Use of data linkage to improve communicable disease surveillance and control in Australia: existing practices, barriers and enablers

Stacey L. Rowe,[1,2] Nicola Stephens,[1] Benjamin C. Cowie,[1,3] Terry Nolan,[4] Karin Leder,[2] Allen C. Cheng[2]

Abstract

Objectives: To review the use of data linkage by Australian state and territory communicable disease control units, and to identify barriers to and enablers of data linkage to inform communicable disease surveillance and control activities.

Methods: Semi-structured telephone interviews were carried out with one key informant from communicable disease control units in all eight Australian states and territories between October 2017 and January 2018.

Results: Key informants from all Australian states and territories participated in the interview. A variety of existing practices were identified, with few jurisdictions making systematic use of available data linkage infrastructure. Key barriers identified from the review included: a lack of perceived need; system factors; and resources. Existing regulatory tools enable data linkage to enhance communicable disease surveillance and control.

Conclusions: We identified considerable variation in the use of data linkage to inform communicable disease surveillance and control activities between jurisdictions. We suggest that routinely collected, disparate data are systematically integrated into existing surveillance and response policy cycle to improve communicable disease prevention and control efforts.

Implications for public health: Existing gaps in communicable disease surveillance data may affect prevention and control efforts. Data linkage is recognised as a valuable method to close surveillance gaps and should be used to enhance the value of publicly held health data.

Key words: data linkage, communicable disease, surveillance, infectious diseases

In Australia, communicable diseases are monitored primarily through the surveillance of scheduled notifiable conditions operating under the governance of state and territory health departments. The system is underpinned by robust public health legislation at the jurisdictional level requiring medical practitioners and/or pathology services to notify a range of communicable diseases of public health importance. A subset of jurisdictional data, excluding identifiers, are transmitted to the National Notifiable Disease Surveillance System (NNDSS) under the auspices of the *National Health Security Act 2007*,[1] operated by the Australian Department of Health.

The notification system has two purposes: to enable a swift public health response to prevent further cases of disease; and to monitor disease epidemiology over time to inform public health policy. Gaps in surveillance data may have an impact on communicable disease prevention and control efforts and may limit optimal development of evidence-based policy. Three important examples are Aboriginal and Torres Strait Islander status (missing in 50% of NNDSS cases), mortality (missing in 56% of cases) and vaccination status (missing in 17% of relevant cases).[2] These gaps impact on estimates of disease burden in important populations[3-5] and hinder evaluation of vaccination programs.[16] Traditionally, missing communicable disease surveillance data are addressed through active case follow-up, whereby cases and/or notifying practitioners are interviewed by trained public health practitioners to obtain the missing data. However, this approach is resource intensive and often impracticable, particularly in view of the large volume of cases being notified to jurisdictional health departments.[6,7] Data linkage therefore represents another method for closing surveillance gaps.

Data linkage (also known as record linkage, data matching or integration) involves bringing together information about an individual or an event from disparate sources (Box 1).[8,9] Use of data linkage in Australia has been steadily increasing as a relatively inexpensive method of conducting epidemiological research or monitoring health at a population level.[10] It is often used to improve understanding of chronic diseases including cancer, cardiovascular and respiratory diseases, as well as for injury, and mother and child health research, where long-term monitoring, treatment or care is required.[10-13] Most data linkage studies in

1. Department of Health and Human Services, Health Protection Branch, Victoria
2. School of Public Health and Preventive Medicine, Monash University, Victoria
3. Doherty Institute for Immunity and Infection, WHO Collaborating Centre for Viral Hepatitis, Victoria
4. Melbourne School of Population and Global Health, The University of Melbourne, Victoria
Correspondence to: Ms Stacey L. Rowe, Department of Health and Human Services, Health Protection Branch, 50 Lonsdale Street, Melbourne, VIC 3000; e-mail: stacey.rowe@dhhs.vic.gov.au
Submitted: April 2018; Revision requested: July 2018; Accepted: October 2018
The authors have stated they have no conflict of interest.

Aust NZ J Public Health. 2019; 43:33-40; doi: 10.1111/1753-6405.12846

Australia are conducted in New South Wales (NSW) and Western Australia (WA), where linkage infrastructure is well developed and utilised.[10,11,13] In the communicable diseases setting, data linkage can be used to evaluate public health interventions and guide policy development.[14] In Australia, data linkage methodologies have been used to quantify case ascertainment in communicable disease registers;[15] to improve estimates of selected communicable diseases;[16-19] to better describe the burden of selected communicable diseases among Aboriginal and Torres Strait Islander people;[3-5,20] to improve communicable disease morbidity and mortality estimates;[21-23] and to evaluate immunisation programs.[24-27] These are all important examples of how data linkage can be applied to support communicable disease research in Australia, and results from these studies are often used to inform policy making and program improvement.[14] However, there is comparatively little use of data linkage by jurisdictional health departments to support communicable disease surveillance and control activities in an ongoing manner or to systematically address underlying gaps in communicable disease data collections. This is despite often having record-level access to a wide range of publicly held and usefully relevant health datasets to support their operations.

In 2016, the Australian Productivity Commission (Commission) – the Australian Government's independent research and advisory body – was tasked with undertaking an investigation into the cost and benefits for improving data availability and use in Australia. The scope of the review included the sharing and linking of publicly held data for national economic development and improved service provision, including that of health services.[10] In 2017, the Commission recommended that Australia establish "enduring linkage systems" to increase the utility of linked health data for the development of public policy, to minimise duplication, and for jurisdictional comparisons and cross-jurisdictional research.[10] In response, we sought to understand how communicable disease control units in each Australian state and territory use data captured in disparate datasets to inform communicable disease surveillance and control activities, and whether data are routinely linked. The aims of the study were to identify the use of data linkage by communicable disease control units in each Australian jurisdiction, and to determine barriers to, and enablers of, the use of data linkage to support communicable disease surveillance and control activities.

