

A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system

C D'Arcy J Holman, A John Bass, Diana L Rosman, Merran B Smith, James B Semmens, Emma J Glasson, Emma L Brook, Brooke Trutwein, Ian L Rouse, Charles R Watson, Nicholas H de Klerk and Fiona J Stanley

Abstract

Objectives: The report describes the strategic design, steps to full implementation and outcomes achieved by the Western Australian Data Linkage System (WADLS), instigated in 1995 to link up to 40 years of data from over 30 collections for an historical population of 3.7 million. Staged development has seen its expansion, initially from a linkage key to local health data sets, to encompass links to national and local health and welfare data sets, genealogical links and spatial references for mapping applications.

Applications: The WADLS has supported over 400 studies with over 250 journal publications and 35 graduate research degrees. Applications have occurred in health services utilisation and outcomes, aetiologic research, disease surveillance and needs analysis, and in methodologic research.

Benefits: Longitudinal studies have become cheaper and more complete; deletion of duplicate records and correction of data artifacts have enhanced the quality of information assets; data linkage has conserved patient privacy; community machinery necessary for organised responses to health and social problems has been exercised; and the commercial return on research infrastructure investment has exceeded 1000%. Most importantly, there have been unbiased contributions to medical knowledge and identifiable advances in population health arising from the research.

INFRASTRUCTURE TO ENABLE research has become a major planning focus in many countries, including the United States, Canada, the United Kingdom, members of the European Union and Australia.¹ Several countries have undertaken national audits of their stock of research infrastructure, which in the health sector includes biomedical laboratories, imaging equipment, biobanks and supercomputers.¹

Concepts about enabling national infrastructure for research on population health and wel-

C D'Arcy J Holman, MB BS, MPH, PhD, FACE, FAFPHM, Professor of Public Health

James B Semmens, BSc, DipEd, MSc, PhD, Director

Emma J Glasson, BPsych, BSc(Hons), PhD, Research Fellow

Emma L Brook, MPH, Research Associate

Brooke Trutwein, BHLthSc, Research Officer

Centre for Health Services Research, School of Population Health, The University of Western Australia, Perth, WA.

A John Bass, BSc(Hons), MSc, PhD, Senior Research Fellow

Ian L Rouse, BSc(Hons), GradDipHlthSc, PhD, Professor

Charles R Watson, BSc(Hons), MB BS, MD, FAFPHM,

Professor

Division of Health Sciences, Curtin University of Technology, Perth, WA.

Diana L Rosman, BSc, GradDipComp, MMedSc, Manager, Health Information Linkage Branch

Merran B Smith, BSc, BA, MSc, PhD, PGradDipMan, General Manager, Information Collection and Management,

Western Australian Department of Health, Perth, WA.

Nicholas H de Klerk, BSc, MSc, FEEd, PhD, Professor

Fiona J Stanley, AC, MB BS, MSc, MD, FFPHM, FAFPHM,

FFCCH, FRACP, FRANZCOG, Director

Telethon Institute for Child Health Research, Perth, WA.

Correspondence: Professor C D'Arcy J Holman, Centre for Health Services Research, School of Population Health, The University of Western Australia, 35 Stirling Highway, Perth, WA 6009. dholman@meddent.uwa.edu.au

Aust Health Rev 2008; 32(4): 766-777

What is known about the topic?

Data linkage systems can provide infrastructure support for health services research. However, comprehensive systems of pre-linked health data covering entire populations are uncommon.

What does this paper add?

The Western Australian Data Linkage System illustrates how a comprehensive, population-based system can be designed and implemented. The benefits go beyond anticipated scientific outputs to include improvements in health policy and practice, developing communities, conserving patient privacy, economic stimulation of the research sector, improving the quality of databases and making research less costly.

What are the implications for practitioners?

Given the extent of benefits and relatively low costs, health data linkage systems may become normal infrastructure for health services research in nations like Australia within the next decade.

fare have existed for over half a century. In 1946, Dunn proposed the idea that each person creates a "Book of Life", starting with birth, ending in death and composed of the records of important health and social events.² Dunn developed the concept of collating the records into a personal file and named the process record linkage. He predicted that statistical analysis of linked records would be useful to health and welfare agencies in evaluating their service programs. He suggested also that linking files of data would establish the accuracy of the source records.²

All of Dunn's predictions have been proven to be correct in theory, albeit in practice comprehensive national systems of linked health and welfare data remain uncommon. Successful implementation demands high levels of community capital and inter-agency cooperation before the many diverse stakeholders will relinquish their immediate interests in favour of a more futuristic and public-spirited common vision. It was not surprising, therefore, that the international review by Roos et al described only a handful of information-rich environments worldwide, which have been constructed from the linkage of multiple, large, population-based, administrative data sets.³ They included the Manitoba Population Health Information System,⁴ Oxford Record Linkage

Study,⁵ Scottish Record Linkage System,⁶ Rochester Epidemiology Project,⁷ the Centre for Health Services and Policy Research in British Columbia,⁸ and the Western Australian Data Linkage System (WADLS).⁹

This report describes the strategic design, the steps to full implementation and the outcomes achieved by one of the world's few comprehensive, population-based data linkage systems, the WADLS, which has linked over 30 health data collections for a population of 2.0 million living in the western one-third of the Australian continent. Our report coincides with the tenth anniversary of the WADLS, which was instigated in 1995 and first became fully operational in 1998.

