

# Linking the prairies to the outback: What can New South Wales learn from the Manitoba Data Linkage Project?

Marian Shanahan and Kees Van Gool

CHERE, University of Technology, Sydney and Central Sydney Area Health Service, Level 6, Building F, 88 Mallett Street, Camperdown, NSW 2050, Ph: 02 9351 0900

## Abstract

Establishing a population-wide linked data project is a complex task and requires long term vision and investment. It is therefore necessary to ensure that in the event of NSW undertaking such an exercise, costly mistakes are avoided, and that our research capabilities are improved and subsequently, informed resource allocation decision making is enhanced. Using the experiences gained in the design and establishment of the Manitoba Population Health Information System (POPULIS), this paper will examine what value linked data can add to Australian and NSW health care policy debates. Starting with a description of POPULIS, we analyse the lessons learned from the Manitoba experience and assess whether its successes can be replicated in NSW. The aim of this paper is to draw out the strengths and weaknesses of the current NSW (and Australian) data collections and test whether linkage enhances the likelihood that some important policy questions can be addressed. The paper also proposes possible options for overcoming some identified limitations. Although linked data can be used for a variety of research efforts, the focus of this paper is its use in health services and health economics research. In particular, we examine how a future NSW data linkage project could assist in addressing issues of resource allocation, access, equity, utilisation and improve our understanding of the health care system's impact on health outcomes. Whilst there is little doubt that linked data has the potential to assist in our understanding of improving efficiency and equity within the health care system, the extent to which this occurs in reality depends on the validity and accuracy of the data and the links between data.

## Introduction

In the two years since the NSW Health Council made its recommendations for a state-wide unique personal identifier (UPI), the NSW Department of Health has made steady progress in developing and implementing a state-wide UPI. In addition, the Health Department is working on developing an electronic health record that will allow various health care providers to access, with the patient's consent, vital medical information (Group 2001). Having a UPI would also enhance NSW's ability to link its various administrative data bases. Work at the national level is happening concurrently with projects such as the HealthConnect program. It is envisaged that HealthConnect be a voluntary national health informa-

tion network that would allow information held in electronic records which would take the form of health summaries written by health care providers in a variety of settings.

Based on the NSW Health Department's publications and presentations on this topic, the emphasis appears to be on the use of health data linkage to improve clinical care and outcomes (Horvath and Kidd 2001; Williams 2001). Whilst there are obvious benefits to patients and clinicians from improved data, we would also like to emphasize the important role that data linkage can play in informing health care policy and in decision making.

The Manitoba administrative linked data project provides several examples of how such a system could address important policy issues, including comparing health expenditures, comparing waiting times, monitoring population health interventions, workforce planning and the development of clinical practice guidelines. The Medical Care Supplement June 1999, Vol. 37 No 6 provides a good overview of the variety of projects undertaken using the Manitoba administrative linked data.

Many of these studies show the extent to which data linkage can play an important role in policy development and analysis. Given the recent data linkage activities in Australia<sup>1</sup>, now would seem an opportune time to investigate what data linkage can offer, what are its requirements and what lessons can be learned from projects such as that operating in Manitoba.

This paper begins with an overview of the Manitoba project (additional description and references can be found elsewhere in these proceedings in the paper by LL Roos) and provides some information on the current status of data sets in NSW. Following this, some examples are used to explore the capacity of available data from NSW and Australia to answer some important policy questions. Potential weaknesses of current data collections are elucidated and suggestions on how these may be overcome are provided. For each example, comparisons are initially restricted to data collections already in place in NSW and Australia. Then, after having identified potential gaps and limitations in existing collections, we provide some suggestions on how to overcome these limitations.

In assessing NSW's ability to answer some important policy questions we also examine whether the potential linked data set meets the criteria set by Roos et al (Roos, Walld et al. 1996) in Manitoba. Roos identified four essential requirements for a successful linked data set that can answer policy relevant ques-

tions. The data set should :

- be linked by individual patient/consumer
- have the ability to conduct group or area level analysis
- have the ability to study policy questions at different levels from aggregate to local; and
- include data on health status, socio-economic status, supply and utilisation.

In the course of this discussion we acknowledge the necessity of protecting patient privacy and confidentiality and recognize that there are significant challenges in this area but we believe that with options such as encryption and establishment and enforcement of strict protocols, these issues can be overcome. Hence, for the purpose of moving this discussion forward we will assume that privacy and confidentiality can be adequately protected.