Methods

Individuals who were responsible for managing or overseeing surveillance systems in communicable disease control units in each Australian state and territory were selected as key informants in this study. Semi-structured telephone interviews were carried out with these key informants between October 2017 and January 2018.

Interview questions were grouped into the following themes: availability and use of existing data linkage infrastructure; availability and use of disparate datasets (either linked or unlinked); existing governance arrangements to support data linkage; and existing or planned priorities relating to the use of data linkage. Key informants were also asked to identify barriers to and enablers of the use of data linkage in the communicable disease setting.

Notes taken during interviews were reviewed and clarification sought from key informants if required. Responses from each jurisdiction relating to their availability and use of existing data linkage infrastructure, as well as their availability and use of disparate datasets were tabulated. Commonly cited barriers to and enablers of data linkage were synthesised into overarching themes where possible.

In this study, we use the term 'jurisdiction' interchangeably with the term 'state and territory'.

Results

Representatives from all eight Australian states and territories participated in the interview (see Acknowledgements, noting that the key informant for Victoria was author, SR). We identified a variety of existing practices in each jurisdiction's use of data linkage.

Integration of notifiable communicable disease data into existing linkage infrastructure

Communicable disease notification data were incorporated into a consolidated master linkage system (MLS) operated by Data Linkage Units (Box 1) in five of the eight Australian states and territories: the Australian Capital Territory (ACT), New South Wales (NSW), Queensland (QLD), Victoria (VIC) and Western Australia (WA), see Table 1(A).

Additionally, in NSW the MLS – maintained by the Centre for Health Record Linkage (CHeReL) – is used to create a separate Communicable Disease Register (CDR). The CDR is distinct from their local notifiable disease register, Notifiable Conditions Information Management System (NCIMS), and contains de-identified linked data integrating NCIMS and Human Immunodeficiency Virus (HIV) database with three selected disparate datasets containing emergency department presentations, hospital admissions and death registrations. It also contains records from these datasets based on predefined International Classification of Disease (ICD-10-AM) and Systematized Nomenclature of Medicine (SNOMED) codes relating to notifiable communicable diseases even if a corresponding person record is not captured on NCIMS.[28] CDR was created under the Public Health and Diseases Registers provision of the *Public Health Act*[29] to support communicable disease policy, planning and operations.

Availability of data from disparate datasets

Most commonly, communicable disease data were linked to administrative datasets (those containing hospital-based care such as admitted episodes and emergency department presentations) and vital statistics (deaths). None of the jurisdictions reported having their communicable disease notification data routinely linked to national datasets such as the Medicare Benefits Schedule (MBS), the Pharmaceutical Benefits Scheme (PBS) or the Australian Immunisation Register (AIR), although some jurisdictions (NSW, QLD, VIC, WA) reported current or planned *ad hoc* linkage and analysis of the MBS and PBS. Several jurisdictions reported having regular access to disparate datasets or databases that could be interrogated on a case-by-case basis (herein referred to as unlinked data) to support communicable disease surveillance and control activities. All jurisdictions reported having access to AIR, and six reported having selected access to local hospital administrative systems (ACT, Northern Territory [NT], QLD, South Australia [SA], Tasmania [TAS]) and WA, Table 1[B]).

Uses of data from disparate datasets

Use of the datasets varied considerably by jurisdiction (Table 1[C]) and none reported using linked data for the purpose of updating missing data in local communicable disease registers (sometimes referred to as 'data cleansing' [Box 1]).

Among the five jurisdictions that reported having their data incorporated into an MLS, few cited routine applications of the linked data. WA first incorporated communicable disease notification data into an MLS in 2007. However, the data obtained through the linkage process are seldom used operationally nor are the returned, linked data incorporated into their local communicable disease notification register. The ACT and NSW incorporated communicable disease notification data into an MLS in 2014. In the ACT, however, the communicable disease data are not routinely refreshed (updated), so linked content data from disparate datasets are not used. In NSW, linked data from the CDR is routinely used to improve sexually transmissible disease estimates among Aboriginal and Torres Strait Islander people.[30] Other planned uses included identification of co-infections and other comorbidities associated with communicable diseases.[28] Because CDR contains only de-identified data and is distinct from the NCIMS, data obtained via the linkage process are not incorporated into NCIMS nor transmitted

Table 1: (A) Data linkage infrastructure and availability of linked communicable diseases data; (B) Linked and unlinked data types available and used in the setting of communicable disease surveillance and control; (C) Uses of linked or unlinked data by communicable disease control units.

	State / Territory							
	ACT	**NSW**	**NT**	**QLD**	**SA**	**TAS**	**VIC**	**WA**
(A) Infrastructure								
Data linkage unit	CHeReL	CHeReL	SA-NT DataLink	DLQ	SA-NT DataLink	TDLU	CVDL	WADLB
Year communicable disease data incorporated into MLS	2014	2014	n/a	2017	n/a	n/a	2017	2007
Years for which communicable disease data are available	2000 -	1993 – Dec 2016	n/a	1989 -	n/a	n/a	1990 -	1990 -
Frequency of data refresh	ad hoc	annual	n/a	tbd	n/a	n/a	quarterly	annual
(B) Data types **Linked ● Unlinked ○**								
ED presentations		●					●	●
Hospital admissions	○	●	○	○	○	○	●	●/○
Death registrations	○^	●^	○^		○^		●	●
Laboratory data+	○		○	○				
Immunisation records	○	○	○	○	○	○	○	○
PBS / MBS records								
(C) Use								
Aggregate analyses	○	●/○	○	●/○	○	○	●/○	●/○
Data cleansing†	○	○	○	○	○	○	○	○
Contact details	○		○	○	○	○		○
Aboriginal and Torres Strait Islander	○		○	○		○		○
Alive / Died	○		○	○	○	○		○
ED or hospital attendance	○		○	○	○	○		○
Risk factors, comorbidities	○		○	○	○			○
Manifestation, outcome	○		○	○	○			○
Vaccination status	○	○	○	○	○	○	○	○
Laboratory results	○		○	○				
Public health investigations	○	●/○	○	●/○	○	○	●/○	●/○
Case ascertainment					○			

