Using linked data to more comprehensively measure the quality of care for stroke – understanding the issues

Dominique A Cadilhac,[1,2] Vijaya Sundararajan,[1] Nadine Andrew,[1] Monique F Kilkenny,[1,2] Felicity Flack,[3] Phil Anderson,[4] James Boyd,[5] Judith Katzenellenbogen,[6] Amanda G Thrift,[1,2]

1 Department of Medicine, Southern Clinical School, Monash University, Australia

2 Stroke Division, Florey Institute of Neuroscience and Mental Health, Melbourne Brain Centre, Australia

3 Population Health Research Network, Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, Australia

4 Data Linkage Unit, Australian Institute of Health and Welfare, Australia

5 Population Health Research Network Centre for Data Linkage, Centre for Population Health Research, Curtin University, Australia

6 Combined Universities Centre for Rural Health, Australia

E-mail: dominique.cadilhac@monash.edu

Videos of the workshop panels described in this contribution can be found online at: **goo.gl/V07vm**

National workshop on performance, care quality and telemedicine is the common heading for all videos readers should look for.

Abstract

This article brings together a summary of the current issues in the utilisation of available Australian data to better understand the quality of health care. Data linkage is an important method that can help to reduce the burden of collecting data. However, issues related to governance and ethics, and technical processes such as data quality and analysis, can be significant impediments to undertaking data linkage projects. These issues may be exacerbated when attempting to link non-government clinical registry data with routinely collected government data. We outline these issues and some possible solutions using the condition of stroke as a case study. This article represents a summary of the proceedings of national workshop on the topic of data linkage and stroke.

