

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

Gidding, HF^{1,2*}, McCallum, L¹, Fathima, P³, Snelling, TL^{3,4,5,6}, Liu, B¹, de Klerk, N³, Blyth, CC^{3,4,7,8}, Sheppard, V⁹, Andrews, RM⁵, Jorm, L¹⁰, McIntyre, PB², and Moore, HC³ on behalf of the ACIR Investigator Team¹¹

Submission History

Submitted:	28/04/2017
Accepted:	17/08/2017
Published:	18/09/2017

¹School of Public Health and Community Medicine, UNSW Medicine, The University of New South Wales, Sydney, NSW, Australia

²National Centre for Immunisation Research and Surveillance, Westmead, NSW, Australia

³Wesfarmers Centre of Vaccines and Infectious Diseases, Telethon Kids Institute, The University of Western Australia, Perth, WA, Australia

⁴Department of Infectious Diseases, Princess Margaret Hospital, Perth, WA, Australia

⁵Menzies School of Health Research, Charles Darwin University, Darwin, Northern Territory, Australia

⁶School of Public Health, Curtin University, Perth, WA, Australia

⁷School of Medicine, University of Western Australia, Perth, WA, Australia

⁸Department of Microbiology, PathWest Laboratory Medicine WA, Princess Margaret Hospital, Perth, WA, Australia

⁹Communicable Diseases, Health Protection NSW, NSW Ministry of Health, NSW, Australia

¹⁰Centre for Big Data Research in Health, UNSW Medicine, The University of New South Wales, Sydney, NSW, Australia

Abstract

Introduction

Several countries have developed national immunisation registers, but only the Nordic countries have linked their registers to other health data in order to comprehensively evaluate the 'real world' effectiveness of vaccines. Nordic countries can link datasets deterministically using the national person identifier, but most countries, including Australia, don't have such an identifier to enable this type of linkage.

Objectives

To describe the process for assembling a linked study cohort that will enable the conduct of population-based studies related to immunisation and immunisation policy.

Methods

National death and immunisation databases along with state health data (notifications of vaccine preventable diseases, perinatal data, hospital admissions and emergency department presentations) up until December 2013 were probabilistically linked (using demographic details) for children born between 1996 and 2012 in two states: Western Australia and New South Wales (42% of Australia's population, combined).

Results

After exclusions there were 1.95 million children in the study cohort (live born children with both a birth and perinatal record which represents 97.5% of all live births in the state perinatal data collections - our source population) and 18.0 million person years of follow up (mean: 9.2 years per child). The characteristics of children in the cohort were generally similar to those only included in state perinatal databases and outcome measures were in keeping with expected figures from unlinked data sources. However, the lack of a dynamic national population register meant immigrants could not be included.

Conclusions

We have been able to develop a similarly comprehensive system to the Nordic countries based on probabilistic linkage methods. Our experience should provide encouragement to other countries with national immunisation registers looking to establish similar systems.

Introduction

The Australian Childhood Immunisation Register (ACIR) was the world's first purpose-built national immunisation register and is still one of only a limited number of national registers world-wide [1, 2]. The ACIR provides population-based estimates of coverage and timeliness of all vaccines publically funded through Australia's National Immunisation Program and includes all children registered on Australia's universal

public health scheme (Medicare). While the ACIR has played a pivotal role in measuring and improving vaccination coverage in Australia [3, 4], it contains only limited demographic information (child's age, sex, Aboriginal and Torres Strait Islander [hereafter respectfully referred to as Indigenous] status, and residential postcode). Thus identification of specific populations with low coverage and factors associated with poor uptake is limited. In addition, unlinked ACIR data can only be used to estimate vaccine effectiveness using ecological study

¹¹The ACIR Investigator Team also includes: Richmond P, Effler P, Menzies R, Hull B, Joseph T, Edmond K, Corben P, and Sheridan S

*Corresponding Author:

Email Address: hgidding@unsw.edu.au (H.F. Gidding)

designs, which lack control for potentially important clinical and demographic confounders of vaccine effects such as geographic location, gestational age, ethnicity and comorbid medical conditions [5-8].

Linkage of individual-level data from population-based immunisation registers to other routinely collected health and demographic data is a more robust approach for post-licensure evaluation of immunisation programs than using unlinked data [9]. The size and representative nature of such assembled cohorts provide the power and diversity to characterise specific sub-populations not previously identifiable. Linkage with other databases to create an historical cohort also allows for direct calculation of disease rates by vaccination status, while controlling for potential confounders, to accurately measure vaccine effectiveness. However, few countries have undertaken comprehensive linkage of their immunisation register to other health databases.