Notes:

^ Selected death registrations only. ACT receive respiratory deaths using a methodology previously described.[64]

+ Denotes access to one or more databases maintained by a public laboratory.

† "Data cleansing" involves the large-scale comparison or matching of two or more sets of personal data (either held by the same organisation or by different organisations), for the purposes of updating one or both of the sets.[63] In this context, data cleansing refers to the updating of missing surveillance data from local notifiable communicable diseases registers with content data gleaned through linkage with or interrogation of disparate datasets.

CHeReL – Centre for Health Record Linkage; SA-NT DataLink – South Australia and Northern Territory DataLink; DLQ – Data Linkage Queensland; TDLU – Tasmanian Data Linkage Unit; CVDL – Centre for Victorian Data Linkage; DLB – Data Linkage Branch; MLS – Master Linkage System; n/a – not applicable; tbd – to be determined; ED – Emergency Department; PBS - Pharmaceutical Benefits Scheme; MBS - Medicare Benefits Scheme.

to the NNDSS. QLD was unable to cite any existing uses of routinely returned linked data as communicable disease notification data had only recently been incorporated into a local MLS (2017). Similarly, VIC incorporated communicable disease notification data into a local MLS in 2017 and indicated that routine use of linked data was planned for 2018, including using linked data to improve ascertainment of Aboriginal and Torres Strait Islander identification, and to ascertain hospital-based care utilisation mortality outcomes.

Most jurisdictions indicated that data linkage methodologies were used for *ad hoc* public health investigations[31] or descriptive epidemiological analyses.[3,32,33] These uses involved engagement of a Data Linkage Unit,[3] development of user-built data linkage programs/look-up algorithms,[32,33] or use of publicly available data linkage software.[34] More commonly, linked data are used to address specific *ad hoc* research questions that are initiated and carried out by research institutions in collaboration with, or distinct from, communicable disease units.[4,5,21,22,25,35]

In contrast, jurisdictions indicated that data obtained from unlinked databases are regularly used for data cleansing (Box 1) to supplement or improve existing communicable disease surveillance data. The most commonly reported use of data among jurisdictions reporting access to unlinked hospital administrative databases (ACT, NT, QLD, SA, TAS, WA) was to obtain missing demographic and clinical data relating to a notifiable disease, such as Aboriginal and Torres Strait Islander status. Data gleaned in this manner is then manually incorporated into local notifiable disease registers. Other uses included the ability to obtain updated case contact details, and to identify clinical (e.g. manifestation and outcome) or risk factor and comorbidity data. Hospital datasets were sometimes used to ascertain case deaths. Some jurisdictions also reported having access to unlinked mortality data (ACT, SA, NT), whereby selected death registrations (those potentially due to a notifiable communicable disease) are manually cross-checked against the notifiable diseases register, which is then updated accordingly. In SA, unlinked mortality data are also used to identify previously un-notified cases of notifiable communicable disease (i.e. to improve case ascertainment).

Existing governance arrangements

Four jurisdictions (ACT, NSW, VIC, and WA) had agreements or protocols in place between communicable disease units and Data Linkage Units, clearly documenting the procedural and governance arrangements surrounding integration of communicable disease data into an MLS. Two states and territories – the ACT and NSW that are incorporated into the MLS maintained by the CHeReL – had data dictionaries describing the availability of linked notifiable disease data and corresponding variables on a publicly available website.[36] Only one state (WA) had detailed specifications relating to identification of relevant linked records with a temporal component. Hospitalisation records are identified and linked to cases of communicable diseases only if a hospitalisation (for any diagnoses) occurred within or on 10 days before or after specimen collection date (or onset date) of the notified disease. Similarly, linked death registrations were only identified and returned if the death was registered up to four days prior or 56 days following the date of onset (or specimen collection date if onset date not known). Conversely, in NSW's CDR, person-specific linkages between cases of communicable diseases and disparate datasets occur for all cases regardless of notification or onset dates.

Existing and planned priorities

Jurisdictions considered use of data linkage to improve completeness of Aboriginal and Torres Strait Islander status as a priority. Some jurisdictions (QLD, NSW, VIC) identified cross-jurisdictional data linkage to minimise duplicate notifications in the NNDSS as a priority. Two states (WA and VIC) identified as a priority the use of hospital-based care and death registration data to systematically improve morbidity and mortality estimates associated with communicable diseases. Additionally, VIC identified as a priority the use of data linkage with AIR to systematically improve completeness of immunisation history for relevant vaccine preventable diseases. Three states (QLD, WA, VIC) identified the use of PBS data as a priority to evaluate uptake and effectiveness of pharmaceutical products against selected communicable diseases.

Barriers to and enablers of data linkage

A range of barriers and enablers were identified. These were broadly grouped into the following themes: Perceived need; Regulatory tools; System factors; and Resources.

