

BIG DATA IN AN INDIGENOUS HEALTH CONTEXT: OPPORTUNITIES AND OBSTACLES

Gail Garvey,¹ Nikki Percival,¹ Liz Izquierdo,² David Moodie,³ Suzanne Moore¹

1. Wellbeing and Preventable Chronic Diseases Division, Menzies School of Health Research, Charles Darwin University, Northern Territory, Australia.

2. School of Nursing, Queensland University of Technology, Queensland, Australia.

3. Editorial Consultant, Seesaw Publishing, Northern Territory, Australia.

Email: Gail.Garvey@menzies.edu.au

Abstract

The ability of health researchers to unearth previously unsuspected health risks, trends and commonalities at a population level through matching information across different datasets is well attested. However, as more of this type of research is conducted, the spotlight is being shone on the barriers to accessing these data. Less well known are the complexities experienced by researchers working with datasets in an Aboriginal and Torres Strait Islander health context. We present the insights of a number of researchers, clinicians and public sector representatives who have extensive experience of data linkage in the Aboriginal and Torres Strait Islander health sector, on key issues and practical and ethical implications of utilising big datasets. Obstacles are further highlighted in the experiences of a national multicentre cancer cervical screening study. While researchers must at all times respect the individuals whose information is contained within these datasets, and abide by the legislative structures governing their use, measures to streamline data linkage processes are required. Realising the potential of existing health data that previously has not been available may underpin significant improvements in indigenous health and ultimately life expectancy.

Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying and information privacy. (wikipedia.org/wiki/Big_data: accessed 14th April 2016)

The growing ability of health researchers to unearth previously unsuspected health risks, trends and commonalities at a population level through matching information across different datasets is well attested.^{1,2} However, as more of this type of research is conducted, the spotlight is increasingly being shone on the barriers to accessing and using these data.³ Less well known are the complexities experienced by researchers working with data sets in an Aboriginal and Torres Strait Islander health context.

The complexity of conducting research across multiple centres in Australia is discussed widely by the research community. A number of publications have highlighted these difficulties, including the length and complexity of the ethics approval process, but the situation remains a time-consuming and challenging component of any project of this type.^{3,4} While it is imperative that a rigorous and thorough ethical review process is maintained, the

current system is exhaustive and costly in terms of the resources that are taken up to ensure compliance, and the delays in obtaining multiple approvals to proceed. As most research in Australia is publicly funded, all taxpayers should be comfortable that their tax dollars are being judiciously utilised.

The opportunities and obstacles that present when using large data sets and data linkage in indigenous health research, and how this approach is contributing to indigenous health research in Australia, has also been the topic of discussion, including at a roundtable conducted by Australia's National Institute for Aboriginal and Torres Strait Islander health research, the Lowitja Institute. This paper draws on the perspectives of the authors and from those gathered from semi-structured interviews and an online survey conducted with eight individuals - three researchers, three government health bureaucrats, one data clinician from a non-government organisation, and one chair of a research ethics committee who is also a researcher. All have extensive experience of data linkage in the Aboriginal and Torres Strait Islander health sector and a good understanding of the key issues and practical and ethical implications of utilising big datasets. These issues are further highlighted in the case study of a national multicentre cervical cancer project.

Key issues and practical implications

Indigenous identification

Not all datasets include a variable on indigenous status and in some historical datasets indigenous status was not gathered routinely or uniformly, making availability and reliability of data on indigenous identification particularly challenging. For example, indigenous status information may not be available for a baby and may be derived from the indigenous status of the mother. However, some datasets may not routinely include indigenous status information of both parents. This is changing now, but will take some time before the data become truly reliable.

Data linkage can improve Aboriginal identification as there is a greater likelihood that indigenous status will be recorded in one or more datasets. In the case of babies, if indigenous status of the mother was not captured at hospital admission, but she gave birth and received other services with Aboriginal status recorded, data linkage can increase confidence that the person/s involved are Aboriginal. The various Australian jurisdictions have developed different processes for data linkage that impact on the extent to which indigenous identification can be ascertained. Western Australia (WA) is held up as a national exemplar. For example, the WA Data Linkage System connects a wide range of datasets spanning up to 50 years. In collaboration with Telethon KIDS and Indigenous academics, a method to combine this information about indigenous identity has been developed so that a 'Getting Our Story Right' indigenous flag can be added to any approved data extract for analysis.