Australia is a federation of six states and two territories, with most administrative health databases maintained by the health departments within each state or territory. However, Medicare-related data, which includes immunisation records from the ACIR, fall under the jurisdiction of the Australian (Commonwealth) Government. Establishment of data access guidelines in mid-2014 facilitated development of a framework for conducting probabilistic linkage between Commonwealth data and state health databases [10]. This paper aims to describe the assembly of a complex multistate-Commonwealth linked data system that will enable the conduct of population-based studies related to immunisation and immunisation policy, and can be used as a reference for studies that might make use of this linked data resource.

Methods

Study setting

This study integrated health data from two states with established data linkage capacity (New South Wales [NSW] and Western Australia [WA]), with the ACIR and national death data. NSW is the most populous state (7.6 million), while WA (2.6 million) has a higher proportion of the population residing in remote regions (0.6% v 7%). Together they represent 42% of Australia's population with an approximate annual birth cohort of 125,000 which is comparable to that of Belgium or Sweden [11].

Study design

We constructed a retrospective cohort for children born between 1996 and 2012 with linked health and vaccination data up until December 2013 (Figure 1) in order to construct a data system that could be used for studies to evaluate coverage, effectiveness and timeliness of childhood vaccinations, at the population level and in specific subgroups that could not previously be identified.

Description of data sources used for linkage

Perinatal and birth records: These data were from state perinatal data collections and birth registries. Perinatal data included demographic, maternal medical and obstetric history,

and information on the labour and delivery of all births, and was recorded by the attending midwife or medical practitioner. The birth registers included demographic details on both parents and the baby (including full name) as recorded by the parents.

Vaccinations: The ACIR included details of vaccinations (type, brand and date of administration) given by recognised immunisation providers to children <7 years of age residing in Australia who were enrolled in Medicare (99% of all children by 12 months of age). Medical contraindications and conscientious objection were also recorded.

Death data: Death data (from the National Death Index) included demographic information, date and cause of death (including contributing causes from 1997) on all deaths in Australia. Causes were coded using the International Classification of Diseases (ICD) coding system.

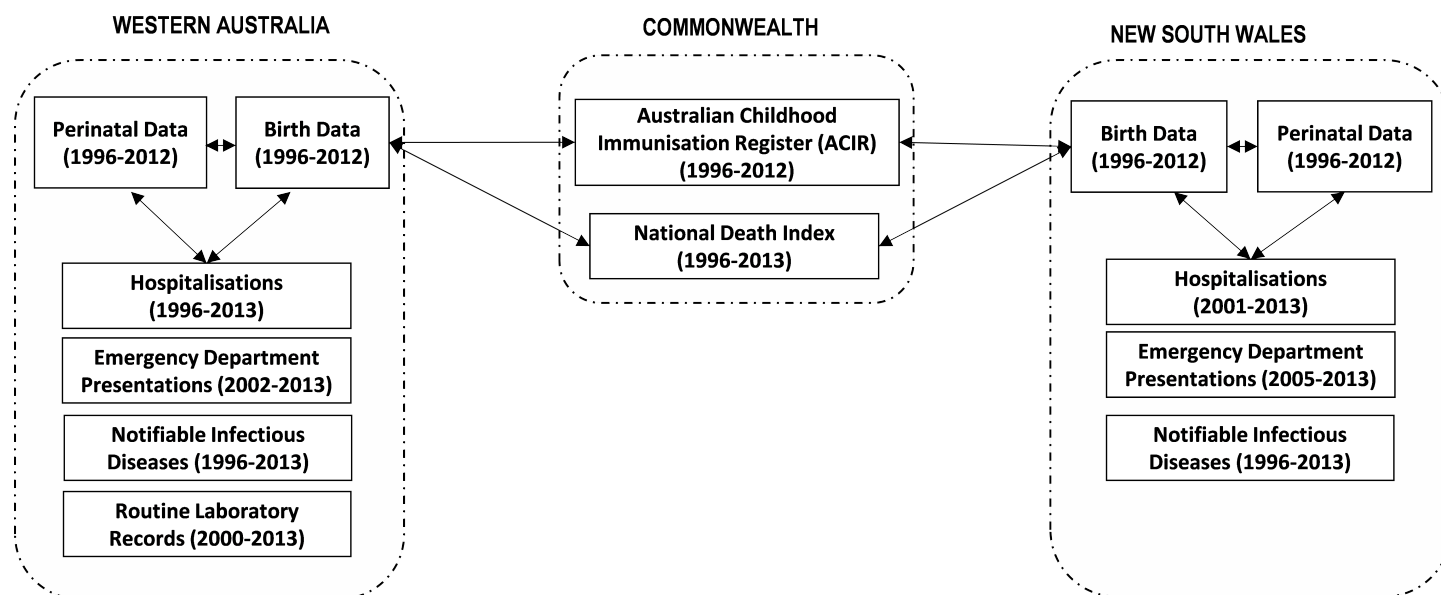
Vaccine-preventable disease diagnoses (notifications) and testing: State-based public health legislation makes it mandatory for clinicians and laboratories to notify any new vaccine-preventable disease diagnoses. Recorded information included: disease, date of onset, diagnosis method and organism serotype (where appropriate). Data on routine microbiology tests performed in all government-funded laboratory services in WA from 2000 were also linked to the cohort. Data included date, time, specimen type, test and result (listing any organisms identified).