Developmental history

Hobbs' legacy 1970–94

We define data linkage as "the bringing together from two or more different sources, data that relate to the same individual, family, place or event". The potential for a systematic approach to this activity in Australia was first proposed by Hobbs in 1970.¹⁰ Based on his experience in Oxford, he suggested the introduction of medical record linkage studies on a national or state-wide basis and the implementation of a pilot linkage scheme in Western Australia.

Data linkage schemes began in WA from the late 1970s, but were limited to specific purposes. Due to the foresight of Hobbs and others,¹⁰ WA possessed high quality personal identifiers for morbidity, mortality and health service data sets, covering the entire population back as far as 1966. Researchers from the Telethon Institute for Child Health Research created a WA Maternal and Child Health Research Database, linking information from midwives' notifications, birth registrations, death certificates, hospital inpatient morbidity, birth defects and cerebral palsy data to perform epidemiological perinatal and paediatric studies.^{11,12} The WA Road Injury Database linked state-wide police, hospital and death records of road crash casualties.¹³ However, lack of political will, and limitations in resources and computing

power precluded the development of full, population-based data linkage in WA until 1995, when an infrastructure grant to establish a data linkage unit was awarded by the WA Lotteries Commission to The University of Western Australia.⁹

Lotteries revolution 1995–98

Several factors happily conspired to create a window of opportunity where previous attempts at comprehensive linkage had failed. Firstly, there had been movements of senior personnel with public health backgrounds between the government and academic sectors, creating a sense of community that transcended agency boundaries. Secondly, a past decade of health promotion campaigns had raised the profiles of health research leaders which, together with the acceptance achieved by the Busselton Health Studies and other large-scale epidemiological projects in vascular diseases, cancer and asbestos-related diseases, served to engender community trust. Thirdly, the new Lotteries Commission funding program for infrastructure overcame past obstacles, whereby data linkage had been excluded from “research” funding by granting bodies and from “services” funding by government health authorities. Fourthly, an agreement to accommodate the WA Data Linkage Unit (DLU), consisting of university staff and equipment, within the State Health Department in proximity to the core data sets overcame earlier privacy concerns and gave ownership to both sectors, while drawing on the diversity of government and academic financial systems as a resource for business flexibility. Finally, advances in the capacity and affordability of computing equipment were also an enabling factor.

The DLU focussed initially on creating data manipulation tools and designing its system concepts and structures. This phase was followed by a staged construction of the master linkage key, designed to optimise the early availability of linked data for demonstration projects of interest to health planners and clinicians.^{14–17} By 1998, the objective of the Lotteries grant to link 6.5 million births, deaths, hospital separations, midwives’ notifications of confinements, cancer noti-

fications and mental health service contacts, covering the 15 years 1980–95, had been achieved at a cost of about A\$1 million.⁹

WADLS consolidation 1999–2000

By 1999 the State Health Department had become sufficiently confident in the DLU to step in as principal funder in about a 3 : 1 mix with the academic sector. The DLU, now with mixed university and government staffing, came under joint administration by the funding partners and a consumer representative was appointed to the Management Committee. A policy of access by any *bona fide* researcher with institutional ethics committee approval and a WA collaborator accelerated the number of linked data requests. Today, the number of linked data requests averages 60 annually. Open access and a commitment to translation of research results into policy, practice and public education outcomes engendered participation by senior clinicians, notably the Royal Australasian College of Surgeons,¹⁵ consumer advocates and a broad constituency of researchers within and outside the state. In 1999 *Health+Medicine* commenced as an 8-page weekly supplement in WA’s major morning newspaper, and a constructive relationship developed, ensuring regular coverage of research results arising from linked data as well as several articles explaining the WADLS to the general public.^{18–20} In the same year, the team appointed the first consumer advocate to take up a position in an Australian university. The independence of the appointee was assured through their selection by the WA Health Consumers’ Council. The advocate’s brief was to increase community and consumer participation in research based on linked data, an objective spearheaded by the Duty to Care research program on people with mental illness.^{21,22}

By 2000, links to all available WA administrative data back to 1966 were incorporated into the master linkage key, and a new phase of monthly updates covering the core data sets was established. The master linkage key was extended to cover many data collections managed by health service agencies other than the WA Department of Health, as well as population-based research

databases held in the academic sector. The motivation for these extensions was often that the data custodians desired access to linked hospital morbidity, cancer and mortality data for follow-up of their cohort or survey participants.

Developments from 2001

Twenty-first century growth in the WADLS has been driven not only by the periodic updates of links to data sets already participating in the system, but even more so by expansions of the system across several new fronts.