### Data in Manitoba and NSW

Manitoba, one of ten Canadian provinces, is situated in the geographic centre of Canada. All provinces of Canada, including Manitoba, are required by the Canada Health Act to provide public administration of the health care system, universal coverage of hospital and medical services, portability of coverage between provinces, and equitable access to hospital and medical care. In return for complying with the Canada Health Act, the federal government provides the provinces with money and some taxation rights.

Manitoba is also home of Manitoba Centre for Health Policy (MCHP) which houses the Manitoba Health Research Data Repository – the “Prairie” part of this paper. The Manitoba Health Research Data Repository is composed of a number of databases. The population registry is a key database of the MCHP. It includes a unique lifetime identification number for each person in the province, as well as information on residence by 6 digit postal code (postal code may be as small as a city block or as large as a rural municipality), marital status, date of birth, family relationships and sex. The Repository, which has information on Manitoba residents going back to 1970, includes all inpatient hospital discharge abstracts, abstracts on outpatient procedures, medical doctor claims for payment, birth and deaths registry, pharmaceutical data, nursing home data, disease registries, information on medical providers (age, sex, location and type of practice and any specialist qualifications) and home care data. In addition, in recognition of the Data Repository uniqueness, stringent protocols for protecting data confidentiality have been established. Such protocols have played an important part in other agencies agreeing to deposit data with MCHP. The agencies include the Office of Vital Statistics, and Cancer Care Manitoba. Since its inception special sub-files have been created, incorporating de-identified data from other research, for example, from the Manitoba Longitudinal Study on Aging.

The use of a consistent set of identifiers (with identification numbers of both patients and doctors scrambled to ensure confidentiality) permits researchers to build histories of individuals

across files. For instance, individuals who are discharged from hospital can be linked to the medical claims file in order to determine if adverse events are being treated in doctors’ offices. Individual enrolment in the Manitoba health insurance system, migration into and out of the province, and mortality can be traced from 1970 onward using the longitudinal population registry.

The Data Repository has not always had its current structure. It originated as a database to collect longitudinal physician utilisation data but has grown over the years with the addition of various databases – sometimes in a planned way and other times fortuitously, for example, as researchers undertook a project or when the existence of new database became known. The use of a new database always requires careful scrutiny and a steep learning curve for researchers. Some databases are only used for specific projects, with no ability for ongoing linkage but most are available for future use.

The Manitoba administrative linked data has been used to answer many questions, and to provide information for policy makers. Examples of what can be done include:

- studies of the population based system –using individual identifier can assess access to health care
- provide data on supply of services (hospital beds, nursing home beds, GPs, specialists)
- provide information on access to new technology – ie surgery done using laparoscopies versus traditional methods
- cross sector analyses using dollars as the measurement

However, the lack of individual data on health status, employment status, education and other demographic information has meant that aggregated data available from census data have been used to compile socio economic information at a small area level. Such information can be used to examine patterns of illness in the population, and how people in different areas of the province use health care services.

It is clear that administrative data has many strengths but it is important to note that these data are insensitive to clinical issues such as severity of pain, laboratory results, limitations of the co-morbidities included in the data and without individual (or family) demographic information make it hard to research many issues.

As is the case in most Australian States and Territories, the NSW health care system has an extensive list of data collections (Table 1). They include acute inpatient care data collections, registries for specific diseases such as cancer, as well as surveys such as the NSW Health Survey. In recent years, more information has been sought from non-acute care and community care sectors through the development of collections such as SNAP (sub- acute and non acute patient classification) and HACC (Home and Community Care) data collection.

Table 1 provides a partial list of some of the NSW data sets, national and non-health department collections that are currently available. Rather than list all available data sets, our aim is to present a range of data that is available. The questions we are focusing on are:

- will the introduction of a unique personal identifier (UPI), as proposed by NSW Health, answer the type of policy questions addressed through the Manitoba linked data project? And
- will the use of the UPI ensure that policy questions of a health economic nature can be answered?

NSW (examples of existing but unlinked data bases)	Manitoba (examples of linked/able databases)
<ul style="list-style-type: none"> <li>- Hospital data sets – ISC</li> <li>- SNAP</li> <li>- Birth/death/marriage registry</li> <li>- MBS/ PBS</li> <li>- Cancer registry</li> <li>- NSW Health &amp; Older Persons Survey</li> <li>- Women’s Health Aus.- Newcastle</li> <li>- Private Health Insurers</li> <li>- NSW Midwives Data Collection</li> <li>- Neonatal ICU Data Collection</li> <li>- Waiting time data collection</li> <li>- Child Dental Health Survey</li> <li>- Labour Force Data Collection</li> <li>- Home and Community Care</li> <li>- Emergency Dept Data Collection</li> </ul>	<ul style="list-style-type: none"> <li>- Registry – Unique identifier, DOB, sex, marital status, postal code</li> <li>- Inpatient hospital discharge abstracts (all patients)</li> <li>- Physician payment data base</li> <li>- Outpatient procedure abstracts</li> <li>- Personal care home data</li> <li>- Pharmaceuticals</li> <li>- Vital Statistics Registry</li> <li>- Cancer Care</li> <li>- Public Access Census Data</li> <li>- Family Services data</li> <li>- Individual disease registries</li> </ul>