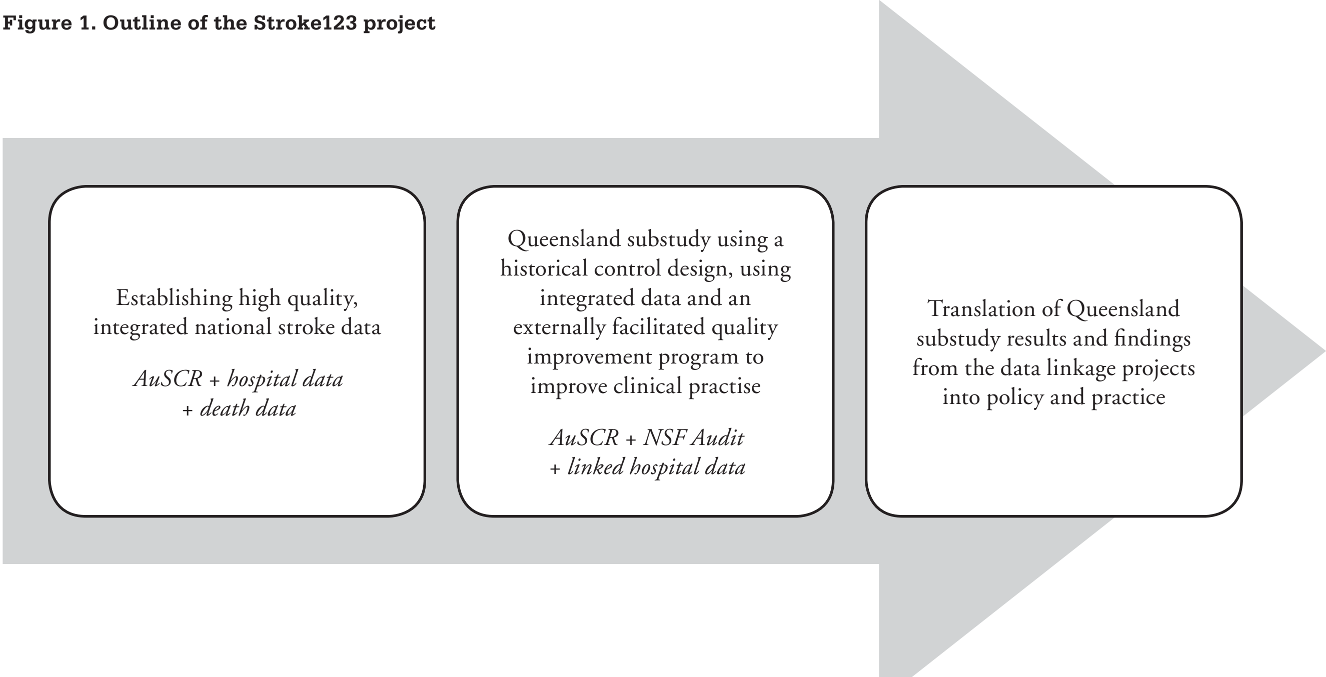
Introduction

Monitoring the quality of health care is a major focus for funders, providers and patients. In Australia, as in other countries, there are various systems in place to collect data, but there is inconsistency in definitions and data collection, and consequent duplication and potential underutilisation of data. Data linkage can be used to reduce the burden of data collection while also allowing a better understanding of the quality of health care and identifying where improvements might be focused. However, issues related to governance, data ownership, ethical considerations and data quality and analysis can create impediments to undertaking data linkage projects. In this article, we summarise these issues and the possible solutions using the condition of stroke as a case study.

In Australia, stroke is the leading cause of adult disability and the second greatest cause of death.[1] Variation in clinical care for stroke contributes to the societal burden of this disease,[2-4] but routinely collected data describing the clinical variation for stroke and its relationship to health outcomes are limited. Various methods for collecting stroke data for monitoring the quality of hospital care have evolved, mainly outside the government sector. Nationally, there are two main programs. The first is the National Stroke Audit Program established in 2007 by the National Stroke Foundation (NSF), designed to measure adherence to the Clinical Guidelines for Stroke Management. This comprises a 40 – 80 case retrospective review of clinical care and a survey of hospital resources at each participating hospital.[5, 6] The second is the Australian Stroke Clinical Registry (AuSCR), established in 2009 in order to collect data on all patients hospitalised with acute stroke and transient ischaemic attack. Patients are given the choice to opt-out, and follow-up data are collected between 90-180 days after the index stroke onset date.[7] Participation by public and private hospitals in both these programs is voluntary. The NSF audit program and AuSCR are complementary and use the same definitions for variables that overlap. Currently, 42 hospitals have ethical approval for AuSCR and there are over 11,000 registered episodes of care; the opt-out rate is <3%, and 88% of eligible registrants have completed follow-up (www.auscr.com.au).

Although each of these national programs has their own purpose and merits, multiple collections create duplication of effort. The ability to reliably link data from other databases can reduce the data collection burden and maximise the use of existing data.[8] This can be facilitated with AuSCR data since personal identifying information is available. However, the governance and technical issues for linking government and non-government datasets are complex.

Figure 1. Outline of the Stroke123 project



AUSCR: Australian Stroke Clinical Registry; NSF: National Stroke Foundation

In 2011, a four-year National Health and Medical Research Council (NHMRC) Partnerships for Better Health grant (1034415), Stroke123, was awarded to a consortium of researchers, government and non-government organisations. The main objective of this project is to build on existing Australian programs to improve acute stroke care (Figure 1). A major aim is to establish high-quality, linked data for stroke using data from AuSCR, the NSF audit and routinely collected hospital and death data. A dedicated National Stroke Data Linkage sub-committee was established to progress this work and comprises several Stroke123 investigators and experts in this area. Our focus here is to outline the issues regarding linking non-government clinical registry data with routinely collected government data. The linkage of AuSCR and NSF audit data is not considered here. We present the issues from the perspective of stroke, but these would be relevant to a range of other diseases.

Methods

On 10 April, 2013 the National Stroke Data Linkage subcommittee held a 3½ hour workshop. The two interactive sessions commenced with short presentations to outline background information and current issues on governance and technical aspects of data linkage. The presentations were followed by small group and/or panel discussions moderated by two of the authors (DAC and AGT).

Results

Thirty-two invited attendees from across Australia participated. The main issues discussed at the workshop are presented below and outlined in the Table.