A feature of the Australian health system is the dual responsibilities of the state and Australian Commonwealth governments for different facets of financing and service delivery. A result is that significant data relating to the health of Western Australians, including residential aged care and claims to the national Pharmaceutical Benefits Scheme (PBS) and Medicare Benefits Scheme (MBS) are held by the Australian Government. The absence of these data in the original WADLS was its most conspicuous shortcoming. Groundwork in the form of dialogue and relationship building for cross-jurisdictional links to the commonwealth data sets had commenced in 1998, facilitated by positions held by team members on national boards and movements of senior staff between government agencies. A pilot project to link the hospital, death, MBS and PBS records of 148 000 diabetic patients was approved in 2001. A best practice protocol was developed that took into account privacy concerns and enabled data custodians to retain control over the uses of the information in their care.²³ This protocol, which was reviewed without objection by the Australian Privacy Commissioner, has since been adopted widely in linkage projects across Australia. The success of the diabetes pilot project and wider recognition of the significance of the data linkage developments in WA led to a commonwealth-state agreement in 2003 for an all-population cross-jurisdictional linkage covering MBS, PBS and aged care data back to 1990 and prospectively. The first linked aged care data were released for research in 2005. In the same year, links to available MBS and PBS data were finalised

and data were later released for research early in 2007.

A further collaboration between the Telethon Institute for Child Health Research and five state government agencies is built on the privacy preserving best practice linkage protocol. This collaboration focuses on vulnerable children and youth and their developmental trajectories in relation to education, health, disability and crime, with intimate involvement by policy makers in the analyses by researchers.