Morbidity: Hospital admissions and emergency department presentations can provide information about the severity of vaccine preventable disease infections and presence of comorbid conditions. Hospitalisation data covered all inpatient separations (discharges, transfers and deaths) in each state. Data included dates of admission and separation, the primary diagnosis code (first-listed diagnosis), up to 50 (NSW) or 20 (WA) secondary diagnosis codes (coded using the ICD coding system), and codes for procedures performed during the stay. Emergency department data contained date and symptoms (coded using ICD codes, Systematised Nomenclature of Medicine Clinical Terms [SNOMED CT] or free text) on presentation to public hospital emergency departments for the majority of metropolitan hospitals in NSW (from 2005) and all hospitals in WA (from 2002).

Data linkage and linkage quality

As there is no unique personal identifier or dynamic national population register for the entire Australian population, probabilistic linkage techniques based on full name, date of birth, residential address and sex were used to link the study databases (Figure 1). Linkage of the state databases occurred at the NSW Centre for Health Record Linkage (CHeReL) and WA Data Linkage Branch (WADLB), while linkage of the identifiers from the state birth registries to the death data and ACIR occurred at the Australian Institute of Health and Welfare (AIHW). State birth registration records were chosen for linkages conducted at AIHW because they included full name of the baby, unlike the perinatal records (which have the full name of the mother but usually not of the baby). Additional

Figure 1: Data sources, years available, and their relationship through linkage*



* For more detail regarding the linkage process, refer to Moore et al [10]

details regarding the linkage process are described in Moore et al [10].

Following the completion of probabilistic linkage, clerical reviews were conducted at each data linkage unit [12, 13]. The state data linkage units then determined which records were a match and only provided those records to the researchers. For the linkages performed at AIHW, the researchers determined which links would be accepted as 'true' links based on the linkage weight cut-off threshold corresponding to the most appropriate matched linkage rate (presumed true matches that are accepted links; similar to a sensitivity except that it is based on the judgement of the clerical reviewer rather than the actual 'truth') and linkage accuracy (how many of the accepted links are presumed to be correct links; similar to a positive predictive value). For linkage of the birth registry to ACIR we chose a matched linkage rate of 99.3% which had a corresponding linkage accuracy of 99.0%. For linkage between the birth registry and death data we chose a matched linkage rate of 99.2% which had a corresponding linkage accuracy of 96.6%.