Perceived need

Perceived need was considered both a prohibiting and enabling factor to the use of data linkage in the communicable disease setting. Many of the jurisdictions (ACT, NT, SA, TAS) indicated that pertinent information relevant for communicable disease surveillance and control activities could easily be gleaned on a case-by-case basis via access to unlinked databases, thus precluding the need for data linkage. All of these informants acknowledged that data linkage would in theory be useful for high-volume conditions,

Box 1: Key concepts and terms relating to data linkage.

Data linkage

Data linkage (also known as data matching or integration) involves the bringing together of information about an individual or an event from disparate sources.[6,9] The term disparate in this context means 'different', 'separate' or 'distinct'.

Data cleansing

'Data cleansing' involves the large-scale comparison or matching of two or more sets of personal data (either held by the same organisation or by different organisations), for the purposes of updating one or both of the sets.[63]

Data linkage infrastructure and networks in Australia

In 2009, the Australian Government's National Collaborative Research Infrastructure Strategy funded the creation of a Public Health Research Network (PHRN). The network comprises Data Linkage Units servicing each Australian state and territory.

Data Linkage Units

There are six data linkage units involved in the PHRN collaboration that have the infrastructure and capability to service all Australian states and territories. Data Linkage Units are located within State or Territory Health Departments (WA, NSW, VIC, QLD) or in university-affiliated institutions (TAS, SA). One Data Linkage Unit (Australian Institute of Health and Welfare) has authority to link national data collections.

Master Linkage System

Also referred to as Master Linkage File, it is a system containing continuously updated links between datasets containing both person and family-based information. These systems are usually maintained and managed by Data Linkage Units.

Master Linkage Key

Also known as a data linkage key or statistical linkage key, a code that is constructed using personally identifying information (such as name, date of birth and address) that uniquely identifies a person or family and is used to link records belonging to the same person from multiple and disparate datasets.

but also indicated that establishing a system to routinely integrate disparate data into local notifiable disease registers was not a current priority.

In contrast, the remaining jurisdictions (WA, NSW, QLD, VIC) – which serve larger population sizes – reported a strong desire for using data linkage to support communicable disease surveillance and control activities in Australia. Most of these jurisdictions also expressed a concomitant desire for the development of a nationally standardised approach for establishing linkages with notifiable communicable diseases, as well as for standardised analytical approaches.

Regulatory tools

Few of the informants cited regulatory tools as a prohibiting or enabling factor for the use of data linkage in the communicable disease setting. NSW made reference to the *NSW Public Health Act 2010*, which allows for the establishment of public health registers for a range of purposes including "to facilitate the care, treatment and the follow up of persons who has or may have an infectious disease, or to facilitate the identification of risk factors or outcomes associated with a disease".[29] VIC made reference to national[1] and state-based[37] legislation governing data collection and use specifically for public health purposes, noting that the powers of the Chief Health Officer under the *Victorian Public Health and Wellbeing Act 2008*[37] were recently used to establish the integration of notifiable disease data (captured on Victoria's Public Health Events Surveillance System – PHESS) into a local MSL. VIC also noted national[38] and state-based[39] privacy legislation governing the handling of personal information, as well as data sharing legislation supporting sharing and use of public sector data to guide policy making, service planning and design.[40-42]

Some of the informants (VIC, QLD, TAS) discussed potential privacy concerns as a potential barrier to the use of data linkage if the linkage was to be used for 'data cleansing' purposes (Box 1). Integrating content data – that is, data sourced from disparate datasets through the linkage process – with notifiable disease registries would essentially render the individuals re-identified. However, all of these informants also expressed a view that the public health benefits of data cleansing would outweigh the risks posed by the re-identification process, citing national and jurisdictional privacy legislation and associated privacy principles as enabling factors.[38,39] They also noted that jurisdictions already have access to a range of publicly held record-level data to support communicable disease surveillance and control activities, so the use of data linkage methodologies in this context simply improves the efficiency of this established practice.

System factors

A range of system factors were identified as potential barriers to the use of data linkage. First, NSW, QLD and VIC specifically made note of Australia's federated political system as a barrier to the use of data linkage to improve completeness of key data items in the NNDSS. Collection of personally identifiable communicable disease data rests with states and territories, and only de-identified communicable disease data are transmitted to the NNDSS. This precludes national data linkage activities by the Australian Department of Health on behalf of the states and territories for public health purposes. Second, some informants noted that the administrative coding (to ICD-10-AM codes) of hospital-based or registry-based data as a barrier to timely use of linked data for improving morbidity and mortality estimates. Third, and among jurisdictions that incorporated notifiable disease data into an MLS, scheduled delays in sending data to, or receiving data from, Data Linkage Units were also reported. These delays impacted on the acceptance of data linkage as a useful tool to augment surveillance and control practices at an operational level.

Resources

All informants identified a lack of resources as a barrier to using data linkage in the communicable disease setting. Many acknowledged that integrating notifiable disease data with disparate data in an enduring manner would be resource intensive at the outset, requiring dedicated staff to work with jurisdictional Data Linkage Units and relevant data custodians to establish data transfer and management protocols. Further, such a system would require staff within communicable disease units with skills to analyse, interpret and report on the resultant linked data. A lack of these skills was identified as a key factor precluding the use of data linkage by communicable disease units. Engagement with external agencies and research institutions with staff who are experienced in the analysis of linked health data was identified as a possible solution to this barrier.

