

# Optimising the use of linked administrative data for infectious diseases research in Australia

Hannah C Moore<sup>a,d</sup> and Christopher C Blyth<sup>a,b,c</sup>

<sup>a</sup> Wesfarmers Centre of Vaccines & Infectious Diseases, Telethon Kids Institute, University of Western Australia, Perth

<sup>b</sup> Department of Infectious Diseases, Perth Children's Hospital, Western Australia

<sup>c</sup> Division of Paediatrics, School of Medicine, University of Western Australia, Perth

<sup>d</sup> Corresponding author: [hannah.moore@telethonkids.org.au](mailto:hannah.moore@telethonkids.org.au)

## Article history

Publication date: June 2018

Citation: Moore HC, Blyth CC. Optimising the use of linked administrative data for infectious diseases research in Australia. Public Health Res Pract. 2018;28(2):e2821810. <https://doi.org/10.17061/phrp2821810>

## Key points

- Population-based data linkage studies can guide policy development, especially in the field of infectious diseases and vaccination policy
- Engagement needs to be improved across the research community with relevant stakeholders, policy makers and community groups for data linkage studies to reach their full potential

## Abstract

Infectious diseases remain a major cause of morbidity in Australia. A wealth of data exists in administrative datasets, which are linked through established data-linkage infrastructure in most Australian states and territories. These linkages can support robust studies to investigate the burden of disease, the relative contribution of various aetiological agents to disease, and the effectiveness of population-based prevention policies – research that is critical to the success of current and future vaccination programs.

At a recent symposium in Perth, epidemiologists, clinicians and policy makers in the infectious diseases field discussed the various benefits of, and barriers to, data-linkage research, with a focus on respiratory infection research. A number of issues and recommendations emerged. The demand for data-linkage projects is starting to outweigh the capabilities of existing data-linkage infrastructure. There is a need to further streamline processes relating to data access, increase data sharing and conduct nationally collaborative projects. Concerns about data security and sharing across jurisdictional borders can be addressed through multiple safe data solutions.

Researchers need to do more to ensure that the benefits of linking datasets to answer policy-relevant questions are being realised for the benefit of community groups, government authorities, funding bodies and policy makers. Increased collaboration and engagement across all sectors can optimise the use of linked data to help reduce the burden of infectious diseases.

## Background

Infectious diseases, in particular respiratory infections, remain a prominent cause of childhood morbidity in Australia. With frequently changing vaccination recommendations and policy changes for seasonal influenza and pertussis, and ongoing clinical trials for new vaccines such as Respiratory Syncytial Virus and parainfluenza virus<sup>1,2</sup>, robust and timely data are critical to support more accurate measurement of the burden of disease. Understanding the relative contribution of multiple aetiological agents to respiratory infections in children can also support policy development. Large amounts of data exist in administrative records. Bringing these disparate pieces of information together through population-based data linkage can enable robust studies that facilitate vaccination policy development and implementation. On 9 May 2017, a group of epidemiologists, public health physicians, clinicians, microbiologists and policy makers from around Australia gathered for a 1-day symposium at the Telethon Kids Institute in Perth, Western Australia (WA). They discussed current research using linked administrative data in Australia and future directions for population-based studies using administrative data for respiratory infection research.

## Availability of administrative data for linkage projects

Data for respiratory infection research are available across the majority of Australian jurisdictions (Table 1). Many recently completed or current studies have focused on hospital morbidity or death data, which is limited to the severe end of the disease spectrum. In WA, the value of linking routine pathology data to hospitalisation, notifications and other morbidity datasets has been realised.<sup>3</sup> Not all respiratory pathogens (or nonrespiratory pathogens) are notifiable, and inclusion of routine pathology data provides the only accurate source of detection for these pathogens.

