC population research registry at CHSPR provides a foundation for population-based PHC research, both at the population level and by subgroups within the population. This information, in turn, facilitates system-level performance measure-

ment and research on PHC. The BC population research registry provides a model for future PHC information systems and highlights the importance of complete data with which to study populations and their use of services.

ACKNOWLEDGEMENTS

The British Columbia Ministry of Health Services provided funding for this project as well as some of the funding for the special issue in which this article is published. Conclusions are those of the authors and reflect no official endorsement by our funders. We would like to acknowledge the contributions of Dr. Rob Reid to our original report on this topic.

Correspondence may be directed to: Dr. Anne-Marie Broemeling, Director, Information Support and Research, Interior Health Authority, 104 - 1815 Kirschner Road, Kelowna, BC V1Y 4N7; tel.: 250-870-4649, fax: 250-870-4605; e-mail: anne-marie.broemeling@interiorhealth.ca.

REFERENCES

- Billings, J., G.M. Anderson and L.S. Newman. 1996. "Recent Findings on Preventable Hospitalizations." *Health Affairs* 15(3): 239–49.
- Black, C., K. McGrail, C. Fooks and L. Maslove. 2005. *Data, Data, Everywhere: Improving Access to Population Health and Health Services Research Data in Canada*. Vancouver and Ottawa: Centre for Health Services and Policy Research/Canadian Policy Research Networks.
- British Columbia. 2003. *Alternative Payments to Physicians: A Program in Need of Change*. Victoria, BC: Office of the Auditor General.
- British Columbia. 2006. *Chronic Disease Management*. Victoria, BC: Ministry of Health. Retrieved January 5, 2009. <<http://www.health.gov.bc.ca/cdm/index.html>>.
- British Columbia. 2007. *BC Stats: Population Projections – BC and Regional*. Victoria, BC: Ministry of Labour and Citizens' Services. Retrieved January 5, 2009. <<http://www.bcstats.gov.bc.ca/DATA/pop/pop/popproj.asp>>.
- Broemeling, A.-M., D. Watson and C. Black. 2005. *Chronic Conditions and Co-morbidity among Residents of British Columbia*. Vancouver: Centre for Health Services and Policy Research.
- 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): 270–73.
- Health Canada. 2003. *National Diabetes Surveillance System*. Retrieved January 5, 2009. <<http://www.phac-aspc.gc.ca/ccdpc-cpcmc/diabetes-diabete/english/ndss/index.html>>.
- Health Canada. 2005 (January 27). *Pan-Canadian Information Privacy and Confidentiality Framework*. Retrieved January 5, 2008. <<http://www.hc-sc.gc.ca/hcs-sss/pubs/ehealth-esante/2005-pancanad-priv/index-eng.php>>.

- Johns Hopkins ACG Case-Mix System. 2003. Baltimore: The Johns Hopkins University Bloomberg School of Public Health. Retrieved January 5, 2009. <<http://www.acg.jhsph.edu>>.
- Manitoba Centre for Health Policy (MCHP). 2005. *Concept Dictionary*. "Ambulatory Care Sensitive (ACS) Conditions." Retrieved January 5, 2009. <http://umanitoba.ca/faculties/medicine/units/mchp/resources/concept_dictionary.html>.
- Ontario Ministry of Health and Long-Term Care. 2005. *Resource Manual for Physicians: Alternate Payments Plan Program*. Toronto: Author.
- Reid, R.J., R.G. Evans, M.L. Barer, S. Sheps, K. Kerluke, K. McGrail, C. Hertzman and N. Pagliccia. 2003. "Conspicuous Consumption: Characterizing High Users of Physician Services in One Canadian Province." *Journal of Health Services Research and Policy* 8(4): 215–24.
- Reid, R., L. MacWilliam, N.P. Roos, B. Bogdanovic and C. Black. 1999. *Measuring Morbidity in Populations: Performance of the Johns Hopkins Adjusted Clinical Group (ACG) Case-Mix Adjustment System in Manitoba*. Winnipeg: Manitoba Centre for Health Policy and Evaluation.
- Reid, R.J., N.P. Roos, L. MacWilliam, N. Frohlich and C. 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.
- Roos, L.L., V. Menec and R.J. Currie. 2004. "Policy Analysis in an Information-Rich Environment." *Social Science and Medicine* 58(11): 2231–41.
- 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., R.A. Soodeen, R. Bond and C. Burchill. 2003. "Working More Productively: Tools for Administrative Data." *Health Services Research* 38(5): 1339–55.
- Watson, D.E., S. Peterson, E. Young and B. Bogdanovic. 2009. "Methods to Develop and Maintain a Valid Physician Registry in Evolving Information Environments." *Healthcare Policy* 5(Sp): 77–90.