Assembly of the study cohort

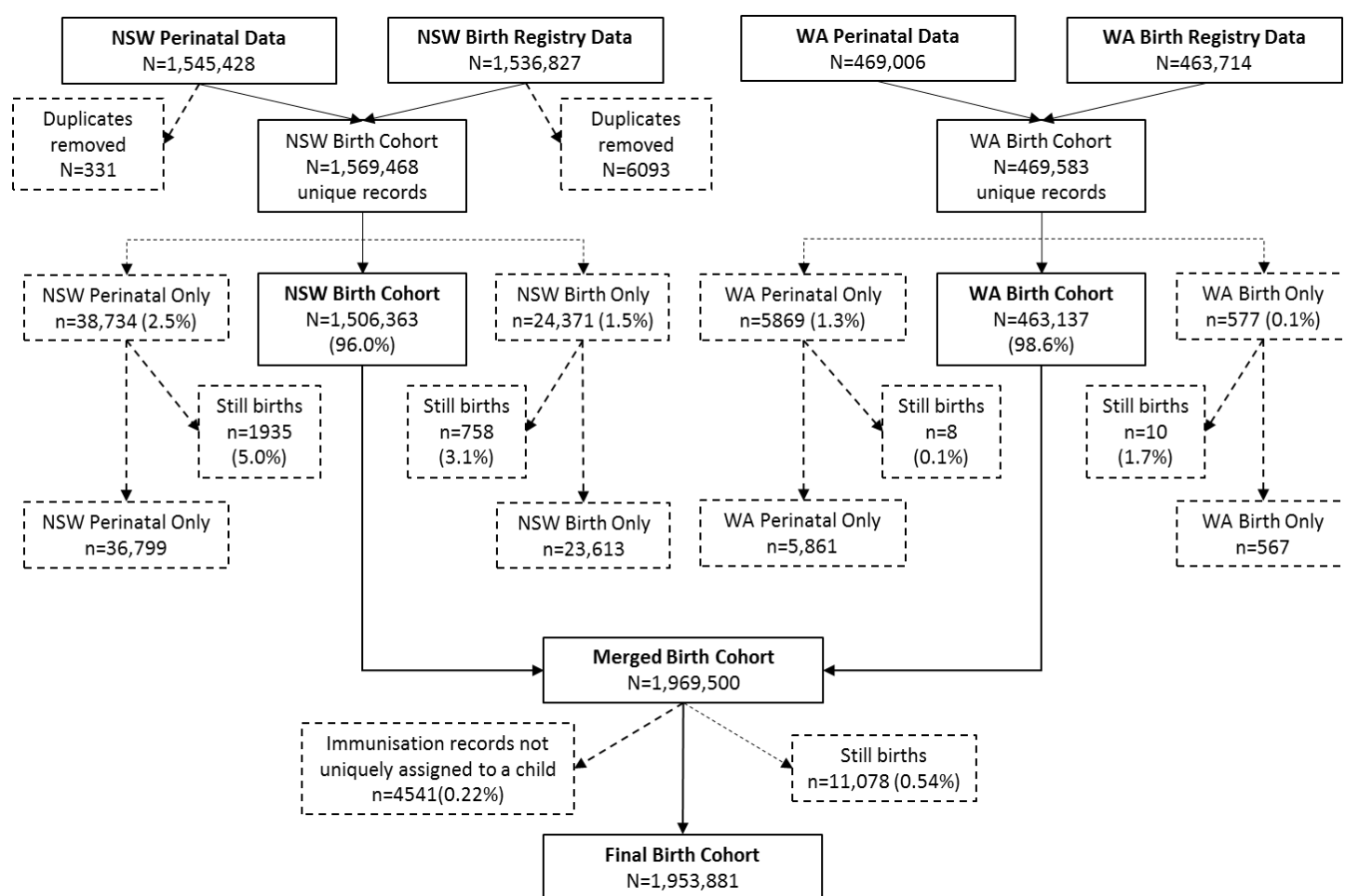
Our source population was all live-births in NSW and WA between 1996 and 2012. Figure 2 outlines the cleaning process and exclusions required to obtain our study cohort. First, exact and non-exact (where the value for at least one variable didn't match between records with the same child identifier) duplicates were removed from the NSW birth registry and perinatal databases (there were no duplicates in the corresponding WA databases) to obtain one record per child in each database. For non-exact birth registry duplicates ($n=3,804$; 0.25% of birth records), the record with the most recent date of registration was kept, or if these dates were the same ($n=66$), then one record was randomly chosen. For non-exact perinatal data duplicates ($n=331$; 0.02% of the perinatal records) the record

that matched the birth registry data was chosen, but for 123 children (246 records) the correct record could not be determined so we excluded these children from the cohort. Of the remaining children, only those with both a perinatal and birth registry record (97.5% of live births in the perinatal data collections) were included in the study cohort. A further 0.22% of the study cohort who did not have a unique person identification number (PIN) on the ACIR were excluded. Finally, stillbirths were excluded based on the date of death ($n=10,946$, plus an additional 124 children with a death record where the date of death was before the date of birth), or where no date of death was provided but there was a birth registry ($n=6$) or perinatal record ($n=2$) indicating the child was stillborn or a neonatal/perinatal death.

Linkage of vaccination and birth registry data

Every child enrolled in ACIR is given a unique PIN which enables deterministic linkage to their 'set' of vaccination records. Therefore, for our study, probabilistic linkage to the birth registry records was performed at the child (PIN) level. There were 26,973,736 linkages accepted as 'true' links after restricting the ACIR records to those from PINs with linkage weights above the cut off threshold to achieve the estimated linkage quality described above. Further exclusions were then made to ensure a unique 'set' of vaccination records per birth registration. For 4,766 birth registrations (0.25% of the children with at least one ACIR record) the PIN matched to more than one child on the birth registry. These children and their 115,658 ACIR records were flagged for subsequent removal of those remaining in the assembled study cohort as described above and in Figure 2. There remained 21,179 children (1.1% of the children with at least one ACIR record) in the birth registry with more than one PIN for whom we either chose the PIN with the closest match on date of birth to the cohort ($n=12,984$), or else the PIN with the highest linkage weight if the date of

Figure 2: Flowchart of the assembly of the study birth cohort showing exclusions



Notes: NSW New South Wales; WA Western Australia. Births=birth registrations from the Registry of Births Deaths and Marriages; Perinatal data=Perinatal Data Collection (NSW) or Midwives Notification System (WA)

birth was identical ($n=6,421$), or else the PIN and its set of records were randomly assigned if both the date of birth and linkage weight were identical ($n=1,774$). This process of assigning only one PIN per child resulted in 314,358 vaccination records being excluded. An additional 18,411 duplicate ACIR records were also removed. When the cleaned ACIR database of 26,525,309 records was linked to the study cohort (children with both a perinatal and birth registration record, see Figure 2), 26,247,927 records linked to a cohort member, with 95.5% of the cohort children having at least one vaccination record.