Routine laboratory data can also be useful in interpreting trends in microbiological testing, especially with the recent increase in molecular testing<sup>4,5</sup>, as they include information on test negatives rather than only positive detections from notifiable data. Correct interpretation of microbiological data, taking into account testing algorithms and changes in test sensitivity, can be challenging. However, challenges can be overcome by effective communication and collaboration with microbiologists. Our experience linking routine laboratory data with hospitalisation data in WA has highlighted

**Table 1.** Availability of administrative datasets used for existing infectious diseases research projects across Australia

Dataset	Jurisdiction (linkage authority)					
	WA (WADLB)	NT (SA/NT)	Qld (Qld Health)	NSW (CHeReL)	Vic (VDL)	National (AIHW)
Birth/death register	✓	✓	✓	✓	✓	
Perinatal	✓	✓	✓	✓	✓	
Hospital	✓	✓	✓	✓	✓	
Emergency department	✓		✓	✓	✓	
Neonatal intensive care	✓			✓		
Notifiable infections	✓	✓	✓	✓	✓	
Routine pathology	✓		✓			
Immunisation	✓(antenatal)	✓(child)	✓(child)			✓
General practice						✓

AIHW = Australian Institute of Health and Welfare; CHeReL = Centre for Health Record Linkage; Qld Health = Queensland Health; SA/NT = SA NT (South Australia Northern Territory) Datalink; VDL = Victorian Data Linkage; WADLB = Western Australian Data Linkage Branch  
 Note: National = Australian Government  
 WADLB: [www.datalinkage-wa.org](http://www.datalinkage-wa.org); SA NT Datalink: [www.santdatalink.org.au](http://www.santdatalink.org.au); Qld Health: [www.health.qld.gov.au/hsu/link/datalink](http://www.health.qld.gov.au/hsu/link/datalink);  
 CHeReL: [www.cherel.org.au](http://www.cherel.org.au); VDL: [www2.health.vic.gov.au/about/reporting-planning-data/the-centre-for-victorian-data-linkage](http://www2.health.vic.gov.au/about/reporting-planning-data/the-centre-for-victorian-data-linkage);  
 AIHW: [www.aihw.gov.au/our-services/data-linkage](http://www.aihw.gov.au/our-services/data-linkage)

the inadequacies of hospital discharge coding for accurately identifying infectious diseases. The same applies to burden of disease estimates for non-notifiable pathogens, which are critical for vaccine development.<sup>6</sup> There is now scope for more nationally representative studies to be conducted. With the availability of data-linkage infrastructure throughout Australia, including other jurisdictions with routine pathology data, infectious diseases data-linkage studies could be improved.

## Barriers to cross-jurisdictional linkage

Inclusion of immunisation data from the Australian Childhood Immunisation Register (now known as the Australian Immunisation Register) is critical for infectious disease population-based research. Primary care data available through the Australian Government's universal health insurer, Medicare, has not been used to assess the burden of infectious diseases at the lower end of the disease severity scale but has been successfully used for diabetes research.<sup>7</sup> These national datasets are available for linkage through approved integrating authorities: the Australian Bureau of Statistics, the Australian Institute of Health and Welfare and the Australian Institute of Family Studies. In recent times, access to, and linkages between, state and national data have been hampered by lengthy approval processes.<sup>8</sup> These involve multiple ethics bodies and linkage units, with processing times from project development to data receipt of more than 2 years. Even projects using only state-held data can be delayed through the approval and linkage processes, sometimes because of the reluctance of data custodians to release data. To achieve timely policy decisions such as changes to vaccination programs, particularly for at-risk groups, delays in the conduct of projects need to be addressed.

With an increasing workforce of record linkage-trained health researchers, the availability of nonhealth datasets (e.g. education data), and the increasing realisation of the benefits of linking data for population-based research, the demand for data linkage for public health research is outweighing the capability of our data-linkage infrastructure. There needs to be a streamlined and collaborative approach to data-linkage projects throughout Australian jurisdictions. The Population Health Research Network is working towards this goal. There also needs to be better communication from the research community – public health physicians, clinicians and epidemiologists applying for data – to individual data custodians, so that custodians can better understand how data will be used for research.