Discussion

In this study, we identified limited use of data linkage by Australian state and territory communicable disease control units to inform local surveillance and control activities. None of the jurisdictions reported routinely making use of data linkage methodologies to improve completeness of missing notifiable disease surveillance data, nor to systematically quantify case ascertainment or reporting bias in local notifiable disease registers. Some jurisdictions reported semi-regular or *ad hoc* usage of aggregated linked health data to improve burden of disease estimates for selected communicable diseases. A majority of jurisdictions reported manually integrating data from unlinked and disparate datasets on a case-by-case basis. All jurisdictions expressed a desire for a more systematic approach to addressing gaps in communicable disease surveillance. Commonly cited priorities included using data linkage to improve completeness of Aboriginal and Torres Strait Islander, vaccination status and mortality outcomes in local and national communicable disease registers.

The inconsistent use of both linked and unlinked data has implications for the interpretation of national surveillance data, and may partially explain the considerable variability previously identified relating to the completeness of important data items captured nationally in the NNDSS[2] and between jurisdictions.[7,43] Variations in the manner in which data are sourced were evident from this review, such as for determination of mortality status. Most jurisdictions did not integrate local death registration data with notifiable disease registries: One had a data linkage process set up to do so, but the mortality data were not routinely incorporated into their local notifiable disease register; and another integrated mortality data through a systematic albeit manual process of cross-checking these two data sources. Similarly, variations in the determination of Aboriginal and Torres Strait Islander status by jurisdictions were also identified. Some jurisdictions reported using disparate datasets to improve completeness of this data item, but none did so via data linkage methodologies. Variations in data collection methods for this data item has implications for accurate burden of disease estimates among Aboriginal and Torres Strait Islander people and for consequent delivery of

targeted prevention and control initiatives to those in greatest need.[3] Data linkage is recognised as a valuable tool for improving Aboriginal and Torres Strait Islander identification, and has been identified to assist in the measurement of the Council of Australian Government Closing the Gap Performance Indicators.[44,45] Gaps in communicable disease surveillance data relating to vaccination status among relevant vaccine preventable disease notifications also remain.[2,7,43] This is despite all jurisdictions having access to AIR. The current challenges faced by communicable disease control units in integrating vaccination data from AIR into notifiable disease registries are largely logistical. Data from AIR must be looked up manually on a case-by-case basis and then transcribed into local notifiable disease registries. If an AIR record for the case is not available, vaccination status may be gleaned from the case's general practitioner, treating clinician, or from the case directly (via self-report). While this commonly occurs for selected low-incidence vaccine preventable diseases – and with good completeness – it is impracticable to carry out this laborious process among high-incidence vaccine preventable diseases. This may be one explanatory factor for gaps in vaccination status among notifiable disease registers. An alternative reason may also relate to the completeness of data within AIR, which has been shown to underestimate childhood immunisation coverage by up to 5%.[46] The use of data linkage would address the first challenge, but not the second. In our study, none of the jurisdictions reported using data linkage methodologies to improve vaccination status in local notifiable disease registries. We contend, that while the use of data linkage may not be practicable during an immediate public health response, gaps in surveillance data relating to vaccination status could be addressed retrospectively to improve completeness of this key variable in local and national surveillance systems, thereby also improving cross-jurisdictional evaluation of national immunisation programs.[24] Using the AIR in this manner would accord with the stated purposes of the register as set out in the *Australian Immunisation Register Act*.[47] Minimising inconsistencies in surveillance practices between jurisdictions could be addressed by adopting nationally consistent linkage and analytical approaches should be prioritised.

Barriers associated with cross-jurisdictional linkage identified by others include issues with timely data access, project governance and administration, as well as data security, privacy, infrastructure and capability.[10,14,48-50] We identified additional barriers including a lack of perceived need, system factors and resources. Jurisdictions with small populations reported regularly using unlinked disparate data to fill surveillance gaps, which precluded a need for data linkage methodologies to be used for data cleansing purposes. While these practices serve small jurisdictions well for rare or low-incidence notifiable conditions, this approach is unlikely to be appropriate for jurisdictions serving larger population sizes, or for high-incidence notifiable conditions such as chlamydia and gonococcal infection, campylobacter and salmonellosis, influenza, pertussis, and varicella zoster virus.[7] Improved completeness of surveillance data relating to these and other vaccine-preventable conditions is essential for vaccine program evaluation, vaccine safety and adverse event monitoring, which could be met through the systematic application of data linkage methodologies.

An additional barrier identified in this study related to the de-identified form in which communicable diseases data are transmitted to the NNDSS, thus precluding systematic data linkage for data cleansing purposes occurring at a national level. In view of this barrier, many of the key informants noted that data linkage – for the purpose of data cleansing – could in theory be conducted by the jurisdictions, with the resultant data incorporated into local notifiable disease registries for transmission to the NNDSS. The process of incorporating data gleaned from linkage activities back into local notifiable disease registers essentially renders the data 're-identified'. Despite this, many of the key informants also noted that data obtained in this manner would accord with the local and national public health legislation, which governs the collection, use and disclosure of information relating to communicable diseases for the public health purposes.[1] Such legislation also sets out a regime for jurisdictions to use data to enhance the understanding of the epidemiology of communicable diseases. These legislative tools were all identified as enabling factors to the use of data linkage in this manner. There is, however, conflict between this use of data linkage for public health and safety purposes, versus the traditional use of linked data for research purposes. Data linkage systems are built around the concept of preserving an individual's privacy while simultaneously releasing the value of stored data for de-identified health research.[51] This best practice approach is appropriate for the release of data to researchers but does little to address underlying gaps in local and national surveillance systems. Further work will be required to better understand public awareness and attitudes relating to the use of data linkage in this manner. Previous research exploring public attitudes to the linkage and sharing of health and administrative data in the UK identified that, in general, people are supportive of their data being used in this manner if there is a personal or societal benefit to do so.[52-54]

