Access to routinely collected data for population health research: experiences in Canada and Australia

David Henry,1,2,3 Paulina Stehlik,1 Ximena Camacho,2 Sallie-Anne Pearson4

1. Centre for Research in Evidence-Based Practice, Bond University, Queensland
2. Melbourne School of Population and Global Health, University of Melbourne, Victoria
3. Institute for Clinical Evaluative Sciences, Canada
4. Medicines Policy Research Unit, Centre for Big Data Research in Health, UNSW, New South Wales

In 2008, researchers at St Michael’s Hospital in Toronto, Canada, demonstrated an association between neighbourhood walkability and the prevalence of obesity and diabetes.1 They did so by linking the physician billing claims, hospital discharge data and residential postcodes for all residents of Toronto. They applied a validated algorithm to the physician claims and hospital data to identify people with diabetes and assigned a walkability index to each postcode. This revealed striking gradients, with relatively low rates of diabetes in the highly walkable central zones of the city and higher rates in the less walkable suburbs. This did not prove that poorly walkable neighbourhoods caused diabetes. A randomised trial to establish causality was not feasible. So, the researchers conducted a controlled cohort study where they measured diabetes incidence in more than 200,000 recent Canadian immigrants who were free of diabetes when they moved to different areas of Toronto.2 After adjusting for age and socioeconomic status, they found the incidence of diabetes was 58% higher in immigrants who moved into the least walkable neighbourhoods, relative to those who moved into the most walkable neighbourhoods. A weaker but still significant association between neighbourhood walkability and incidence of diabetes was found among long-term Toronto residents. These findings have influenced the design of new housing developments in Ontario.

Use of routinely collected health data in Canada

This diabetes mapping study arises from extensive experience of using routinely collected health data in population health research in Canada. Researchers in Ontario and the neighbouring province of Manitoba have excellent access to population-level data, including primary care and specialist billing data, prescription dispensing claims, hospital discharge diagnoses, social service usage, vital statistics and census information, which can be enriched through linkage to electronic medical records, registries, surveys, laboratory data, Indigenous and immigrant status. All are linkable at the level of the individual.3,4

While many of the collections, particularly administrative (payment) data, lack key information such as body weight, clinical measurements and lifestyle factors these gaps can be filled to varying degrees by the information from electronic medical records, clinical registries and national surveys. In Canada, studies like the diabetes mapping example are possible through national and provincial support for key developments including: i) robust and transparent governance of linked comprehensive population-level data, enabling timely approval of large programs of research generating hundreds of individual projects;5,6 ii) funding of data repository infrastructure that generates affordable marginal project costs;7 iii) validation of linked administrative data, and creation of algorithms and code sets that are shared across researcher groups;8 iv) the establishment of distributed networks of independent data centres conducting collaborative research of national importance, for instance the investigation of safety of prescription medicines;9 v) involvement of policy makers as knowledge users;10 and vi) broad public support for the use of personal health information for research by data institutes and university researchers.11

A key success factor in Manitoba and Ontario was the establishment of dedicated independent centres where linked health data from the whole population are used to study a wide range of health problems. The Manitoba Centre for Health Policy (MCHP) and the Institute for Clinical Evaluative Sciences (ICES) act as secure stewards, making routinely collected data available in research-ready form to large numbers of researchers working within and outside the centres.12,13 Data are not released, and all analyses are done in a secure environment with external access through a virtual private network. The work has been enabled by legislation that allows the centres to work with linked unidentifiable data without the consent of the individuals involved. Crucially, the centres take on the responsibilities of the original data custodians, who are not involved in the approval of studies involving secondary uses of the data. Both centres have strong data-sharing and governance agreements with many data custodians. Two national bodies, Statistics Canada and the Canadian Institute for Health Information (CIHI), share data to support external research. The former provides key census-derived data down to postcode level and conducts and shares data from regular comprehensive national surveys.14 CIHI develops and maintains data standards, and acts as a national steward of multiple provincial data-sets (and some registries), making them available in clean linkable research-ready form.15 Staff at the MCHP and ICES carry out validation work, including medical chart reviews at institutions across the provinces. The validation data, algorithms and code-sets are part of the intellectual property that is made available to everyone working with the data. Training is provided to analysts and epidemiologists, and at ICES there is a scholars’ program for scientists.16 Many postgraduate research students complete the practical phases of their work in these centres. Both institutes conduct and facilitate a wide range of studies, including health and social policy analyses, health services research, public health research, pharmaco-epidemiology, health economics and clinical studies in areas ranging from mental health to chronic diseases and cancer. Examples of published studies are provided in the Supplementary File.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
while the states are responsible for hospital and clinic services, and there is a significant private sector. The central role of the Federal Government means that linkage of key data-sets requires cross-jurisdictional data-sharing, with the associated complexity and delays in project approvals. This is usually not necessary in Canada, with some exceptions, such as data that identify First Nations individuals and immigrants.21,22

The publics in both countries appear to support broadly the use of routinely collected health data in health and medical research.11,27 In Australia, King and colleagues (2012) reported that 73% of survey respondents would agree to their medical records being used for medical research.27 In the presence of extra security measures a minority of respondents (25%) would feel some concern about threats to their privacy. In Canada, Willison and colleagues found that over 80% of survey respondents supported the use of their medical records for health research.11 They expressed much greater trust in disease foundations, hospitals and universities than in pharmaceutical manufacturers and insurance companies.