Federal fragmentation

A national approach to best practice in data linkage needs to be undertaken. It has been suggested that the varying jurisdictional approaches have contributed to the problem of under identification, leading to calls for more complementary and unity in the desire to use data that is collected from people for the benefit of the people. Under Australia's federal system of government, states/territories have control of health services, which has resulted in large amounts of data being collected and stored by them using divergent methods.

For national-scale research projects, the differing processes for accessing data between states/territories may also create additional issues. For example, there is fear that the fragmented approach will result in many silos making it difficult to streamline data access and linkage and thus impact on efforts to develop better access to data for all types of population based research.

There are also concerns about who owns and controls the data. Big datasets should be viewed in light of the potential benefits to Indigenous Australians and the current system, where government and/or state and territory departments hold and control these large datasets, can be a specific

barrier to sharing information and linking data. Challenges in accessing and sharing these datasets may also lead to mistrust among the community which is completely understandable given past injustices. Systematic and ethical processes for sharing information must occur, but systems must be established that enable the use of these data to assist in the development of better policies, planning, management and delivery of health services to Aboriginal and Torres Strait Islander people.

National collaboration

In recent years there has been a concerted effort to build better linkages between datasets in different jurisdictions, and to make data collection more uniform. These include the establishment of the Population Health Research Network in 2009,⁵ the publication in 2012 of national best practice guidelines for collecting indigenous status in health data sets,⁶ and a subsequent evaluation of these guidelines released in 2013.⁷ These initiatives are facilitating improvements in data linkage and strong support for their continuation within the existing system. For example, data linkage infrastructure is being developed across Australian states and territories through the Population Health Research Network. This includes technical development of data linkage systems modelled on those existing in WA and NSW. In addition, consistent access policies and research protocols have been developed and a secure data access environment is now operational through the Secure Unified Research Environment.

There are a number of other positive developments in linking and sharing data in the indigenous health context. For example, data custodians are increasingly aware of the importance of data linkage in enhancing indigenous identification across datasets with a view to generating reliable data for closing the gap in indigenous disadvantage.

National studies are likely to have more power for change in the long-run, but in the meantime there is a need to recognise the jurisdictional divide in order to work in the current climate.

Approvals processes

Ethical approval processes are time consuming and complex, adding additional challenges to linking datasets. All health research projects must go through ethical approval processes, and projects involving Indigenous Australians may also require additional approvals. Projects also need to be cleared by jurisdiction-based data custodians, all operating under different legislative regimes. The process may take several months to complete. For example, the NSW Ministry of Health has a partnership agreement with the Aboriginal Health and Medical Research Council. Under this agreement, projects that propose to use information on Aboriginal

ETHICS IN CANCER

people are referred to both the Aboriginal and Medical Research Council and also their Ethics Committee for approval prior to data release.

Although a centralised agreement for ethics applications and reporting of progress and outcomes would mitigate a lot of researcher fatigue and frustration, streamlining the approvals process must be balanced against the need to ensure cultural respect and that Indigenous people and communities are fully informed of research proposals using their health and health-related information. It is also important that Aboriginal people feel safe about providing their indigenous status with the knowledge that the data will be used for them with the aim of improving indigenous health and not against them (as in the past). There is a much greater possibility of data sharing and linkage if there is greater input and control of data by Aboriginal people.

Education and resourcing

More education and training around data collection and data linkage projects, and greater resourcing for a data linkage workforce would address, at least in part, some of the aforementioned obstacles and maximise opportunities to improve the health outcomes of Indigenous Australians. For example, data collection agencies should have culturally competent staff collecting data from Indigenous Australians and engage in respectful discussions regarding ownership of personal and community information with Aboriginal and Torres Strait Islander peoples and community organisations. Additionally, individuals should be informed about such data collections, their importance, how their data will be used, stored and the potential contributions it may add to improving health, planning and service delivery.

Box 1: Data linkage units in Australia

The Public Health Research Network is a collaboration of six state/territory data linkage units in Western Australia, New South Wales, South Australia/Northern Territory, Tasmania, Queensland and Victoria, and two national linkage units, namely the Centre for Data Linkage based in Western Australia and the Australian Institute of Health and Welfare in Canberra.⁵ The Data Linkage Unit (DLU) in WA has existed for more than 20 years and others have commenced in the past 10 years.^{5,6} Researchers can apply for datasets related to the same individual to be linked and provided without identifiers, ensuring that privacy is not breached. The process requires initial approval from the data custodian, then ethical approval through various human research ethics committees (HRECs) in each state or territory and, in the case of research for Indigenous people, Aboriginal HREC approvals may also be required. The process of obtaining HREC approval differs in each jurisdiction— some require a national ethics application form; others accept a form if already approved in another jurisdiction, and still others require a specific application form.