**Table 1 Examples of NSW and Manitoba Data Collections**

#### Possible policy questions

Linked administrative data opens up the possibility of answering important policy relevant questions. Examples of such policy relevant questions are listed in Table . The next section of this paper then use case studies to determine whether or not NSW would be in a position to answer these questions if a UPI was able to link all data collections identified in Table 1.

- Do different incentives lead to more effective use of services?
- What are the economic and health costs of waiting times for surgical procedures?
- Do changes in funding mix alter the equity of access to the system?
- Do changes in funding mix alter the efficiency of the system?
- What are the economic and health costs of medical errors?
- What are the resource implications of an ageing population?
- Is health care supply congruent with health needs?
- Does workforce/healthcare substitution impact on health outcomes and health expenditure?
- Can we obtain more accurate information on risk adjustment?
- Does moral hazard exist amongst privately insured patients?
- Could compliance with clinical guidelines result in greater efficiency (and/or better outcomes) in the health care system?

**Table 2 Possible research questions if linked data was available**

#### Do different incentives lead to more effective use of services?

In recent years there have been a number of initiatives in Australia using altered incentives to either providers or consumers of care that have aimed to encourage (or discourage) certain behaviour. Examples include the practice incentive payment, the general practice immunisation incentive, special hospital funding to reduce waiting lists and the 30% private health insurance rebate. In the Manitoba context, Brownell (Brownell, Roos et al. 2001) documented how administrative data bases could be used to monitor access to beds, high profile surgeries (eg open heart surgery, hip replacements, and cataracts), and quality of care following sizable funding cuts to hospitals. In the NSW context the addition of the UPI, the use of private health insurance data, and linkages to the HIC data, there would be sufficient information to explore some of these questions using individual level data. However the obvious limitations are the lack of diagnosis in the MBS data and the challenges of accessing Private Insurance data at the individual level. These limitations would be important if the economic impact of shifts in service provision between private and public sectors were to be explored.

#### What are the economic and health costs of extended waits for surgery?

The information on waiting lists for elective surgery is fairly comprehensive. The NSW Department of Health provides web based information for patients and referring general practice on waiting times by hospital, doctor and procedure. It reports these in terms of the length of waiting time that 50<sup>th</sup> and 90<sup>th</sup> percentile of patients experienced (see: [www.health.nsw.gov.au/im/ims/wtc/index.htm](http://www.health.nsw.gov.au/im/ims/wtc/index.htm)). This information is derived from the NSW Health Waiting Times Collection, which is a census of all admitted patient services in public, private and psychiatric hospitals. A summary of information collected as part of the waiting times data is shown in Table 3.

System Generated Items	Demographics	Wait List Booking Details
To provide information on the facility at which the patient is to be admitted a record of the waiting list entry. Data items include: -Reporting facility -Booking identifier -Record update date	To provide information on the client's demographic details. Fields include: -Date of birth -Sex -Medical record number -Client's address	To provide information on dates, length of waiting times, procedures, urgency and doctor. Data fields include: -Date listed on list -Indicator of procedure -Registration number of doctor -Specialty of doctor -Intended length of stay -Accommodation status (private, public, workers compensation etc) -Provisional admission date -Planned admission date -Planned procedure date -Date and time of delay -Reason for delay -Date and time of decline -Reason for decline -Listing status -Urgency status

**Table 3 NSW Waiting Times Collection Data fields**

Source: NSW Health

The primary reason for the waiting data collection is to manage lists effectively, that is, to ensure that patients are prioritised according to clinical need. However, waiting lists are often seen as an indicator of insufficient resources. This view is promulgated by media commentators and politicians alike. If this view of waiting lists is accepted then the information requirement will be different to the current data collection.

If extended waiting times are seen as an indicator of insufficient resources then the research question that ought to be asked is what are the costs and benefits of reduced waiting times. To address this question, it would be necessary to measure the health effects of extended waits as well as any additional health care costs incurred during a patient's wait. In addition, the opportunity costs associated with redeploying resources from other sectors in order to reduce waiting times would need to be measured. This is not an easy area to research, especially given the difficulty in measuring waiting times accurately.