Privacy, governing legislation and the public good

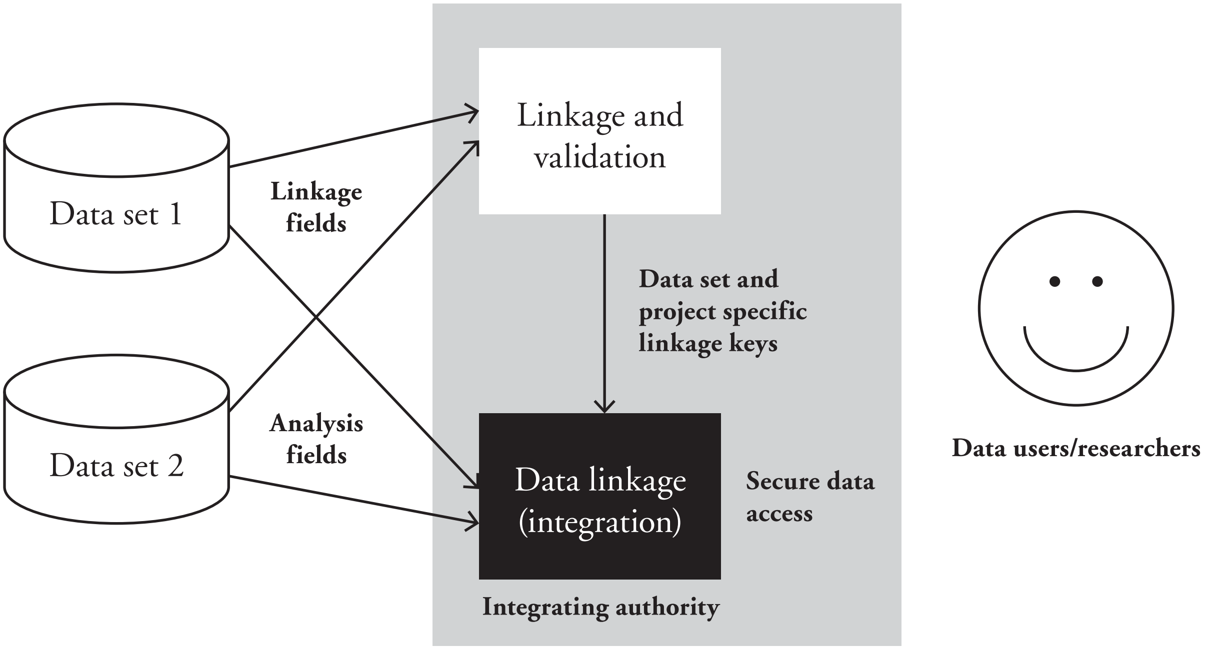
Information privacy concerns the right of individuals to control their personal information. Usually we respect a person's right to information privacy by asking for their consent to use their information for research. However, this is not an absolute right and must be balanced with competing public interests. A range of empowering and privacy legislations provide exemptions for the use of personal information in research without consent. This requires approval by a human research ethics committee and appropriate strategies to minimise the risks to privacy.

Minimising risks to privacy: separating roles

Privacy risks may be minimised by using the two-stage separation model of data linkage,[9] which involves the separation of identifying data (name, address, date of birth, gender) from content data.[10] The identifying data are used for linkage and the content data without personal identifiers are provided to researchers for analysis. This model also requires the separation of roles and responsibilities; i.e. the people conducting the linkage should not also be the researchers conducting the analysis.

A two-stage model, as practised in Australia, consequently preserves privacy (Figure 2). However, there are complexities when data custodians have dual roles. Many clinical registries, such as AuSCR, have emerged from research institutes where their data custodians are often also leading researchers in their field. At this point, although the data custodian of the clinical registry may be a principal investigator for a study, they should not directly be involved in linked data analysis. Therefore, separation of roles in the analytic process is important to establish.

Figure 2. Data linkage under the Integrating Authority framework



Governance and ethics
Addressing the governance and ethical issues can be complex and time consuming. Research projects using linked data require approvals from the data custodian of each data collection included in the project, the relevant human research ethics committee(s) and the research institution(s) conducting the research. Obtaining these approvals can take many months and may result in additional expenses, so timelines and researcher expectations must take into account these complexities.