Case study

Cervical cancer incidence and mortality have halved in Australia since the introduction of the National Cervical Screening Program in 1991, yet Indigenous women remain twice as likely to get cervical cancer and four times more likely to die from it.⁷ The program is unable to report on cervical screening participation for indigenous women as indigenous status is not universally recorded in the state Pap test registries which provide monitoring data to the program. In 2011, we commenced a national project to link Pap test registries datasets within each state and territory to health datasets containing an indigenous identifier, in order to assess participation of indigenous women in the program. Ethical approval was required from 10 state-based HRECs and three Aboriginal HRECs. In addition, regulatory approval was required from seven data linkage units (DLU) and the custodians of 24 datasets.

Unreasonable time and financial cost for ethical and linkage approval

The time from initiation to completion of the ethics committee approval process ranged from two to 32 months, and final approval to link and access all datasets took five years. In one jurisdiction, a data custodian provided conditional approval pending ethics committee approval; by the time the HREC approval was received, a new employee held the data custodian position and the conditional approval was deemed to be invalid, requiring the approval process to start afresh. The first set of data was obtained in December 2013 and one data set remains outstanding as of April 2016 (figure 1). While the professionalism, support and thoroughness of almost all individuals involved has been exemplary, the process has been fraught with duplication, ineffective regulation and delay.

Figure 1: Time from commencement of application or data acquisition to completion

	2011	2011	2012	2013	2014	2015	2016
HREC	April				Dec		
Data Linkage Unit		Sept					Jan
Data acquisition				Dec			*

* one state outstanding as of April 2016

Over 400 days of person time, at a cost in excess of \$200,000, were spent obtaining HREC and DLU approval. Datasets contained different variables or the same variables with different naming conventions; the researchers worked with the DLU to obtain variables that were necessary to answer the project's research questions. In some jurisdictions, a request for each variable had to be justified and negotiated and, where changes to the original request were necessary (either researcher or DLU driven), an amendment was required by the relevant ethics committees and/or data custodians.

Results

The first results, based on population data (1,334,795 women aged 20-69 years) from one jurisdiction, show that Indigenous women have a 20-point lower screening participation rate than other Australian women, with no improvement over time,⁸ and a higher rate of high-grade cervical abnormalities.⁹ Had the process been more efficient and less protracted, results for the whole country would have been available by now, information which could have underpinned interventions to reduce cervical cancer occurrence in Indigenous women.

Discussion and conclusion

This paper highlights some of the obstacles encountered by researchers using data linkage to answer important research questions regarding the health of Indigenous Australians. Australia has many publicly funded data holdings, including clinical dataset registries, administrative databases and survey data, access to which can lead to improvements in public health. Although population level data exist, access is so complex that researchers are taking longer to achieve results that can underpin interventions and improve outcomes for Indigenous Australians. This is an ethical concern for many researchers.

Ethical implications

By its very nature, data linkage allows researchers to use data that has been de-identified and it is therefore highly unlikely that an individual's personal health data could be made public. The Privacy Act provides a mechanism to allow such research to go forward as long as the relevant HREC approvals are in place.¹⁰ The use of big data and conduct of data linkage projects should also be guided by the values and ethics in conducting Aboriginal and Torres Strait Islander research.¹¹

A recent National Health and Medical Research Council report stated: "It is particularly important that the use of Aboriginal and Torres Strait Islander data maximises opportunities to improve health outcomes for this population group."¹² While researchers must at all times respect the individuals whose information is contained within these data, and abide by the legislative structures governing their use, measures to streamline data linkage processes are required. Key to this is the critical need for researchers to establish and build relationships with

indigenous groups to ensure that indigenous status is accurately recorded in health and census data in the first place and to facilitate/navigate/expedite approval and compliance requirements. Mistrust between indigenous people, communities, data custodians and researchers, which could be addressed through better education about why data are being collected, how data are being used and stored, who benefits and how findings will be disseminated.^{13,14} Further, establishing a national set of guidelines for sharing de-identified data collected from indigenous communities has the potential to prevent unnecessary duplication in data collection and maximise health benefits for indigenous people.¹⁵

Multi-jurisdictional data linkage is in its infancy in most of Australia and the ability of services to provide linked data in a timely manner varies across the country. There is no doubt that data linkage projects have an increasingly important role to play in health care planning and providing a more complete picture of the health of Aboriginal and Torres Strait Islander health outcomes, without the time and cost burden of gathering additional and often duplicate data.