Censoring person time

The sensitivity of the National Death Index is estimated to be between 89% and 95% [14–16]. Therefore, we developed an algorithm based on all of the linked databases to enhance the sensitivity of ‘death’ ascertainment. This was done to reduce the risk of including person time for children with the potential to appear both healthy and unvaccinated. Linked death records from the National Death Index accounted for 93% of all ascertained deaths with the remaining 7% identified based

on an indication in their emergency, hospital, perinatal or ACIR records that the child had died.

Other study variables of interest

The linkage of multiple databases provided information on a number of maternal and infant characteristics (Table 1). We used the demographic information on the perinatal database except where the value was unknown, missing or indeterminate (e.g. for sex), when we used the birth registration values. Indigenous status was derived using three algorithms proposed by Christensen et al [17] (for always Indigenous, ever Indigenous, and using a multi-stage median algorithm [our base case]), which provides an application of the theory outlined in ‘National Best Practice Guidelines for Data Linkage Activities Relating to Aboriginal and Torres Strait Islander People’ [18].

Ethical approval

Ethical approval was obtained from: The Aboriginal Health and Medical Research Council Ethics Committee (approval

ID: 931/13), AIHW Ethics Committee (approval ID: EC 2012/4/62), Department of Health Human Research Ethics Committee (approval ID: 1/2013), Department of Health - WA Human Research Ethics Committee (approval ID: 2012/75), NSW Population and Health Service Research Ethics Committee (approval ID: HREC/13/CIPHS/15), and Western Australian Aboriginal Health Ethics Committee (approval ID: 459).

Results

There were 1,953,881 children assembled in the study cohort (Figure 2). Of these children, 7,962 (0.4%) were recorded as having died during the study follow-up period (January 1996 until December 2013). Accounting for deaths, the study included a total of 18.0 million person-years of follow up (mean: 9.2 years per child). The characteristics of the study cohort were generally similar to those for all live-born children included in the state perinatal databases (our source population; Table 1). However, there were some epidemiologically important differences between the study cohort and those children who did not have a birth registration record: compared to the children with only a perinatal record, our cohort included fewer Indigenous children, younger mothers, mothers who smoked during pregnancy, and mothers living in remote regions or with a low Socio-Economic Indexes for Areas (SEIFA) score.

Unlinked health and vaccination records that were expected to link to the study cohort were not available. Therefore, to examine the quality of the linkage processes we compared several outcomes obtained from our linked data to key indicators that were available from other sources. We found that vaccination coverage, mortality, and disease notification rates were broadly comparable (see Appendix 1 for details). Slight differences were anticipated given the different calculation methods and that linked data is expected to provide a more accurate estimate of the outcome measures.

Discussion

Of the 14 countries we were able to identify with national immunisation registers, only the Nordic countries have published studies using comprehensive linkage systems encompassing a national immunisation register, demographic, birth, maternal, hospital, and vaccine preventable disease notification data. We have been able to develop a similarly comprehensive system, but unlike other countries undertaking linkage of their national register, our system relies on probabilistic linkage of demographic variables rather than deterministic linkage based on a unique person identifier. Whilst a unique person-level identifier is most efficient for linkage between databases, most countries, including Australia, do not have such an identifier that is present on all of the relevant databases. Nevertheless, we have been able to achieve a high level of linkage accuracy using probabilistic linkage techniques demonstrating that this method could be suitable for other countries that do not have a universal person identifier.

There are three interrelated limitations with our system. First, in contrast to the Nordic countries, Australia does not have a dynamic national population register. Therefore, our

study cohort was static (birth records in NSW and WA) and we were not able to include children who immigrated after birth or identify children in our cohort who migrated out of NSW or WA. In 2011 (most recent census year available), 1.2% of all residents in NSW or WA migrated out of these states to another part of Australia [19] and 0.8% migrated overseas [20]. Thus, some children in the cohort will be missing data on their vaccination and health outcomes. The unobserved loss to follow-up will accumulate with increased person time of observation. However, it is anticipated that most planned analyses will focus on comparing disease rates in vaccinated and unvaccinated infants less than 5 years of age and thus the cumulative impact is expected to be less than 10%. More importantly, we do not expect loss to follow-up to differ between vaccinated and unvaccinated children, thus relative comparisons should be unbiased. With respect to immigration, in the latest Australian census (2011), 33% and 28% of NSW and WA residents, respectively, were recorded as having been born overseas [20]. Therefore, our study cohort may not be generalisable to immigrant children. Second, we were unable to capture hospital encounters for our cohort that occurred outside of each state. However, a recent study linking hospitalisations across four states (NSW, WA, South Australia and Queensland) found that only 2.75% of hospitalisations in NSW residents and 0.2% of hospitalisations in WA residents occurred in the other states [21]. The third limitation is that because perinatal records often don't include the full name of the baby, only children with a birth registry record could be linked to the national databases and therefore included in our cohort. This led to certain differences between the perinatal data and our study cohort (Table 1). Similar differences have been reported in other studies linking the state birth registrations and perinatal data collections in WA (for Indigenous births only), NSW and Queensland [22, 23, 24] and should be kept in mind when generalising the data to all births. However, the study is representative of all registered births and includes a large cohort involving 97.5% of all live births (89.8% of all Indigenous births) in the NSW and WA perinatal data collection. In addition, comparisons between groups in the cohort should be valid [25].