## Solutions to overcome the barriers

The above-mentioned barriers need to be overcome to ensure future research projects can address critical research questions in a timely fashion. The option to establish large-scale data repositories, which regularly link data held in both state and national collections, needs to be explored and would greatly increase time efficiencies in projects involving multiple sources of data. Governance structures for access to data held within such repositories would need to be developed, with guidelines for who owns the data, who approves access, and funding mechanisms to support ongoing linkage infrastructure.

One reason for delays in multijurisdictional data linkage projects, and the apparent reluctance to release data, is data security and data sharing across borders. There are now multiple solutions for safe data access, which maintain confidentiality and privacy. These technologies either allow for access to unit record data through secure remote data access laboratories for accredited research personnel (e.g. the Secure Unified Research Environment established by the Sax Institute in 2012<sup>9</sup>), or creation of virtual pooled datasets for complex analyses, from individual study locations through a secure web portal (e.g. ViPAR<sup>10</sup> and DataSHIELD<sup>11</sup>). These data-sharing mechanisms need to be embraced by the research community, data custodians and linkage units. Researchers commit a great deal of time to data coding and cleaning algorithms to analyse data. Through improved data sharing, researchers should be encouraged to share their coding algorithms and work collaboratively to achieve their desired results and the best outcomes for research. For the infectious diseases research community, forming a network or consortium could be one way of ensuring researchers are acknowledged for their intellectual input as well as decreasing the duplication of effort in cleaning, coding and analysing linked data in isolation.

## The future

The recent Productivity Commission report about data availability and use has acknowledged that lack of trust in data access processes is stifling the use and value of Australia's data.<sup>12</sup> Establishing a Data Sharing and Release Act and a National Data Custodian to guide and monitor access and use arrangements are key recommendations of the report.<sup>12</sup> Also, 23 recommendations from a recent review of data linkage in WA call for developing data-linkage legislation and a specialised Human Research Ethics Committee providing expertise about data-linkage projects.<sup>13</sup>

In future, there needs to be greater communication and collaborative effort to optimise the use of administrative data for research. Those in the infectious diseases and vaccine research fields could be leaders in this area. Multidisciplinary teams that include epidemiologists, biostatisticians, microbiologists, clinicians and community groups are needed to ensure robust policy-relevant studies are conducted that take into account the limitations and nuances of individual datasets. Other applications where linked administrative data can be of benefit, beyond burden of disease studies for the infectious diseases field, include mathematical modelling studies to investigate disease transmission, informing adaptive clinical trials, and improving clinical practice and management of paediatric infectious diseases. Researchers need to work together to ensure data-linkage projects are being conducted in the best way to benefit the Australian community, with the aim of preventing or reducing the burden of infection in children.

## Acknowledgements

We acknowledge the guest speakers at the symposium who contributed to the discussion forming the themes of this commentary: Ross Andrews, Menzies School of Health Research; Stephen Lambert, University of Queensland; Sheena Sullivan, World Health Organisation Collaborating Centre for Reference and Research on Influenza; Louisa Jorm and Heather Gidding, UNSW Sydney; Pia Hardelid, University College London; Rachel Reeves, University College London and Public Health England; James Boyd, Curtin University; Kim Carter and Tom Snelling, Telethon Kids Institute; Ben Cowling, University of Hong Kong; Kathryn Glass, Australian National University; and Peter McIntyre, National Centre for Immunisation Research and Surveillance. The symposium was funded by the Wesfarmers Centre for Vaccines & Infectious Diseases at the Telethon Kids Institute, established by a foundation grant from Wesfarmers Ltd, and a University of Western Australia Research Collaboration Award. HM and CB are funded by National Health and Medical Research Council Fellowships (1034254 and 1111596).

## Peer review and provenance

Externally peer reviewed, not commissioned

## Competing interests

None declared

## Author contributions

The symposium was cohosted by HM and CB. HM wrote the first draft of the manuscript and made revisions following peer review. CB provided further comments. Both authors approved the final version for submission.