The cornerstone of good public health surveillance is the systematic collection, analysis, interpretation and use of data to inform public health policy.[55] Traditional data sources to inform communicable disease surveillance are notifications made by medical practitioners and laboratories, augmented by active case investigation and follow-up. However, as evidenced by this review, a range of disparate and routinely collected data sources are commonly used. We suggest that a more systematic approach be taken to make better use of available data to improve communicable diseases surveillance and control activities in Australia, whereby commonly used disparate data are routinely integrated and used in the surveillance and response policy cycle.[56] Specifically, we suggest that:

- Data from disparate but routinely collected data sources identified in this review are integrated into existing communicable disease surveillance systems at the jurisdictional level.
- Standardised methods for linkage and analyses of the resultant linked data are developed, piloted and implemented to improve data quality and ascertainment.
- Engagement between data custodians, linkage analysts and policy-makers within government, as well as researchers, patients and community groups external to government, is prioritised to improve interpretation, transparency and use of data to support prevention and control efforts.
- Intelligence gleaned from data linkage activities is incorporated into existing communicable disease policy and planning

networks, and national reporting and monitoring frameworks.

Such an approach would align with the national framework for communicable disease control, which identified that linkages with existing datasets should form part of a modernised surveillance system.[57] Further, it would improve availability and use of data for policy and program development as demonstrated in other parts of the world. The UK, for example, has committed to integrating a range of disparate datasets with bloodborne virus surveillance systems to inform the care and management of people tested for and diagnosed with these conditions.[58] And several countries in Europe routinely link national immunisation registers with communicable disease surveillance systems to inform vaccine safety and effectiveness studies.[59] Finland, for example, has established real-time data linkage between their national immunisation and communicable disease registers to facilitate comprehensive and timely vaccine effectiveness studies.[60]

This study was limited by its methodological design, in that we interviewed only one key informant from each jurisdiction. As such, the views expressed by our key informants may not represent the jurisdictions at large. Further, there may be data linkage activities being conducted to support communicable disease surveillance and control within jurisdictions of which our key informants were unaware. Finally, we acknowledge that the use of data linkage methodologies to improve completeness of missing data in notifiable disease registries is not a panacea for the issue of missing data, in general, and note that data items of interest gleaned from disparate datasets may themselves be missing or erroneous.

Conclusion

Australia has a robust surveillance system in place to prevent, monitor and control a range of communicable diseases of public health importance.[2] Despite this, gaps in surveillance data remain,[2,3,7] impacting on the system's utility to support the development of evidence-based prevention and control efforts. Data linkage is recognised as a valuable method to close surveillance gaps and enhance the value of publicly held health data.[10,24,48,61,62] To our knowledge, this is the first time that jurisdictional practices relating to the use of data linkage in the communicable disease setting in Australia have been described.

We identified a range of barriers to and enablers of the use of data linkage in the communicable disease setting in Australia. In the absence of a nationally integrated data linkage system[10] or interoperable communicable disease surveillance system,[57] we suggest enhancements to state and territory-based communicable disease registers whereby data from disparate and routinely collected sources are systematically integrated and used in the surveillance and response policy cycle. The development of standardised linkage and analytical approaches, as well as improved multi-sectoral and community-level engagement, will be important for the successful use of data linkage in the communicable disease setting.

Acknowledgements

The authors gratefully acknowledge the key informants interviewed as part of this study: Rebecca Hundy (ACT), Nick Rose (NSW), Peter Markey (NT), John Marques and Sonia Harmen (QLD), Emma Denehy (SA), Kerryn Lodo (TAS), Caroliene Giele (WA). We also gratefully acknowledge Angela Young of the Public Health Research Network.