Conclusion

In summary, Australia now has a comprehensive population-based system for evaluation of the childhood immunisation program. Despite the lack of a unique personal identifier or dynamic national population register, we were able to achieve a high level of linkage accuracy and believe our cohort is generalisable to all registered births in Australia. Our experience should provide encouragement to other countries with national immunisation registers looking to establish similar systems. This would enable robust post-implementation evaluation of country-specific vaccination schedules and international comparisons to inform optimisation of immunisation policies.

In 2016, the ACIR was expanded to record vaccinations given to people of all ages (the Australian Immunisation Register; AIR). Researcher access to linked AIR and Medicare data would not only enable evaluation of whole-of-life immunisation policies but also provide information about population mobility, and thus a more dynamic study cohort. We also plan to

Table 1: Demographic characteristics of the study cohort and comparison with the all live births in the Perinatal Data Collection (representing our source population) and live births with unlinked perinatal records

Demographic characteristics	Study Cohort N=1,953,881		Perinatal record with no birth record N=42,660		All perinatal records N= 1,996,541		
	n	%	n	% ^a	n	%	
<i>Parental characteristics</i>							
<i>Maternal age (years)</i>							
≥35	391,237	20.0	5,912	13.9	397,149	19.9	
30-34	627,111	32.1	9,857	23.1	636,968	31.9	
25-29	565,215	28.9	12,069	28.3	577,284	28.9	
20-24	289,110	14.8	10,652	25.0	299,762	15.0	
<20	81,208	4.2	4,124	9.7	85,332	4.3	
missing	-	-	46	0.1	46	0.0	
<i>Paternal age (years)^b</i>							
≥35	654,098	33.5					
30-34	588,710	30.1					
25-29	399,465	20.4					
20-24	153,285	7.9					
<20	26,254	1.3					
missing	132,069	6.8					
<i>Maternal country of birth^b</i>							
Australia	1,277,984	65.4					
Not Australia	530,553	27.2					
Not stated	145,344	7.4					
<i>Plurality</i>							
Singleton	1,897,476	97.1	40,701	95.4	1,938,177	97.1	
Multiple	56,405	2.9	1,959	4.6	58,364	2.9	
<i>Gestational/delivery characteristics</i>							
<i>Maternal smoking during pregnancy</i>							
Yes	284,458	14.6	17,342	40.7	301,800	15.1	
No	1,665,836	85.3	25,230	59.1	1,691,066	84.7	
Not stated	3,587	0.2	88	0.2	3,675	0.2	
<i>Gestational age</i>							
≥37 weeks	1,817,917	93.0	37,769	88.5	1,855,686	92.9	
32-36 weeks	117,491	6.0	3,976	9.3	121,467	6.1	
< 32 weeks	18,151	0.9	875	2.1	19,026	1.0	
missing	322	0.0	40	0.1	362	0.0	
<i>Mode of delivery</i>							
Vaginal	1,190,119	60.9	29,421	69	1,219,540	61.1	
Instrumentation	225,278	11.5	2,979	7	228,257	11.4	
Caesarean	537,908	27.5	10,225	24	548,133	27.5	
Not stated	576	0.0	35	0.1	611	0.0	

Demographic characteristics	Study Cohort N=1,953,881		Perinatal record with no birth record N=42,660		All perinatal records N= 1,996,541		
	n	%	n	% ^a	n	%	
<i>Baby characteristics</i>							
State of birth							
	WA	461,577	23.6	5,861	13.7	467,438	23.4
	NSW	1,492,304	76.4	36,799	86.3	1,529,103	76.6
Gender							
	Male	1,003,815	51.4	21,458	50.3	1,025,273	51.4
	Female	950,066	48.6	21,146	49.6	971,212	48.6
	Missing/indeterminate			56	0.1		
Indigenous status							
Base case ^c							
	Indigenous	97,789	5.0	10,683	25.0	108,472	5.4
	Non-Indigenous	1,835,206	93.9	21,099	49.5	1,856,305	93.0
	Missing	20,886	1.1	10,878	25.5	31,764	1.6
Always Indigenous							
	Indigenous	46,450	2.4	9,031	21.2	55,481	2.8
	Non-Indigenous	1,886,545	96.6	22,751	53.3	1,909,296	95.6
	Missing	20,886	1.1	10,878	25.5	31,764	1.6
Ever Indigenous							
	Indigenous	135,287	6.9	10,942	25.6	146,229	7.3
	Non-Indigenous	1,797,708	92.0	20,840	48.9	1,818,548	91.1
	Missing	20,886	1.1	10,878	25.5	31,764	1.6
Season of birth							
	Summer	470,367	24.1	14,191	33.3	484,558	24.3
	Autumn	497,046	25.4	9,364	22.0	506,410	25.4
	Winter	493,409	25.3	8,981	21.1	502,390	25.2
	Spring	493,059	25.2	10,124	23.7	503,183	25.2
Birthweight							
	<1500g	15,516	0.8	761	1.8	16,277	0.8
	1500-2499g	95,193	4.9	3,902	9.2	99,095	5.0
	2500-3499g	1,004,462	51.4	23,806	55.8	1,028,268	51.5
	3500-4499	804,583	41.2	13,562	31.8	818,145	41.0
	≥4500g	33,654	1.7	578	1.4	34,232	1.7
	missing	473	0.0	51	0.1	524	0.0