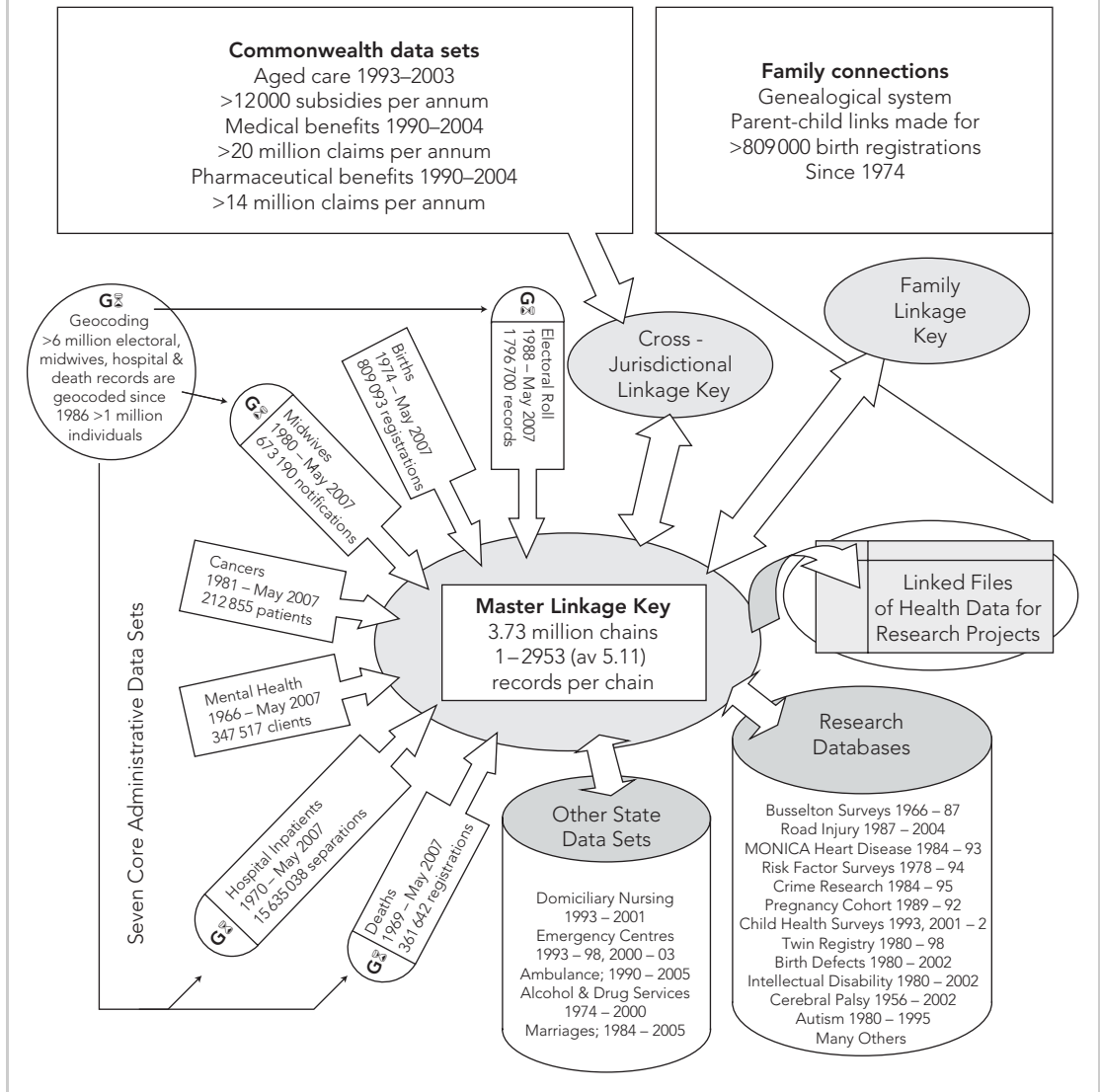
Genealogical linkage

A new breakthrough occurred in 2002 with the instigation of the Family Connections Genealogical Register to store family links between genealogically-related individuals.²⁴ It followed a successful 18-month community awareness program, Genomics, Society and Human Health, which explained the research potential to the general public. Stage I of the register commenced in 2003 and has created a single inter-generational family linkage based on electronic birth, marriage and death registrations back to 1974 for births and 1984 for other registrations. Stage II will encode and link earlier registration records once these are computerised; and in stage III a public appeal will seek to fill remaining gaps. One aim is to identify family links of all nuclear families that include at least one offspring who was born in WA from 1950 onwards. Family links will be used in combination with links to health records to support population-based genetic and human genome research with emphasis on characterisation of gene-gene and gene-environment interactions. The latter will include the utilisation and outcomes of genetic services. Links already exist to 200 000 DNA specimens stored from representative community health surveys, and proposals are now being championed independently of the WADLS to pilot a population-based WA human genome survey.²⁵

Other linkage innovations

Links were made to the state electoral roll, as a seventh core data set, to improve the capacity for accurate longitudinal censoring for outward

I Scope of the Western Australian Data Linkage System as at May 2007



migration, selection of control groups and to support studies of illness-related migration.²⁶ Geocoding of residential addresses on hospital morbidity, midwife notifications, mortality and electoral roll records was introduced to add spatial references to the data. This facilitated links to environmental factors and improved the accuracy of social disadvantage and remoteness indices.²⁷ It enabled the geographic distribution of patients

to be defined, and supported research into socio-demographic determinants of health service utilisation and outcomes.²⁸⁻³¹ A further enhancement has been agreement from government departments to create links to data sets outside the traditional boundaries of health, such as information held in the education, community services and juvenile justice domains for use in child health research.

An effect of these developments and enhancements was to diversify the sources of funds to operate the DLU, such that substantial funding for activities other than linkage of the core data sets now arises from sources other than the State Health Department. An effective business plan and the adoption of fee-for-service arrangements augur well for long-term financial viability.

Widespread appreciation of the social, scientific, economic and privacy-protection benefits of the WADLS led to a successful application to the WA Government for Centre of Excellence support commencing in 2005. The broad base of advocacy available to the WADLS by this stage was embodied by over 50 letters to the government, some referring in detail to positive experiences in using the system, written by consumer groups, senior clinicians, service agencies and laboratory-based as well as population health research groups. The Centre of Excellence has broadened the team's focus to the whole of Australia and the Indo-Pacific Region, especially with respect to expert advisory and training services. The team hosts Australia's only national coursework program on the analysis of linked health data and completed a review and recommendations for the Sax Institute that led to the establishment of the Centre for Health Record Linkage (CHeReL) in New South Wales.³²

Present scope of the Western Australian Data Linkage System

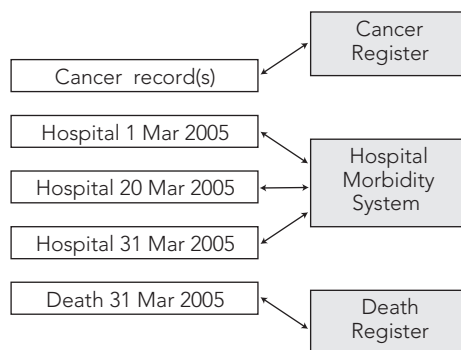
The WADLS in 2008 is a complex, multi-set system for the creation, storage, update and retrieval of links between health and welfare-related data. It uses computerised probabilistic matching and clerical review to create a dynamic master linkage key between over 30 population-based administrative and research health data collections in WA (population 2.0 million). The linkages mean that the historical population of recent decades (3.7 million individuals going back as far as 1966) can be used for research into all major diseases, their risk and protective factors and the utilisation and outcomes of health services. The scope of the system as at May 2007 is shown in Box 1.

The DLU, recently renamed the Health Information Linkage Branch, now employs around 20 staff, consisting of linkage, systems and client support groups with specific responsibilities for quality control and liaison with data custodians and researchers.

A key aspect of the design of the WADLS is the separation of linkage-related processes from those operating on sensitive clinical and service data. The WADLS is neither a centralised data repository nor a distributed virtual data warehouse. Rather, the information retained in the system consists of pointers to the source data. Examples of such data are a single record in the hospital morbidity data system; a set of records for a single patient in the cancer register; or a single record in the death register. The elements for a deceased individual might therefore include a cancer registry entry, three hospital separation records and a death registration, as depicted in Box 2. This introduces the concept of a *chain of links*, in which pointers to health-related data elements are ordered chronologically for each person. Thus one chain points to all of the known events in the life of a single person, much like Dunn's Book of Life.² The chain itself does not contain any clinical information.

The inclusion of pointers to residential aged care, MBS and PBS benefits in the WADLS has increased the length of most individual chains.