References

1. National Health Security Act (Vic) 2007, s 174 of 2007. Compilation No. 15 (28 September 2007, 2008).
2. Gibney KB, Cheng AC, Hall R, Leder K. Australia's National Notifiable Diseases Surveillance System 1991–2011: Expanding, adapting and improving. *Epidemiol Infect*. 2017;145(5):1006-17.
3. Rowe SL, Cowie BC. Using data linkage to improve the completeness of Aboriginal and Torres Strait Islander status in communicable disease notifications in Victoria. *Aust N Z J Public Health*. 2016;40(2):148-53.
4. Mak DB, Watkins RE. Improving the accuracy of Aboriginal and non-Aboriginal disease notification rates using data linkage. *BMC Health Serv Res*. 2008;8:118-.
5. Watkins RE, Mak DB, Giele CM, Clews S. Aboriginal and non-Aboriginal sexually transmitted infections and blood borne virus notification rates in Western Australia: using linked data to improve estimates. *BMC Public Health*. 2013;13(1):404.
6. Gibney K, Franklin L, Stephens N. Infectious diseases notification practices, Victoria 2013. *Commun Dis Intell Q Rep*. 2016;40(3):E317-E25.
7. NNDSS Annual Report Working Group. Australia's notifiable disease status, 2014: Annual Report of the National Notifiable Diseases Surveillance System. *Commun Dis Intell Q Rep*. 2016;40(1):E48-E145.
8. Hobbs MST, McCall MG. Health statistics and record linkage in Australia. *J Chronic Dis*. 1970;23(5):375-81.
9. Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public health research. *Annu Rev Public Health*. 2011;32(1):91-108.
10. Productivity Commission. *Data Availability and Use, Inquiry Report*. Canberra (AUST): Government of Australia; 2017.
11. Tew M, Dalziel KM, Petrie DJ, Clarke PM. Growth of linked hospital data use in Australia: A systematic review. *Aust Health Rev*. 2017;41(4):394-400.
12. Public Health Research Network. *Proof of Concept Collaborations* [Internet]. Crawley (AUST): University of Western Australia PHRN; 2011 [cited 2018 Jan 19]. Available from: http://www.phrn.org.au/for-the-community/what-we-have-learnt/proof-of-concept-collaborations/
13. Brook EL, Rosman DL, Holman CDAJ. Public good through data linkage: Measuring research outputs from the Western Australian Data Linkage System. *Aust N Z J Public Health*. 2008;32(1):19-23.
14. Moore H, Blyth C. Optimising the use of linked administrative data for infectious diseases research in Australia. *Public Health Res Pract*. 2018;28(2):e2821819.
15. Lim FJ, Blyth CC, Levy A, Fathima P, de Klerk N, Giele C, et al. Using record linkage to validate notification and laboratory data for a more accurate assessment of notifiable infectious diseases. *BMC Med Inform Decis Mak*. 2017;17:86.
16. Gibson A, Jorm L, McIntyre P. Using linked birth, notification, hospital and mortality data to examine false-positive meningococcal disease reporting and adjust disease incidence estimates for children in New South Wales, Australia. *Epidemiol Infect*. 2015;143(12):2570-9.
17. Clothier HJ, Vu T, Sundararajan V, Andrews RM, Counahan M, Tallis GF, et al. Invasive pneumococcal disease in Victoria: A better measurement of the true incidence. *Epidemiol Infect*. 2008;136(2):225-31.
18. 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.
19. McCallum LK, Liu B, McIntyre P, Jorm LR. Estimating the burden of pertussis in young children on hospitals and emergency departments: A study using linked routinely collected data. *Epidemiol Infect*. 2014;142(4):695-705.
20. Moore H, Burgner D, Carville K, Jacoby P, Richmond P, Lehmann D. Diverging trends for lower respiratory infections in non-Aboriginal and Aboriginal children. *J Paediatr Child Health*. 2007;43(6):451-7.
21. Gidding HF, Amin J, Dore GJ, Ward K, Law MG. Hospital-related morbidity in people notified with hepatitis C; A population-based record linkage study in New South Wales, Australia. *J Hepatol*. 2010;53(1):43-9.
22. Gidding HF, Dore GJ, Amin J, Law MG. Trends in all cause and viral liver disease-related hospitalizations in people with hepatitis B or C: A population-based linkage study. *BMC Public Health*. 2011;11(1):52.
23. Amin J, Law MG, Bartlett M, Kaldor JM, Dore GJ. Causes of death after diagnosis of hepatitis B or hepatitis C infection: A large community-based linkage study. *Lancet*. 2006;368(9539):938-45.
24. Gidding HF, McCallum L, Snelling T, Liu B, de Klerk N, Blyth C, et al. Probabilistic linkage of national immunisation and state-based health records for a cohort of 1.9 million births to evaluate Australia's childhood immunisation program. *Int J Popul Data Sci*. 2017;2(1):1-13.
25. Regan AK, Moore HC, de Klerk N, Omer SB, Shellam G, Mak DB, et al. Seasonal trivalent influenza vaccination during pregnancy and the incidence of stillbirth: Population-based Retrospective Cohort Study. *Clin Infect Dis*. 2016;62(10):1221-7.
26. Regan AK, de Klerk N, Moore HC, Omer SB, Shellam G, Effler PV. Effect of maternal influenza vaccination on hospitalization for respiratory infections in newborns: A Retrospective Cohort Study. *Pediatr Infect Dis J*. 2016;35(10):1097-103.
27. Fathima P, Blyth CC, Lehmann D, Lim FJ, Abdalla T, de Klerk N, et al. The Impact of pneumococcal vaccination on bacterial and viral pneumonia in Western Australian children: Record Linkage Cohort Study of 469589 births, 1996–2012. *Clin Infect Dis*. 2018;66(7):1075-85.
28. Rose N. Interview: data linkage in the setting of communicable disease control, 4 January 2018, personal communication.
29. Public Health Act (NSW) 2010.