Demographic characteristics	Study Cohort N=1,953,881		Perinatal record with no birth record N=42,660		All perinatal records N= 1,996,541	
	n	%	n	% ^a	n	%
<i>Geographical characteristics</i>						
SEIFA ^d						
91-100%	171,068	8.8	1,665	3.9	172,733	8.7
76-90%	284,838	14.6	3,129	7.3	287,967	14.4
26-75%	928,020	47.5	16,279	38.2	944,299	47.3
11-25%	303,262	15.5	8,740	20.5	312,002	15.6
0-10%	210,597	10.8	10,870	25.5	221,467	11.1
Missing	56,096	2.9	1,977	4.6	58,073	2.9
ARIA ^e						
Major cities	1,443,393	73.9	23,580	55.3	1,466,973	73.5
Inner regional	300,858	15.4	10,117	23.7	310,975	15.6
Outer regional	121,626	6.2	4,722	11.1	126,348	6.3
Remote	27,000	1.4	1,347	3.2	28,347	1.4
Very remote	8,997	0.5	1,170	2.7	10,167	0.5
Missing	52,007	2.7	1,724	4.0	53,731	2.7

^a Distributions of all variables were significantly different ($p < 0.001$) between the linked (cohort) and unlinked perinatal records using the chi-square test

^b Paternal age was only available in the birth registry and Maternal country of birth was obtained from multiple datasets, so these variables are not able to be compared to unlinked and all perinatal records

^c Based on the multi-stage median algorithm by Christensen et al [17]

^d SEIFA=Socio-Economic Indexes for Areas; Based on state specific scores where 0-10% represents those most disadvantaged and 91-100% represents those least disadvantaged

^e ARIA=Accessibility/Remoteness Index of Australia



extend the linkages described here to include health data from additional jurisdictions, maternal vaccination status (now being recorded on the state perinatal data collections), and to investigate ways to incorporate individual-level socioeconomic and education data (such as the Australian Early Development Census) to be able to measure the health and economic benefits of vaccination more fully.

Access to the study data

The ACIR linkage Investigator team welcome contact from researchers who would like to propose a collaborative project using the data we have assembled. Please contact the lead investigator (H Gidding). Additional ethical and data custodian approvals may be required depending on the project. For confidentiality reasons all analyses must be conducted in the Secure Unified Research Environment (SURE), a secure, remote access facility where the data are stored.

Acknowledgements

This project was funded by the Population Health Research Network (PHRN), a capability of the Commonwealth Government National Collaborative Research Infrastructure Strategy and Education Investment Fund Super Science Initiative, and a National Health and Medical Research Council (NHMRC) project grant (APP1082342). The authors are grateful to the staff at the Population Health Research Network (PHRN) and participating PHRN data linkage and infrastructure nodes (the Western Australian Data Linkage Branch, the NSW Centre for Health Record Linkage, and the Australian Institute for Health and Welfare), and the WA and Commonwealth Departments of Health and NSW Ministry of Health who provided advice and the data. Thank you to Arto Palmu for his helpful comments on the manuscript and to Han Wang and Sarah Sheridan for assistance assembling the comparative data. The Aboriginal and Torres Strait Islander community and members of the Aboriginal Immunisation Reference Group are acknowledged for their contribution to this research project. HM, BL, TS and CB and HG are supported by NHMRC Fellowships.

Statement on conflicts of interest

None declared.