2 Example of a chain of five links for a single individual serving as pointers to separate health episodes or other events



However, unlike hospital records representing discrete episodes of care, many data sets acquired under the best practice protocol consist only of personal identifiers and their variants, making the chains shorter on average than the actual number of health records that they represent.

Access to linked data from the WADLS follows a well documented process that conforms to the National Health and Medical Research Council's guidelines approved under ss 95 and 95A of the *Privacy Act 1988* (Cwlth) and to the current *National Statement on Ethical Conduct in Human Research* published in accordance with s 90 of the *NHMRC Act 1992* (Cwlth). Access is reviewed on a project-by-project basis with the first necessary step being approval of the project by the applicable institutional ethics committee. Separate approvals must also be obtained from all of the agencies that are a custodian of one or more of the data sets included in the project. Before analysts receive linked health data they are first required to agree in writing to a stringent set of conditions regarding data security. These include undertakings not to use the data for anything other than an approved purpose, not to transfer data to a third party and not to attempt to link any of the files provided to an external data set. Users of the facility have been a diverse community, with academic researchers and analysts in government and service agencies almost equally represented (56% and 44% respectively in 1995–2003).

Applications

During the first ten years of full operation, the WADLS has been used to provide data linkage and related services to researchers in academic, community, health industry and government organisations, whose work has yielded over 400 distinct studies. The outputs include over 250 scientific journal publications and over 35 graduate research degrees completed or in progress. A detailed analysis of projects and publications in 1995–2003 is available.³³

Health services and epidemiologic research

Applications of the linked data in health services research, health service management and epi-

demiological research have included studies in the following areas:

Health services outcomes

The WADLS has supported a successful Quality of Surgical Care Project established by the Royal Australasian College of Surgeons,¹⁵ and other safety research such as surveillance for adverse effects of new technology such as minimally invasive surgery or assisted reproduction;^{34,35} economic outcomes such as the impact of survival of low birth weight neonates on cerebral palsy and health care costs;³⁶ or efficacy and prognostic research such as improvements in survival after cancer treatments.³⁷ By enabling the evaluation of clinical outcomes, data linkage has promoted clinical best practice and provided an opportunity to undertake analyses of new and existing treatments involving medical devices, drugs and surgical procedures.³⁸

Health services utilisation

This research has included studies of patterns of care in clinical populations to identify differences between subgroups, time trends or causal relationships in the processes of service delivery. The examples include inequities in access to breast reconstruction in women having undergone mastectomy,³⁹ factors causing trends in Caesarian sections,⁴⁰ increasing re-admissions driven by new technology in the treatment of ureteric calculi,³⁷ and the impact of Australian government financing policies on the use of private health insurance.^{41,42}

Aetiologic research

The WADLS has been used to identify health events during the longitudinal follow-up of exposed cohorts, illustrated by a cohort-cross-over study of travellers' thrombosis in 9.4 million passengers arriving on international long-haul flights, which has provided the most precise assessment of this health risk available to date;⁴³ and by a risk assessment used to establish national occupational exposure standards for crystalline silica.⁴⁴

3 Example of health services research that would have been impossible without pre-existing, systematic data linkage infrastructure

Patient-initiated switching between private and public inpatient hospital payment classifications. *Aust N Z J Health Policy* 2005; 2: 12⁴¹

Nature of the research problem: (1) How to measure the extent of switching between private and public payment classifications during consecutive inpatient hospital episodes. (2) How to measure the effects of risk factors for switching, including initial private/public classification, diagnostic groups, age and sex, time interval between episodes and national private health insurance policy era (eg, pre and post "life-time health cover").

Conceptual model to answer the questions: A "couplet methodology" was devised, whereby units of observation were the sequential pairs of inpatient hospital episodes in the same patient (= couplets). Thus a patient with four separate hospital inpatient episodes had three couplets. Couplets were either concordant or discordant in terms of payment classification. Risk factors for discordance were evaluated using clustered logistic regression and measured attributes of the couplet elements (eg, whether the first hospital episode was private or public; whether the couplet pair straddled implementation of the "life-time health cover" policy).

How data linkage made the research possible: Without systematic data linkage infrastructure, linking the entire 1980–2001 hospital morbidity data set for the population of Western Australia, it would have been impossible to identify and extract the 6376472 hospital episode couplets in this study. It would have also been impossible to identify the groups of couplets arising in the same individual so as to account for clustering error in the regression analysis.

Disease surveillance and needs analysis

This has included the collation and cross-validation of information from multiple sources to identify first-time incident events or prevalent clinical populations for needs analysis.^{16,17,22,45} Among the latter have been studies of the increasing "active prevalence" of cancer due to wider use of chemotherapy to postpone death;⁴⁵ and adverse mortality trends in mental health patients with prevalent ischaemic heart disease.²²

Data linkage was essential to all of these applications and many would have been impossible without pre-existing, systematic data linkage infrastructure due to the prohibitive cost of creating once-off links between large numbers of records (see Box 3 as an illustration).

