

# The Whole Is Greater Than the Sum of the Parts: Using Data Linkage and Cohort Designs to Create Data Synergy at MCHP

Le tout est plus important que la somme de ses parties :  
utilisation du couplage de données et des modèles  
de cohortes pour créer des synergies au Centre des  
politiques de santé du Manitoba

GREGORY S. FINLAYSON, BA, PHD (C)

*Research Scientist*

*Manitoba Centre for Health Policy*

*Department of Community Health Sciences, Faculty of Medicine*

*University of Manitoba*

*Winnipeg, MB*

LISA M. LIX, PHD, P.STAT.

*Centennial Chair & Associate Professor*

*School of Public Health, University of Saskatchewan*

*Saskatoon, SK*

LESLIE L. ROOS, PHD

*Distinguished Professor, Senior Research Scientist*

*Manitoba Centre for Health Policy, University of Manitoba*

*Winnipeg, MB*

**T**HE MANITOBA CENTRE FOR HEALTH POLICY (MCHP) HAS BEEN CALLED A “data-rich environment” (Roos et al. 2004). Indeed, the depth and breadth of the data justify this description – there are currently nearly 100 databases held in the Population Health Data Repository (“the Repository”) at MCHP, with some containing data back to 1970. And the number of databases continues to grow. However, it is not merely the volume of data that is important – rather, it is the synergy that is created when these databases are linked. Every data set at MCHP can be linked to others through a longitudinal registry of

virtually every resident of Manitoba. This comprehensiveness presents unique research opportunities that would not be possible otherwise. All personal identifiers are removed, and the Personal Health Identifier Number (PHIN) that is used to link all databases is scrambled to protect individual privacy. In addition, strict access control and approval processes are in place.

The administrative data at MCHP currently cover four distinct domains of provincial government services: healthcare, education, social services and justice. Within each of these groups are anonymized records of contacts Manitobans have with the particular service. For example, the healthcare database includes virtually all contacts individuals have with physicians, every in-patient hospitalization and every prescription drug that is dispensed. The education databases contain records of class grades, school enrolments and performance on standardized tests. Social services databases include, among others, data on social service recipients, public housing and the Early Development Instrument (EDI). Having the capacity to link these databases and follow individuals over time, with appropriate privacy controls, enables researchers to consider important policy questions. In this brief report, we will review some examples of the types of research that have been possible using these data.

It is important to be able to identify individuals who have certain health conditions, and particularly chronic conditions, for many purposes. Examples include examining the prevalence of these conditions in the population, examining trends in chronic disease and comparing the cost of healthcare for people who have these conditions to those who do not. Through linking the Canadian Community Health Survey with health administrative data, Lix and colleagues (2006b) have developed and validated algorithms that have made this possible. Another highly useful kind of analysis involves linking the health administrative data with clinical data contained in the provincial bone mineral density database. This linkage has allowed Leslie and colleagues (2008) to use these two data sets to develop models to predict fracture risk.

In addition to the linkage capability, these data may be used longitudinally. For example, Lix and colleagues (2006a) followed three cohorts over a three-year period to describe differences in residential mobility patterns for people with diagnosed schizophrenia, a matched cohort with no diagnosed mental disorders and people with diagnosed inflammatory bowel disease. The implications of the findings are important for housing and health policy analysts because mobility may result in discontinuities in health services delivery.

The administrative data housed in the Repository may also be used for multi-level modelling, for example, looking at the factors affecting educational achievement of children given their individual characteristics, the characteristics of their family and those of the neighbourhood in which they live. For example, being able to group individuals into their family units is a tremendous benefit for a variety of research projects: suicide and its effect on family members, and sibling research on the effects of childhood health and socio-economic status on educational achievement and labour force participation are but two examples. Such “nesting” of data has a number of benefits: it avoids the dilemma of the unit of analysis, it handles the lack of independence among observations (e.g., children in the same family) and it can be used to clarify confounding variables at both the within- and between-group levels.

These are but a few examples of work done at MCHP using linked and longitudinal

## The Whole Is Greater Than the Sum of the Parts

data. Of the approximately 92 major research projects taken on since 1991 for the provincial Ministry of Health, 79 would have been entirely impossible without the capacity to link databases – and all but eight used the registry of the provincial population. During this same time period, 491 papers using linked data have been published by MCHP and external investigators, all while maintaining privacy and access controls. The “atlases” produced at MCHP (Fransoo et al. 2009; Martens et al. 2003) always compare changes in rates between two points in time, and many other projects describe events over time. MCHP reports are used to inform policy, to monitor health and the healthcare system and to develop the requisite tools for researchers who need linked data to answer important research questions.

MCHP has been funded to expand the number of databases and to develop secure “remote access arms” that will make these data more accessible to others, while maintaining stringent protection of personal information through anonymization of databases, physical controls, and review and approval processes. These enhancements will further increase the opportunities for researchers to investigate policy-relevant questions.

### REFERENCES

- Fransoo, R., P. Martens, E. Burland, The Need to Know Team, H. Prior and C. Burchill. 2009. *Manitoba RHA Indicators Atlas 2009*. Winnipeg: Manitoba Centre for Health Policy.
- Leslie, W.D., J. Tsang and L.M. Lix. 2008. “Validation of Ten-Year Fracture Risk Prediction: A Clinical Cohort Study from the Manitoba Bone Density Program.” *Bone* 43: 667–71.
- Lix, L.M., A.M. Hinds, G. DeVerteuil, J.R. Robinson, J. Walker and L.L. Roos. 2006a. “Residential Mobility and Severe Mental Illness: A Population-Based Analysis.” *Administration and Policy in Mental Health and Mental Health Services Research* 33: 160–71.
- Lix, L.M., M. Yogendran, C. Burchill, C. Metge, N. McKeen, D. Moore et al. 2006b. *Defining and Validating Chronic Diseases: An Administrative Data Approach*. Winnipeg: Manitoba Centre for Health Policy.
- Martens, P., R. Fransoo et al. 2003. *The Manitoba RHA Indicators Atlas: Population-Based Comparisons of Health and Health Care Use*. Winnipeg: Manitoba Centre for Health Policy.
- Roos, L.L., V. Menec and R. Currie. 2004. “Policy Analysis in an Information-Rich Environment.” *Social Science & Medicine* 58: 2231–41.