Barriers to using routinely collected hospital data
In Australia, there is agreement between the State and Territory Health Authorities and the Commonwealth Government to collect the National Minimum Datasets which include a minimum set of variables for admitted patient (http://meteor.aihw.gov.au/content/index.phtml/itemId/426861) and non-admitted patient emergency department care (http://meteor.aihw.gov.au/content/index.phtml/itemId/426881). Some States collect additional variables for which differences may exist in variable definitions, as well as how they may be used for data linkage. For example, full names and addresses are not currently available for data linkage in Victoria's hospital data collections. Similarly, Queensland collects Medicare numbers, but they are not available for data linkage purposes. Therefore, methods to link data may require a different mix of variables or derivations for each State. This means that establishment of a national data linkage algorithm may be very complex and may require a different process for each State.

Groups that can provide assistance
There are two Commonwealth-approved national integrating authorities involved with data linkage: the Australian Bureau of Statistics and the Australian Institute of Health and Welfare (AIHW). In addition, the Population Health Research Network (PHRN) is a collaborative network developing data linkage capability within and between Australian jurisdictions (see http://www.phrn.org.au/). The Centre for Data Linkage at Curtin University and the AIHW are members of the PHRN and have established secure data linkage infrastructure to enable national and cross-jurisdictional health-related research in Australia. The Centre for Data Linkage has a focus on linking health information, e.g. admission or emergency presentation data from different States. The AIHW have data linkage capability and resources for management and analysis of complex datasets, specialising in Commonwealth data. The PHRN is able to provide advice on the feasibility of linkage projects, with the AIHW providing specific advice about the linkage of Commonwealth data collections. This is important since health departments will not consider projects unless they are well-designed, of sufficient public health importance and feasible.

Linkage accuracy
The matching methodology chosen for a study can have an impact on the accuracy of the overall linkage or completeness of the linked data. For example, a linkage using 'exact matching' will be reliant on the quality of the data used; i.e. if any of the linkage variables are incorrectly recorded between sources, then that match will be lost. In contrast, probability matching can be used across several variables which include discrepancies and still make a match as it quantifies the levels of agreement and disagreement across the fields. However, this additional effort and complexity will not significantly improve the linkage if the data quality is good.

Analysis of linked data
Analysing linked data is complex because data are not always well-structured, accurate or complete. When the same variables have been collected in two linked datasets, these may provide conflicting information, e.g. indigenous status or International Classification of Diseases (ICD) 10 codes. Therefore, it is important to reconcile these differences, at times by using

one source of data as the ‘reference’ or developing an algorithm taking both sources into account. Where discrepancies are identified, these should be fed back to the data custodians as part of their data quality improvement process. Data validation and quality improvement, through identifying discrepancies in datasets, is a secondary but important role provided by linked data. Researchers should also have an ability to share data dictionaries and relevant programming used for undertaking analysis of linked data, e.g. programming for determining the Charlson index score[11] for predicting ten-year mortality using ICD-10 codes available in hospital data.[12]

Validation studies for clinical registries

In the context of registries, validation studies can allow one to test the representativeness of the clinical registry data and to assess whether linkage is complete. Validation studies can also inform whether there are any variables in the clinical registry or researcher collected data, such as AuSCR, that can be dropped, which can reduce the data collection burden. Validation studies may also provide some evidence that complete ascertainment of outcomes on all patients in the resource-intensive clinical registry may be unnecessary, since the registry data may be deemed representative of the disease population of interest. Therefore, morbidity outcome measurement could be limited to a sample of all cases.