Methodologic research

Each new stage in the development of the WADLS has brought new methodological challenges that have exceeded textbook-based knowledge. Specific developments arising from WADLS projects have included a backcasting model to correct for the prevalent pool artifact in first-time incidence rates based on linked multi-event data;⁴⁶ the use of cohort-crossover designs to estimate effects of index procedures and health events on short-term readmission risks;⁴⁷ the

couplet methodology and other new methods for the study of use of private health insurance;⁴¹ and an improved method of risk adjustment, the Multipurpose Australian Comorbidity Scoring System (MACSS).^{48,49}

The developments of cross-jurisdictional and genealogical linkages have brought further methodological challenges. New methods for the analysis of linked MBS and PBS data documented in successful grant applications have included linkage-based exposure indices for the intensity, periodicity, continuity and disease specialisation of primary medical care, as well as new approaches to comorbidity risk adjustment and case-time-cohort studies of medication safety using linked pharmaceutical data.

Benefits

The total marginal cost of implementing a data linkage system in WA, estimated to have been A\$3.7 million over 10 years, has been inexpensive in comparison with biomedical research infrastructure. The advantages and social benefits arising from this investment are now apparent. Some benefits were anticipated from the outset. Others, such as the conservation of patient privacy in research, have only come to light with experience. We have identified the following dis-

tinct areas, in approximate ascending order of importance, where the benefits of data linkage have accrued:

Adding value to existing information assets

Data linkage has generated a greater return on the substantial existing investment in routine administrative data sets, which have often been perceived in the past to be unproductive record-keeping systems. It has added value to existing data sets through the improvement of data quality. Tens of thousands of duplication errors in records have been corrected and other technical glitches with data have come to light, leading to greater accuracy of recording at an administrative level.

Commercial and competitive benefits

The WADLS has endowed its research community with a competitive advantage in attracting funds. The resultant flows of research income from outside WA, including several NIH grants from the United States and many large Australian NHMRC grants, has summed to A\$58 million, and when discounted at 4% per annum back to the years of infrastructure expenditure represented greater than a tenfold return on investment. This revenue has provided employment and stimulated the economy. We believe that WA has also been relatively more successful in attracting and retaining individual research leaders of high quality.

Cost-efficiency of research

Linking existing data has proven to be a cheap and effective alternative to performing *de novo* longitudinal studies and other more traditional approaches to epidemiologic and health services research, especially in mobile populations.^{50,51} This has conserved the limited resources available to support health and medical research, thus enabling more research to be performed with a given research budget.

Conservation of patient privacy

The privacy of individual patients has been conserved by reducing the need for release of names and other personal identifiers to researchers,

because a major reason for the release of identifiers in the past (to find and clerically collate information on patients) has been removed.⁵² Contrasted with consent-based responses to privacy concerns, data linkage systems are superior because they conserve the privacy of all patients regardless of whether or not they would have given consent to the use of information. A consent-based approach conserves the privacy only of those who decline to participate, and this may be at the cost of irreconcilable impracticability of the research. Epidemiological and health service research projects using named data provided from major WA health collections fell from >90% before the WADLS to 36% in 2003.⁵²

Community development

Although difficult to quantify, the WADLS has enhanced the level and quality of interactions between researchers, clinicians, administrators, consumer groups and the mass media. The Quality of Surgical Care Project and the *Health+Medicine* newspaper supplement are examples.^{15,18-20} It has provided a focal point for envisioning future possibilities, for improved cooperation and for rigorous debate about the uses of data and the results of subsequent research.³⁸ This has strengthened the community by exercising the machinery necessary for teamwork and other organised responses to health and social problems.

Contributions to medical and scientific knowledge

A good volume of published, peer-reviewed research results has been forthcoming, facilitated by our policy of open access to *bona fide* researchers. The contributions to knowledge have taken the form of scientific publications and conference presentations, as well as theses and dissertations prepared by research trainees.³³ The scientific outputs have been free from the biases introduced when samples achieve far from complete recruitment or arise from a limited number of institutions that may not be representative of the performance of an entire health system on a complete, geographically-defined population.

Improvements in population health

These have occurred at least at two levels. Contributions to knowledge published in peer-reviewed journals have contributed to worldwide medical and scientific knowledge, which will eventually translate into better programs of preventive services, treatment and care for human populations. Specific local improvements in population health include Australia's first 30-day surgical mortality audit program.^{15,53} Reforms in mental health legislation and service delivery can also be traced directly to WA data linkage-based research.^{21,22,33}

Conclusion

The creation of a data linkage system is a challenge in social organisation. It demands leadership, inter-agency and inter-sectoral cooperation, a dedicated group of users who drive reforms, and perseverance. It is the relative scarcity of these community resources that explains why comprehensive, population-based data linkage systems are so few in the world today. The requirements for technological solutions and resources are overshadowed by the importance of high-quality relationships between leading individuals and between institutional players. These relationships have assisted in development of best practice protocols and procedures to address privacy concerns and in disseminating research results more broadly in support of the public good.

Having surmounted these organisational and community challenges, at least for a brief period in history, few would doubt that the social, scientific, economic and privacy protection benefits in Western Australia have been worth the effort. We anticipate such systems will become the norm in countries like Australia within the next decade.

Competing interests

The authors declare that they have no competing interests.