## References

1. Englund JA, Karron RA, Cunningham CK, Larussa P, Melvin A, Yogeve R, et al. Safety and infectivity of two doses of live-attenuated recombinant cold-passaged human parainfluenza type 3 virus vaccine rHPIV3cp45 in HPIV3-seronegative young children. *Vaccine*. 2013;31(48):5706–12.
2. Higgins D, Trujillo C, Keech C. Advances in RSV vaccine research and development – a global agenda. *Vaccine*. 2016;34(26):2870–5.
3. Moore HC, Lehmann D, de Klerk N, Smith DW, Keil AD, Blyth CC. How accurate are international classification of diseases – 10 diagnosis codes in detecting influenza and pertussis hospitalizations in children? *J Pediatric Infect Dis Soc*. 2014;3(3):255–60.
4. Kaczmarek MC, Ware RS, Nimmo GR, Robson JMB, Lambert SB. Pertussis seasonality evident in polymerase chain reaction and serological testing data, Queensland, Australia. *J Pediatric Infect Dis Soc*. 2016;5(2):214–7.
5. Lim FJ, Blyth CC, Keil AD, de Klerk N, Moore HC. Using record linkage to examine testing patterns for respiratory viruses among children born in Western Australia. *Epidemiol Infect*. 2017;145(8):1688–98.
6. Lim FJ, Blyth CC, Fathima P, de Klerk N, Moore HC. Record linkage study of the pathogen-specific burden of respiratory viruses in children. *Influenza Other Respir Viruses*. 2017;11(6):502–10.
7. Comino EJ, Tran DT, Haas M, Flack J, Jorm L, Harris MF. Validating self-report of diabetes use by participants in the 45 and Up Study: a record linkage study. *BMC Health Serv Res*. 2013;13(1):481.
8. Moore HC, Guiver T, Woollacott A, de Klerk N, Gidding HF. Establishing a process for conducting cross-jurisdictional record linkage in Australia. *Aust N Z J Public Health*. 2016;40(2):159–64.
9. Sax Institute. Sydney: Sax Institute; 2017. SURE; 2017 [cited 2017 Jun 6]; [about 1 screen]. Available from: [www.saxinstitute.org.au/our-work/sure/](http://www.saxinstitute.org.au/our-work/sure/)
10. Carter KW, Francis RW, Bresnahan M, Gissler M, Grønberg TK, Gross R, et al. ViPAR: a software platform for the virtual pooling and analysis of research data. *Int J Epidemiol*. 2016;45(2):408–16.

11. Wilson R, Butters O, Avraam D, Baker J, Tedds JA, Turner A, et al. DataSHIELD – new directions and dimensions. *Data Sc J.* 2017;16:21.
12. Productivity Commission. Data availability and use: Productivity Commission inquiry report No. 82. Canberra: Productivity Commission; 2017 [cited 2017 Jul 26]. Available from: [www.pc.gov.au/inquiries/completed/data-access/report/data-access.pdf](http://www.pc.gov.au/inquiries/completed/data-access/report/data-access.pdf)
13. Data Linkage Expert Advisory Group. A review of Western Australia's data linkage capabilities: developing a whole of government model. Perth: Data Linkage Expert Advisory Group; 2016 [cited 2017 Jul 26]. Available from: [www.jtsi.wa.gov.au/docs/default-source/default-document-library/a-review-of-western-australia's-data-linkage-capabilities---developing-a-whole-of-government-model---december-2016.pdf?sfvrsn=0](http://www.jtsi.wa.gov.au/docs/default-source/default-document-library/a-review-of-western-australia's-data-linkage-capabilities---developing-a-whole-of-government-model---december-2016.pdf?sfvrsn=0)

Copyright: 

© 2018 Moore and Blyth. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Licence, which allows others to redistribute, adapt and share this work non-commercially provided they attribute the work and any adapted version of it is distributed under the same Creative Commons licence terms. See: [www.creativecommons.org/licenses/by-nc-sa/4.0/](http://www.creativecommons.org/licenses/by-nc-sa/4.0/)