Discussion

The opportunity to make better use of existing data through data linkage provides comprehensive and relevant data to answer numerous research questions, generate quality of care estimates, and reduce the need to collect primary data multiple times. The Stroke Data Linkage workshop was an opportunity to share ideas and advance this important work of relevance to the field of data linkage and stroke (in this instance). Importantly, it was recognised that: (i) many different researchers and organisations are grappling with the approval processes; (ii) it is important to separate roles and responsibilities; and (iii) the analytic methods required are complex but not insurmountable. With much still to be learnt in the field of data linkage, there was strong agreement that efforts and ‘know-how’ should be shared to avoid duplication. Researchers need to be aware of the benefits and pitfalls, and continued efforts must be made to improve Australian data linkage processes and governance requirements. Recently, reporting guidelines have been recommended for appraising the quality of reported data linkage studies.[13]

Although linked government data are useful, the data are limited if important clinical care indicators are not routinely collected, since these can provide guidance for where quality improvement should be focused. Within the domain of stroke (and many other conditions where clinical variation exists), clinical indicator reporting is not mandatory. This is the case, despite a stroke indicator set being recommended by the Australian Council on Healthcare Standards since 2008.[14] Because there is no requirement to collect these data, the establishment of stroke clinical care monitoring has been disparate and inefficient. This also creates data linkage inefficiencies because different data custodians exist for these data. A major outcome, and public good, of using linked data is that they directly contribute to improvements in the quality of care. There is international evidence that openly publishing quality of care results may provide an incentive to raise the quality of care at institutions with below-expected care practices or evidence of poor outcomes when case-mix has been taken into account.[15]

Embedding clinical quality monitoring for diseases of importance to Australia will improve the efficiency of health care and use of data, since many of the layers of bureaucracy will be eliminated. Ultimately comprehensive data are required to drive evidence-based improvements to clinical care, thus leading to better health outcomes.

Acknowledgments

The Stroke123 project is funded by a National Health and Medical Research Council (NHMRC) Partnership for Better Health grant (1034415). Partner organisations who contributed funds to this project are the Florey Institute of Neuroscience and Mental Health, the National Stroke Foundation, Queensland Health and Monash University. Significant in-kind support for this project is also provided from: Stroke Society of Australasia; The George Institute for Global Health; New South Wales Agency for Clinical Innovation; Victorian Stroke Clinical Network; Victorian Data Linkage Unit; South Australian Stroke Clinical Network; and Western Australian Stroke Clinical Leads. The Population Health Research Network is also acknowledged in contributing staff time related to data linkage projects. Dominique Cadilhac is support by a NHMRC/National Heart Foundation fellowship grant (610313). Educational grants for the national workshop on *Stroke performance data, Care quality and Telemedicine* hosted by the Florey Institute of Neuroscience and Mental Health & the National Stroke Foundation (from which a session on Data Linkage was undertaken as presented in this paper) was provided by Boehringer Ingelheim, Telstra, Polycom, National Stroke Foundation and Ipsen.

We wish to thank the participants of the workshop who are dedicated to addressing data linkage issues in stroke and the members of the Stroke123 Data Linkage Sub-committee who are progressing this work. We also thank Karen Moss (Florey), Brenda Grabsch (Florey) and Tharshanah Thayabaranathan (Monash University) for assistance with note taking for this meeting.

Table. Summary of issues and considerations

Issues	Considerations	Possible solutions
Governance		
Data custodians	Approvals needed from all custodians of data	Summary of custodian information, requirements and contacts made available at a central source; e.g. PHRN or AIHW or a dedicated website.
	Sustainability of national linked data system over the longer term	Benefits of national linked data systems need to be shown for continued investment by Government.
Ethical		
Consent	Personal information can be used for research without the person's consent in some circumstances	Data linkage projects seeking a waiver of consent must be approved by a Human Research Ethics Committee
Privacy	Contractual obligations relating to privacy and confidentiality	Data linkage performed by an expert (accredited) body (or integrating authority) such as the PHRN or AIHW Maintaining privacy by using data linkage keys which allow records to be linked from different data collections whilst protecting individuals identities
Security	Physical security means the data are locked and access is restricted to the building where the data are housed Plans for disaster recovery of data needed	Password protected, Encrypted One-way access computer, automatic locking Backups and disaster recovery plan Firewall protected, Virus and spyware protected
Storage	Retention should normally be for a specified period Encrypted	Purpose built remote storage environments with virtual access e.g. SURE
Minimising risk	Transport of data is the minimum necessary: password protected and encrypted.	Identifiers and encryption keys separated
Technical		
Matching	Variations exist in the types of data available for linkage between States and datasets e.g. rules for identifying hospitals Spelling and data entry errors can reduce completeness if using exact matching e.g. incorrectly spelt names	Development of a reliable national algorithm for data linkage Use of probabilistic or stepwise deterministic linkage techniques
Analysis	Completeness of data contained in datasets Accuracy of the data linkage procedures used	Data validation studies Use of a central 'reference' dataset Re-estimation of coding validity and capture-recapture rates
Variable definitions	Variable definitions can vary between states and datasets	Centralised and consistent definitions to facilitate consistent coding of variables