References

1 Australian Government Department of Education, Science and Training. National Collaborative

Research Infrastructure Strategy. Strategic road-map. 2006. Available at: http://www.ncris.dest.gov.au/key_documents.htm (accessed Oct 2007).

- 2 Dunn HL. Record linkage. *Am J Public Health* 1946; 36: 1412-6.
- 3 Roos LL, Menec V, Currie RJ. Policy analysis in an information-rich environment. *Soc Sci Med* 2004; 58: 2231-41.
- 4 Roos NP, Black CD, Frohlich N, et al. A population-based health information system. *Med Care* 1995; 33: DS13-20.
- 5 Baldwin JA. The Oxford Record Linkage Study as a medical information system. *Epidemiol Prev Med* 1972; 65: 237-9.
- 6 Kendrick S, Clarke J. The Scottish Record Linkage System. *Health Bull* 1993; 51: 72-9.
- 7 Kurland LT, Molgaard CA. The patient record in epidemiology. *Scientific American* 1981; 245: 54-63.
- 8 British Columbia Centre for Health Services and Policy Research. BC Linked Health Database. Available at: <http://www.chspr.ubc.ca/data> (accessed Oct 2007).
- 9 Holman CD, Bass AJ, Rouse IL, Hobbs MS. Population-based linkage of health records in Western Australia: development of a health services research linked database. *Aust N Z J Public Health* 1999; 23: 453-9.
- 10 Hobbs MST, McCall MG. Health statistics and record linkage in Australia. *J Chron Dis* 1970; 23: 375-81.
- 11 Stanley FJ, Croft ML, Gibbons J, Read AW. A population database for maternal and child health research in Western Australia using record linkage. *Paediatr Perinatal Epidemiol* 1994; 8: 433-47.
- 12 Stanley FJ, Read AW, Kurinczuk JJ, et al. A population maternal and child health research database for research and policy evaluation in Western Australia. *Semin Neonatol* 1997; 2: 195-201.
- 13 Rosman DL, Knuiman MW. A comparison of hospital and police road injury data. *Accident Anal Prev* 1994; 26: 215-22.
- 14 Brameld KJ, Holman CDJ, Bass AJ, et al. Hospitalisation of the elderly during the last year of life: an application of record linkage in Western Australia 1985-1994. *J Epidemiol Community Health* 1998; 52: 740-4.
- 15 Semmens JB, Lawrence-Brown MM, Fletcher DR, et al. The Quality of Surgical Care Project: a model to evaluate surgical outcomes in Western Australia using population-based record linkage. *Aust N Z J Surg* 1998; 68: 397-403.
- 16 Patterson KM, Holman CDJ, English DR, et al. First-time hospital admissions with illicit drug problems in indigenous and non-indigenous Western Australians: an application of record linkage to public