As shown in Table 1, waiting lists data could potentially be linked to other existing data collections. The aim of such linkage would be to measure the resource and health impacts of waiting times. For example, the Commonwealth's Department of Health and Aged Care Medicare and Pharmaceutical data bases (MBS and PBS) could provide information on medical service utilisation and pharmaceutical consumption prior, during and following a period of waiting for elective surgery. Similarly, emergency department data collections could provide information on ED utilisation during this period.

Linking health outcomes to the data file is more problematic. Whilst deaths could be added (as well as additional procedures/diagnosis which may arise from waiting), there is currently no routine data collection which seeks to measure the potential loss in quality of life of individuals whilst on a waiting lists or after having experienced extended waits.

Despite a limited range of proxies being available for health outcomes (eg data from the hospital morbidity data set), this area remains a weakness of Australian health data.

**Do shifts in funding mixes alter the equity of the system?**

In Australia, neither a fully privatised nor fully tax funded system seem to be acceptable. Health care policy over recent decades has focused on the proportion of health expenditure funded through tax, out-of-pocket or insurance premiums. It is important to understand the impact of such shifts in the funding on the equity of health care financing. Unfortunately, the ability to measure the equity impact of altering funding mixes is limited. The next section of the paper examines Australia's current

ability to measure these changes and evaluates the potential for linked data to contribute to increasing the accuracy of such measurement.

An example of an empirical question which could be asked is what did the recent PHI initiatives have on the distribution of the burden of finance?

Table shows that since the introduction of the 30% rebate, the proportion of health expenditure funded through taxes rose from 74.6% to 75.4%. This rise is entirely due to tax expenditure on the PHI rebate. It seems paradoxical that the rise in PHI coverage has resulted in a greater burden of health expenditure falling on the taxpayer, rather than PHI premiums. This is because the rebate is tax funded and paid to both new and existing members.

	Taxes	Out of Pocket	Premiums*	Total
1996-97	71.7%	17.1%	11.2%	100%
1997-98	74.0%	15.9%	10.1%	100%
1998-99	74.6%	16.7%	8.7%	100%
1999-00	75.4%	17.1%	7.5%	100%

**Table 4 Funding contribution by source of funds – 1996 to 2000**

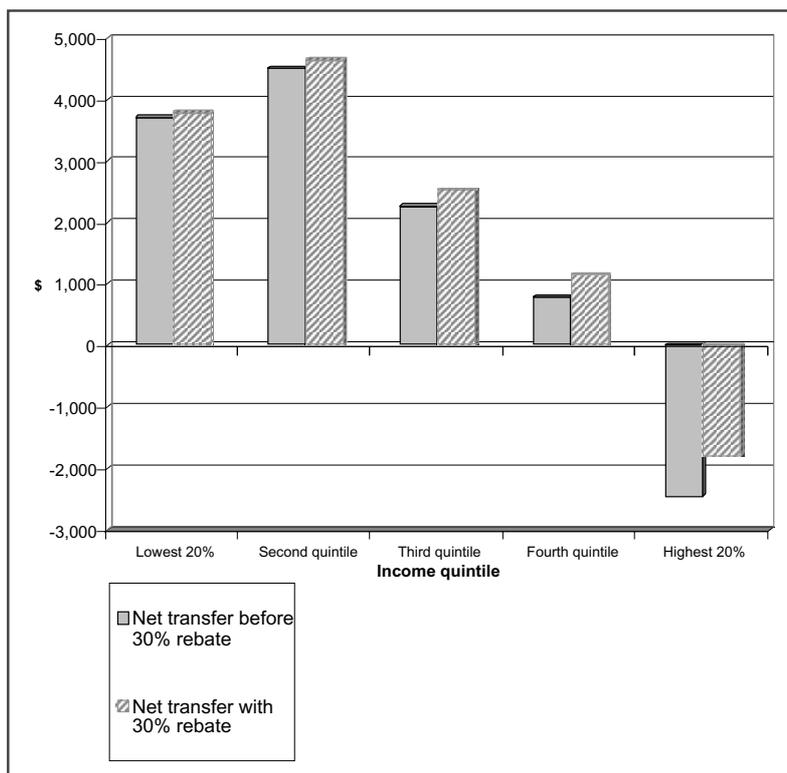
Source: (Rice and Smith 1999)

\* excludes compulsory motor vehicle third party and workers compensation insurance

Following on from the first question, a second empirical question is to determine who might be paying for the funding mix change and who are the beneficiaries? Through its Household Expenditure Survey, the Australian Bureau of Statistics (ABS) estimates the amount of publicly financed health benefits received as well as the total amount of taxes paid. It reports

these by household income quintile.