PHRN: Population Health Research Network; *AIHW*: Australian Institute of Health and Welfare; *SURE*: Secure Unified Research Environment

References

1. **Senes S.** How we manage stroke in Australia. AIHW cat no CVD 31. AIHW, editor. Canberra: *Australian Institute of Health and Welfare*; 2006 February. 60 p.
2. **Cadilhac DA, Carter RC, Thrift AG, Dewey HM.** Why invest in a national public health program for stroke? An example using Australian data to estimate the potential benefits and cost implications. *Health Policy.* 2007;83(2-3):287-94.
3. **Cadilhac DA, Kilkenny MF, Longworth M, Pollack MR, Levi CR.** Metropolitan-rural divide for stroke outcomes: do stroke units make a difference? *Intern Med J.* 2011;41(4):321-6.
4. National Stroke Foundation. National Stroke Audit – Acute Services Clinical Audit Report 2011. Melbourne: *National Stroke Foundation*, 2011 November. Page 60.
5. **Harris D, Cadilhac D, Hankey GJ, Hillier S, Kilkenny M, Lalor E.** National Stroke Audit: The Australian Experience. *Clinical Audit.* 2010;2:25-31.
6. **National Stroke Foundation.** Clinical guidelines for stroke management 2010. Melbourne: *National Stroke Foundation*; 2010 September. 168 p.
7. **Cadilhac DA, Lannin NA, Anderson CS, Levi CR, Faux S, Price C, et al.** Protocol and pilot data for establishing the Australian Stroke Clinical Registry. *Int J Stroke.* 2010;5(3):217-26. Epub 2010/06/12.
8. **Bohensky MA, Jolley D, Pilcher DV, Sundararajan V, Evans S, Brand CA.** Prognostic models based on administrative data alone inadequately predict the survival outcomes for critically ill patients at 180 days post-hospital discharge. *J Crit Care.* 2012;27(4):422 e11-21. Epub 2012/05/18.
9. **Kelman CW, Bass AJ, Holman CD.** Research use of linked health data--a best practice protocol. *Australian and New Zealand journal of public health.* 2002;26(3):251-5. Epub 2002/07/27.
10. **National Health and Medical Research Council.** National Statement on Ethical Conduct in Human Research. Canberra: Australian Government, NHMRC, *Australian Research Council*, Australian Vice-Chancellors' Committee, 2007 October. (Updated May 2013) page 113
11. **Charlson ME, Pompei P, Ales KL, MacKenzie CR.** A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of chronic diseases.* 1987;40(5):373-83. Epub 1987/01/01.
12. **Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al.** Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. *Med Care.* 2005;43(11):1130-9.
13. **Bohensky MA, Jolley D, Sundararajan V, Evans S, Ibrahim J, Brand C.** Development and validation of reporting guidelines for studies involving data linkage. *Australian and New Zealand journal of public health.* 2011;35(5):486-9. Epub 2011/10/07.
14. **Australian Council on Healthcare Standards.** *Australasian Clinical Indicator Report 2004–2011*: 13th edition. Sydney: ACHS, 2012 September. Page 112.
15. **Asplund K, Hulter Asberg K, Appelros P, Bjarne D, Eriksson M, Johansson A, et al.** The Riks-Stroke story: building a sustainable national register for quality assessment of stroke care. *Int J Stroke.* 2011;6(2):99-108.