- health surveillance. *Aust N Z J Public Health* 1999; 23: 460-3.
- 17 Brameld K, Holman D, Thomas M, Bass J. Use of a state data bank to measure incidence and prevalence of a chronic disease: end-stage renal failure. *Am J Kidney Dis* 1999; 34: 1033-9.
 - 18 Holman D. Health system needs check-up too. *The West Australian* 1999 Aug 3: 16.
 - 19 WA tops in data links. *The West Australian* [H+M Supplement] 2003 Nov 19: 6.
 - 20 McKimmie M. Data link system tipped to improve patient safety. *The West Australian* 2005 Nov 23: 7.
 - 21 Lawrence D, Holman CD, Jablensky AV, et al. Excess cancer mortality in Western Australian psychiatric patients due to higher case fatality rates. *Acta Psychiatr Scand* 2000; 101: 382-8.
 - 22 Lawrence DM, Holman CD, Jablensky AV, Hobbs MS. Death rate from ischaemic heart disease in Western Australian psychiatric patients 1980-1998. *Br J Psychiatry* 2003; 182: 31-6.
 - 23 Kelman CW, Bass AJ, Holman CD. Research use of linked health data — a best practice protocol. *Aust N Z J Public Health* 2002; 26: 251-5.
 - 24 Glasson EJ, de Klerk NH, Bass AJ, et al. Cohort profile: The Western Australian genealogical project. *Int J Epidemiol* 2008; 37: 30-35.
 - 25 Western Australian Institute for Medical Research. The proposed Joondalup Family Health Study. <http://www.jfhs.org.au/> (accessed Oct 2007).
 - 26 Moorin RE, Holman CDJ, Garfield C, Brameld KJ. Health related migration: evidence of reduced "urban-drift". *Health Place* 2006; 12: 131-40.
 - 27 Glover J, Rosman DL, Tennant S. Unpacking analyses relying on area-based data: are the assumptions supportable? *Int J Health Geogr* 2004; 3: 30.
 - 28 Hall SE, Holman CD. Inequalities in breast cancer reconstructive surgery according to social and locational status in Western Australia. *Eur J Surg Oncol* 2003; 29: 519-25.
 - 29 Hall SE, Holman CD, Wisniewski ZS, Semmens J. Prostate cancer: socio-economic, geographical and private-health insurance effects on care and survival. *BJU Int* 2005; 95: 51-8.
 - 30 Brameld KJ, Holman CD. The use of end-quintile comparisons to identify under-servicing of the poor and over-servicing of the rich: a longitudinal study describing the effect of socioeconomic status on healthcare. *BMC Hlth Serv Res* 2005; 5: 61.
 - 31 Brameld KJ, Holman CD. The effect of locational disadvantage on hospital utilisation and outcomes in Western Australia. *Health Place* 2006; 12: 490-502.
 - 32 Centre for Health Record Linkage. Linking health data for NSW and the ACT. Available at: http://www.cherel.org.au/CHeReL_flyer.pdf (accessed Oct 2007).
 - 33 Brook EL, Rosman DL, Holman CD. Public good through data linkage: measuring research outputs from the Western Australian Data Linkage System. *Aust N Z J Public Health* 2008; 32: 19-23.
 - 34 Fletcher DR, Hobbs MS, Tan P, et al. Complications of cholecystectomy: risks of laparoscopic approach and protective effects of operative cholangiography: a population-based study. *Ann Surg* 1999; 229: 449-57.
 - 35 Hansen M, Kurinczuk JJ, Bower C, Webb S. The risk of major birth defects after intracytoplasmic sperm injection and in vitro fertilization. *N Engl J Med* 2002; 346: 725-30.
 - 36 Stanley F, Blair E, Alberman E. Cerebral palsies: epidemiology and causal pathways. Cambridge: Mac Keith Press, 2000.
 - 37 Laurvick CL, Semmens JB, Leung YC, Holman CD. Ovarian cancer in Western Australia (1982-98): trends in surgical intervention and relative survival. *Gynecol Oncol* 2003; 88: 141-8.
 - 38 Hall SE, Holman CD, Finn J, Semmens JB. Improving the evidence base for promoting quality and equity of surgical care using population-based linkage of administrative health records. *Int J Qual Health Care* 2005; 17: 415-20.
 - 39 O'Leary CM, de Klerk N, Keogh J, et al. Trends in mode of delivery during 1984-2003: can they be explained by pregnancy and delivery complications? *BJOG* 2007; 114: 855-64.
 - 40 Holman CDJ, Wisniewski ZS, Semmens JB, Bass AJ. Changing treatments for primary urolithiasis: impact on services and renal preservation in 16,679 patients in Western Australia. *BJU Int* 2002; 90: 7-15.
 - 41 Moorin RE, Holman CDJ. Patient-initiated switching between private and public inpatient hospitalisation in Western Australia 1980-2001: an analysis using linked data. *Aust New Zealand Health Policy* 2005; 2: 12.
 - 42 Moorin RE, Holman CDJ. The influence of federal health care policy reforms on the use of private health insurance in disadvantaged groups. *Aust Health Rev* 2006; 30: 241-51. Available at: http://www.aushealthreview.com.au/publications/articles/issues/ahr_30_2_0506/ahr_30_2_241.asp (accessed Aug 2008).
 - 43 Kelman CW, Kortt MA, Becker NG, et al. Deep vein thrombosis and air travel: record linkage study. *BMJ* 2003; 327: 1072.
 - 44 De Klerk NH, Ambrosini GL, Musk AW. Crystalline silica exposure and major health effects in Western Australian gold miners. *Ann Occ Hyg* 2002; 46 (S1): 1-3.

- 45 Brameld KJ, Holman CD, Threllfall TJ, et al. Increasing 'active prevalence' of cancer in Western Australia and its implications for health services. *Aust N Z J Public Health* 2002; 26: 164-9.
- 46 Brameld KJ, Holman CD, Lawrence DM, Hobbs MS. Improved methods for estimating incidence from linked hospital morbidity data. *Int J Epidemiol* 2003; 32: 617-24.
- 47 Brameld KJ, Holman CD. Estimation of excess risk of readmission to hospital after an index inpatient separation. *Med Care* 2003; 41: 693-7.
- 48 Holman CD, Preen DB, Baynham NJ, et al. A multipurpose comorbidity scoring system performed better than the Charlson index. *J Clin Epidemiol* 2005; 58: 1006-14.
- 49 Preen DB, Holman CD, Spilsbury K, et al. Length of comorbidity lookback period affected regression model performance of administrative health data. *J Clin Epidemiol* 2006; 59: 940-6.
- 50 Bartu A, Freeman NC, Gawthorne GS, et al. Mortality in a cohort of opiate and amphetamine users in Perth, Western Australia. *Addiction* 2004; 99: 53-60.
- 51 Calver J, Wiltshire A, Holman CD, et al. Does health assessment improve health outcomes in indigenous people? An RCT with 13 years of follow up. *Aust N Z J Public Health* 2005; 29: 107-11.
- 52 Trutwein B, Holman CDJ, Rosman DL. Health data linkage conserves privacy in a research-rich environment. *Ann Epidemiol* 2006; 16: 279-80.
- 53 Semmens JB, Aitken RJ, Sanfilippo FM, et al. The Western Australian Audit of Surgical Mortality advancing surgical accountability. *Med J Aust* 2005; 183: 504-8.

(Received 5/10/07, revised 14/03/08, accepted 26/06/08) □