30. New South Wales Health. *NSW Sexually Transmissible Infections Strategy 2016 – 2020: Data Report January to December 2016*. Sydney (AUST): State Government of New South Wales; 2017. [cited 2018 Jan 4]. Available from: http://www.health.nsw.gov.au/sexualhealth/Documents/nsw-sti-2016-annual-report.pdf
31. Roscoe T. *Health Alert: 160005 - Hepatitis B Lookback* [Internet]. Melbourne (AUST): Victorian Department of Health and Human Services; 2016 [cited 2018 Jan 4]. Available from: https://www2.health.vic.gov.au/about/news-and-events/healthalerts/hepatitis-b-lookback
32. Malo J, Davis S, Lambert SB. Using data linkage to identify newly acquired hepatitis C notifications in Queensland. *Proceedings of the Communicable Diseases Control Conference*; 2017 Jun 26 – 28; Melbourne, AUST. Canberra: Public Health Association of Australia; 2017.
33. Edwards L, Markey P, Cook H, Trauer J, Krause V. The relationship between influenza and invasive pneumococcal disease in the Northern Territory, 2005–2009. *Med J Aust*. 2011;194(4):207.
34. Lodo K. Causes and rates of hospitalisation among a cohort of Tasmanians notified with hepatitis C: a data linkage study 2014 , Chapter 6 [master's thesis]. Canberra, AU; Australian National University; 2014. 34p.
35. Regan AK, Moore HC, Sullivan SG, De Klerk N, Effler PV. Epidemiology of seasonal influenza infection in pregnant women and its impact on birth outcomes. *Epidemiol Infect*. 2017;145(14):2930-9.
36. Centre for Health Record Linkage. *Data Dictionaries* [Internet]. North Sydney (AUST): CHeReL; 2018 [cited 2018 Jan 4]. Available from: http://www.cherel.org.au/data-dictionaries
37. Public Health and Wellbeing Act (VIC) 2008, s 46 of 2008.
38. Privacy Act (Cth) 1988, s 119 (07/09/2017).
39. Health Records Act (Vic) 2001, s 023 of 2001.
40. Victorian Data Sharing Act 2017, s 60 of 2017.
41. Public Sector (Data Sharing) Act (SA) 2016, Stat. 61 (30 May 2017, 2016).
42. Data Sharing (Government Sector) Act 2015, Stat. 60 [NSW] (24 November 2015, 2015).
43. Pennington K, the Enhanced Invasive Pneumococcal Disease Surveillance Working Group. *Invasive Pneumococcal Disease Surveillance, 1 April to 30 June 2017*. Canberra (AUST): Australian Department of Health; 2017.
44. Australian Institute of Health and Welfare. *Report on the Use of Linked Data Relating to Aboriginal and Torres Strait Islander People*. Canberra (AUST): AIHW; 2013.
45. Australian Institute of Health and Welfare and Australian Bureau of Statistics. *National Best Practice Guidelines for Data Linkage Activities Relating to Aboriginal and Torres Strait Islander People*. Canbera (AUST): AIHW; 2012.
46. Hull BP, Lawrence GL, MacIntyre CR, McIntyre PB. Immunisation coverage in Australia corrected for under-reporting to the Australian Childhood Immunisation Register. *Aust N Z J Public Health*. 2003;27(5):533-8.
47. Australian Immunisation Register Act (Cth) 2015, s 138 (2015), Australia.
48. Moore HC, Guiver T, Woollacott A, de Klerk N, Gidding HF. Establishing a process for conducting cross-jurisdictional record linkage in Australia. *Aust NZ J Public Health*. 2016;40(2):159-64.
49. Andrew NE, Sundararajan V, Thrift AG, Kilkenny MF, Katzenellenbogen J, Flack F, et al. Addressing the challenges of cross-jurisdictional data linkage between a national clinical quality registry and government-held health data. *Aust N Z J Public Health*. 2016;40(5):436-42.
50. Mitchell RJ, Cameron CM, McClure RJ, Williamson AM. Data linkage capabilities in Australia: Practical issues identified by a Population Health Research Network 'Proof of Concept Project'. *Aust N Z J Public Health*. 2015;39(4):319-25.
51. Kelman CW, Bass AJ, Holman CDJ. Research use of linked health data — a best practice protocol. *Aust N Z J Public Health*. 2002;26(3):251-5.
52. Given J, Nelson E, Frances K, Dolk H, Gillian. Public attitudes to data linkage and sharing. *Int J Popul Data Sci*. 2017;1:299.
53. NatCen Social Research. *Public Attitudes to Data Linkage* [Report]. London (UK): University College; 2018.
54. Opus International Consultants. *Public Attitudes to Data Integration* [Report] [Internet]. Wellington (NZ): Statistics New Zealand; 2015 [cited 2018 Aug 24]. Available from: http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/keep-data-safe/public-attitudes-data-integration-2015.aspx
55. Guidelines Working Group. Updated Guidelines for Evaluating Public Health Surveillance System. *MMWR Recomm Rep*. 2001;50(RR-13):1-35; quiz CE1-7.
56. Nusbunga P, White ME, Thacker SB, Anderson MA, Blount SB, Broome CV, et al. Public Health Surveillance: A tool for targeting and monitoring interventions. In: Jamison DT, Breman JG, Measham AR, editors. *Disease Control Priorities in Developing Countries, Disease Control Priorities in Developing Countries*. 2 ed. Washington (DC): World Bank; 2006.
57. Australian Department of Health. *National Framework for Communicable Disease Control*. Canberra (AUST): Government of Australia; 2014.
58. National Institute for Health Research. *Theme C: Project 1: Mapping the CoC for BBV through Data Linkage* [Internet]. London (UK): University College London Research Department of Infection and Population Health Blood Borne & Sexually Transmitted Infections; 2018 [cited 2018 Aug 22]. Available from: http://bbsti.hpru.nihr.ac.uk/our-research/research-themes/theme-c-project-1-mapping-coc-bbv-through-data-linkage
59. Derrough T, Olsson K, Gianfredi V, Simondon F, Heijbel H, Danielsson N, et al. Immunisation Information Systems – useful tools for monitoring vaccination programmes in EU/EEA countries, 2016. *Euro Surveill*. 2017;22(17):30519.
60. Nohynek H, Baum U, Syrjänen R, Ikonen N, Sundman J, Jokinen J. Effectiveness of the live attenuated and the inactivated influenza vaccine in two-year-olds – a nationwide cohort study Finland, influenza season 2015/16. *Euro Surveill*. 2016;21(38):30346.
61. Kelman C, Smith L. It's time: record linkage — the vision and the reality. *Aust NZ J Public Health*. 2000;24(1):100-1.
62. Duckett S, Jorm C, Danks L. *Strengthening Safety Statistics: How to Make Hospital Safety Data More Useful*. Carlton (AUST): Grattan Institute; 2017.
63. Office of the Victorian Privacy Commissioner. *Data Matching in the Public Interest - A Guide for the Victorian Public Sector*. Melbourne (AUST): OVIC; 2009.
64. Muscatello D, Morton P, Evans I, Gilmour R. Prospective surveillance of excess mortality due to influenza in New South Wales: Feasibility and statistical approach. *Commun Dis Intell Q Rep*. 2008;32(4):435-42.