Notwithstanding length and complexity of data linkage projects, the case study presented here is an exemplar of what can be achieved to address a significant gap in reporting of indigenous people's participation in a national cancer screening program that has been in operation for 25 years. A firm foundation has been established – the challenge now is to build on it.

Acknowledgments: Our sincerest thanks to all those who provided input into this article. Thanks also to Dr Bronwyn Morris and Professor John Condon from the Menzies School of Health Research for their review of this article.

Funding: We would like to express our gratitude to the organisations that helped support this work, including the Lowitja Institute, the Centre of Research Excellence in Discovering Indigenous Strategies to improve Cancer Outcomes Via Engagement, Research Translation and Training (DISCOVER-TT CRE, funded by the National Health and Medical Research Council #1041111), and the Strategic Research Partnership to improve cancer control for Indigenous Australians (STREP Ca-CIndA, funded through Cancer Council NSW (SRP 13-01), with supplementary funding from Cancer Council WA. The views expressed in this publication are those of the authors and do not necessarily reflect the views of the funding agencies. Liz Izquierdo was employed by the Lowitja Institute at the time the research was conducted.

References

1. Holman CD, Bass AJ, Rosman DL, et al. A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system. *Australian health review* : a publication of the Australian Hospital Association. 2008;32(4):766-77.

ETHICS IN CANCER

- Rosman D, Spilsbury K, Alan J, et al. Multi-jurisdictional linkage in Australia: proving a concept. *Australian and New Zealand journal of public health*. 2016;40(1):96-7.
- Pariyadan MV, Hotham ED, Turner S, et al. Ethics and compliance hurdles in conducting multicentre low-risk research. *The Medical journal of Australia*. 2015;203(8):324-5.
- Roberts LM, Bowyer L, Homer CS, et al. Multicentre research: negotiating the ethics approval obstacle course. *The Medical journal of Australia*. 2004;180(3):139.
- Population Health Research Network (PHRN). National Data Linkage Units Curtin Health Innovation Research Institute (CHIRI) at Curtin University, Western Australia, 2016 [15 March 2016].
- Holman C, Bass A, Rouse I, et al. Population-based linkage of health records in Western Australia: development of a health services research linked database. *Australian and New Zealand journal of public health*. 1999;23(5):453-9.
- Australian Institute of Health and Welfare. *Cervical screening in Australia 2012-2013*. Canberra: AIHW, 2015.
- Whop LJ, Garvey G, Baade P, et al. The first comprehensive report on Indigenous Australian women's inequalities in cervical screening: A retrospective registry cohort study in Queensland, Australia (2000-2011). *Cancer*. 2016:n/a-n/a.
- Whop LJ, Baade P, Garvey G, et al. Cervical Abnormalities Are More Common among Indigenous than Other Australian Women: A Retrospective Record-Linkage Study, 2000-2011. *PLoS one*. 2016;11(4):e0150473.
- Australian Law Reform Commission. *Australian Privacy Law and Practice (ALRC 108) Section 66.29 Research: Databases and Data Linkage*. Canberra: 2008.
- National Health and Medical Research Council. *Values and Ethics: Guidelines for Ethical Conduct in Aboriginal and Torres Strait Islander Health Research*. Canberra 2003.
- National Health and Medical Research Council. *Principles for accessing and using publicly funded data for health research*. Canberra: NHMRC, 2015.
- National Health and Medical Research Council. *Keeping research on track: A guide for Aboriginal and Torres Strait Islander peoples about health research ethics*. Canberra: NHMRC, 2005.
- National Health and Medical Research Council. *The NHMRC Road Map II: A strategic framework for improving the health of Aboriginal and Torres Strait Islander people through research*. Canberra: NHMRC 2010.
- Wang Z. Data sharing in Indigenous health research: guidelines needed. *The Medical journal of Australia*. 2015;203(1):12-3.