Figure shows the publicly financed health benefits (defined as hospital care, medical clinics, pharmaceuticals and other health benefits) received by households in each income group minus the taxes they paid<sup>2</sup> (Aynonamous 2001). The figure shows the redistributive effect of Australia's public health care system. The solid bars show the redistributive effect of the system prior to the introduction of the 30% PHI rebate. The poorest group is a net beneficiary of the public health care system by approximately \$3,700 per household per year. For the richest group, the amount of health care taxes paid exceeds the amounts of health care benefits received by approximately \$2,450 per household per year.



**Figure 1 Net transfer to or from income quintile – before and after the introduction of the rebate.**

Figure 1 also shows net transfers following the introduction of the 30% PHI rebate. The hatched bars indicate how the rebate has altered the redistributive impact of the publicly financed health care system<sup>3</sup>. The 30% rebate provides a tax break for all households who take out PHI, resulting in the richest groups receiving a larger tax-break than poorer groups.

Whilst Australia already collects some essential data that would help answer issues surrounding equity of financing, some assumptions have been made in conducting the above analysis. For example, the ABS estimates the health benefits received by a household according to average rates of health care utilisation and adjusts these according to group's age, sex

and residence composition. Thus benefits are not calculated on the basis of actual health care use. Actual benefits data by income quintile would be a preferred measure. Furthermore, the taxes paid and attributed to households by the ABS only accounts for around 52% of Australia's total government revenue. Important taxes such as company taxes, some indirect taxes and other revenue measures are not attributed to households. The tax incidence may therefore be very different to the figures shown above.

## Conclusion

This paper has attempted to highlight the data needs for informed policy debates in the areas of incentive payments, waiting lists and financing of health care. The three examples discussed above illustrate that these types of questions are partly answerable using linked data. However, in some instances it may be necessary to collect additional data, especially in relation to outcome data.

As with clinical practice, where evidence based decision making is becoming increasingly important, decision makers require scientific information of the highest possible quality to inform those decisions. Never the less, even if the data is perceived as being less than ideal in terms of addressing policy questions, it is only by attempting to use it that its limitations and strengths will be recognised. The examples used in this paper illustrate some key issues that require attention while the capabilities of the linked data sets are being established. Such issues include collection of and access to information regarding diagnosis, outcomes of care and the quality of data from private and public sources. There is a current opportunity to add to the research capacity of Australia's health information infrastructure. It is therefore timely to consider the issues raised in this paper now.

## Endnotes

1 For example, data linkage projects in place in WA and Qld, infrastructure development in NSW as well as the Commonwealth Department of Health and Ageing's investment.

2 Given that total government health care spending in 1998/99 was just under \$36 billion and total government tax revenue was \$180 billion, approximately 20% of taxes go towards publicly financed health expenditure. This proportion was used to estimate the amount of taxes paid towards health care by each household income group.

3 The 1998 ABS Health Insurance Survey shows that only 20% of the poorest income group take out private health insurance, whereas 76% of the richest households are privately insured. Quotes from Australia's two largest private insurance companies were obtained and average premiums ranged from \$1,214 to \$3,175 per household per year (before the rebate). It is assumed that richest households select higher cost premiums and poor households will select lower premiums.

**References**

- Aynonomous (2001). Co-ordinated Care: funds pooling not the be all and end all. *HealthCover*. 11: 18–21.
- Brownell, M.D. and N.P. Roos, et al. (2001). "Monitoring health reform: a report card approach". *Social Science & Medicine* 52(5): 657–670.
- Group, E. W. (2001). *A NSW Health Strategy for the Electronic Health Record*. Sydney, NSW Health.
- Horvath, D. and M. Kidd (2001). *Information Management and Technology. Health Working as a Team*, NSW Health.
- Peacock, S. and Segal, L., et al. (2001). *Predicting the Expected Costs of Health Care: Methods for Risk Adjustment in Health Care*, Centre for Health Program Evaluation.
- Rice, N. and Smith, P. (1999). *Approaches to Capitation and Risk Adjustment in Health Care*. York, University of York, Centre for Health Economics.
- 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. Walld, et al. (1996). "Record linkage strategies, outpatient procedures, and administrative data". *Medical Care* 34(6): 570–82.
- Williams, P. (2001). *Electronic Health Records. Health Working as a Team*, NSW Health.