Access to routinely collected health data for comprehensive health research in Australia

In Australia, the states govern the use of their own data and support a wide range of research studies. However, these often lack critical information that is only available from Federal Government data. Compared with the largely distributed processes that govern the use of linked whole-of-system health data in Canada, Australia has a more centralised data access process. The National Statistical Service (NSS) lays out the principles governing the linkage and use of Federal Government data.20 Researchers must go through a demanding project-by-project process that includes the approval of data custodians.29 By contrast, in Ontario and Manitoba, the MCHP and ICES subsume the roles of the data custodians of the linked collections; original data custodians are not required to approve individual projects arising from these linkages, making for a faster and more efficient approval process. Historically, there have been only three accredited data integrating authorities in Australia; all are government agencies – the Australian Bureau of Statistics, the Australian Institute for Health and Welfare (AIHW) and the Australian Institute of Family Studies.20

These organisations have important statutory and non-statutory roles and functions, which must limit their capacity to evaluate and respond to large numbers of requests for linked data for the purposes of external research.

Since 2009, the Federal Government has committed more than $46 million to establish and support the Population Health Research Network (PHRN).21 PHRN has distributed funds to six state/territory data linkage units, one national accredited integration authority (AIHW) and a secure national remote-access data research laboratory, the SURE facility hosted by the Sax Institute in NSW. These centres are bound by the data access rules established by the Federal Government for use of its data, leading to slow project approvals.

Despite these challenges there have been excellent examples of innovative use of cross-jurisdictional data linkage, which are testimony to the tenacity of Australian researchers. Large cohort studies have used participant consent to access Federal Government health data. Examples are the Australian Longitudinal Study on Women’s Health and the 45 And Up study.22,23 The Department of Veterans Affairs (DVA) acts as a single payer for the health care of its clients and, like the provinces of Canada, provides comprehensive data without the need for cross-jurisdictional linkage. The DVA has been pro-active in supporting research to improve the health and wellbeing of the veteran community and has helped establish pharmacoepidemiology as a research discipline in Australia.24,25

At state level, the Western Australia Data Linkage System has been a pioneer in the use of linked data,26 and this has sometimes included Federal Government data.27 Data linkage units in other provinces have followed the WA lead, including the Centre for Health Record Linkage (CheReL) in NSW28 and the SA-NT DataLink facility for South Australia and Northern Territory.29,30 The SURE facility has established best practices in the independent stewardship of linked state and federal data.31

In the face of these large investments the number of projects conducted in Australia using linked federal data remains very low compared to what has been achieved in other countries. A recent systematic review, led by one of us (SAP), cataloguing published studies using Pharmaceutical Benefits Scheme data from 1987 to 2013, found that

Comparison of routinely collected health data in Australia and Canada

There are similarities between the two countries. Both have comprehensive publicly funded healthcare systems that generate routine health data-sets documenting use of primary and specialist care, procedures, prescription medicines and hospital visits, and include a variety of registries.

But there are some key differences in the organisation of healthcare and hence data capture. In Canada, each province acts as a single payer for all health services and there is no private sector providing medically necessary care. The Canadian provinces have the lead role in providing both community and hospital-based services, including prescription medicines, although the extent of coverage of the latter varies by province. In Canada, health insurance claims often include a diagnostic code.21 In comparison, Medical Benefits Schedule (MBS) claims data in Australia do not routinely include information on diagnosis.

In Australia, the Federal Government is directly responsible for the funding of primary care and specialist consultations and prescription medicines in the community.
The Productivity Commission recommend legal and structural reforms are aimed at moving from a system based on risk aversion and avoidance, to one based on transparency and confidence in data processes, treating data as an asset and not a threat.45

The Report recommends the Government plans to appoint a National Data Commissioner who will work together with the Australian Privacy Commissioner to oversee the new data access framework. The Data Commissioner will be assisted by a National Data Advisory Council, who will consult widely with community groups and provide advice on a range of issues including ethics and best practices in data privacy and security. Accredited Data Authorities will be established to determine which data-sets are made public, as well as who can access them.

Federal Government response to the Productivity Commission recommendations

The response was released on 1 May 2018.46 The Government has committed $65 million over the next four years to support Australia’s data infrastructure and plans to implement several of the main recommendations of the Productivity Commission. The Government will introduce a Consumer Data Right to provide consumers with greater control over the data that businesses hold about their activities. Government of legitimate use of data (including de-identified health data) will be simplified and streamlined through new legislation that protects individual privacy. The Government plans to appoint a National Data Commissioner who will work together with the Australian Privacy Commissioner to oversee the new data access framework. The Data Commissioner will be assisted by a National Data Advisory Council, who will consult widely with community groups and provide advice on a range of issues including ethics and best practices in data privacy and security. Accredited Data Authorities will be established to determine which data-sets are made public, as well as who can access them.

Conclusions

Without question this is a positive response by Government. But we have one concern and one recommendation. Our concern relates to the governance of the Accredited Data Authorities, where the Government states that “accountability for the risks of sharing and releasing data will remain with data custodians.”47 It is not clear exactly what this means. However, based on our extensive experience of working with routinely collected health data in Canada and Australia, we believe that, if appropriate protections are in place, it is not necessary to involve the original data custodians in evaluating and approving proposed secondary uses of data. Once data have been shared through a secure mechanism and have been linked and de-identified by the Accredited Data Authority, decisions about downstream secondary uses should devolve to the appropriate oversight bodies concerned with privacy, science and ethics.

Finally, we recommend that the proposed Accredited Data Authorities should be independent of government. Government agencies inevitably have conflicting priorities and, in our experience, have difficulty attracting and retaining the necessary complement of highly trained data scientists and methodologists. Independent data research institutes like those in Manitoba and Ontario have the capacity to perform extensive analyses on behalf of government agencies, as well as to facilitate investigator-initiated research, program...