References

- Chin LK, Crawford NW, Rowles G, BATTERY JP. Australian immunisation registers: established foundations and opportunities for improvement. *Euro surveillance*. 2012;17(16).
- Hull BP, Deeks SL, McIntyre PB. The Australian Childhood Immunisation Register - A model for universal immunisation registers? *Vaccine*. 2009;27(37):5054-60.
- Hull BP, McIntyre PB. Timeliness of childhood immunisation in Australia. *Vaccine*. 2006;24(20):4403-8.
- Hull BP, McIntyre PB, Sayer GP. Factors associated with low uptake of measles and pertussis vaccines - an ecologic study based on the Australian Childhood Immunisation Register. *Australian and New Zealand Journal of Public Health*. 2001;25(5):405-10.
- Blyth CC, Jacoby P, Effler PV, Kelly H, Smith DW, Borland ML, et al. Influenza Vaccine Effectiveness and Uptake in Children at Risk of Severe Disease. *The Pediatric Infectious Disease Journal*. 2016;35(3):309-15.
- Snelling TL, Andrews RM, Kirkwood CD, Culvenor S, Carapetis JR. Case-control evaluation of the effectiveness of the G1P[8] human rotavirus vaccine during an outbreak of rotavirus G2P[4] infection in central Australia. *Clinical Infectious Diseases*. 2011;52(2):191-9.
- Szilagyi PG, Fairbrother G, Griffin MR, et al. Influenza vaccine effectiveness among children 6 to 59 months of age during 2 influenza seasons: A case-cohort study. *Archives of Pediatrics & Adolescent Medicine*. 2008;162(10):943-51.
- World Health Organization. Generic protocol for monitoring impact of rotavirus vaccination on gastroenteritis disease burden and viral strains. Geneva: World Health Organization, 2008 WHO/IVB/08.16.
- Hviid A. Postlicensure epidemiology of childhood vaccination: the Danish experience. *Expert Review of Vaccines*. 2006;5(5):641-9.
- Moore HC, Guiver T, Woollacott A, de Klerk N, Gidding HF. Establishing a process for conducting cross-jurisdictional record linkage in Australia. *Australian and New Zealand Journal of Public Health*. 2015;40(2):159-64.
- Australian Bureau of Statistics. Births, Australia, 2014: cat. no. 3301.0, ABS, Canberra; 2015 [10/01/2017]. Available from: <http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3301.02014?OpenDocument>.
- Centre for Health Record Linkage. Quality Assurance. Available from: <http://www.cherel.org.au/quality-assurance>.
- Eitelhuber T. Data linkage - making the right connections: Department of Health, Government of Western Australia; 2016. Available from: <http://www.datalinkage-wa.org.au/about-us/linkage-quality>.
- Kelman C. The Australian national death index: an assessment of accuracy. *Australian and New Zealand Journal of Public Health*. 2000;24(2):201-3.
- Magliano D, Liew D, Pater H, Kirby A, Hunt D, Simes J, et al. Accuracy of the Australian National Death Index: comparison with adjudicated fatal outcomes among Australian participants in the Long-term Intervention with Pravastatin in Ischaemic Disease (LIPID) study. *Australian and New Zealand Journal of Public Health*. 2003;27(6):649-53.
- Powers J, Ball J, Adamson L, Dobson A. Effectiveness of the National Death Index for establishing the vital status of older women in the Australian Longitudinal Study on Women's Health. *Australian and New Zealand Journal of Public Health*. 2000;24(5):526-8.
- Christensen D, Davis G, Draper G, Mitrou F, McKeown S, Lawrence D, et al. Evidence for the use of an algorithm in resolving inconsistent and missing Indigenous status in administrative data collections. *Australian Journal of Social Issues*. 2014;49(4):423.
- 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 Canberra: AIHW; 2012. Available

from: <http://www.aihw.gov.au/publication-detail/?id=10737422216>.

19. Australian Bureau of Statistics. Australian Demographic Statistics, Jun 2014: cat no. 3101.0. ABS, Canberra; 2014. Available from: <http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3101.0Jun%202014?OpenDocument>.

20. Australian Bureau of Statistics. Migration, Australia, 2014-15: cat. no. 3412.0, ABS, Canberra; 2016. Available from: <http://www.abs.gov.au/ausstats/abs@.nsf/mf/3412.0/>.

21. Spilisbury K, Rosman D, Alan J, Boyd JH, Ferrante AM, Semmens JB. Cross border hospital use: analysis using data linkage across four Australian states. *Medical Journal of Australia*. 2015;202(11):582-6.

22. Gibberd AJ, Simpson JM, Eades SJ. No official identity: a data linkage study of birth registration of Aboriginal children in Western Australia. *Australian and New Zealand Journal of Public Health*. 2016;40(4):388-94.

23. Xu F, Sullivan EA, Black DA, Jackson Pulver LR, Madden RC. Under-reporting of birth registrations in New South Wales, Australia. *BMC Pregnancy and Childbirth*. 2012;12:147.

24. Queensland Health. An estimate of the extent of under-registration of births in Queensland. Brisbane; 2014. Available from: <http://www.health.qld.gov.au/hsu/peri/underreg.pdf>.

25. Mealing NM, Banks E, Jorm LR, Steel DG, Clements MS, Rogers KD. Investigation of relative risk estimates from studies of the same population with contrasting response rates and designs. *BMC Medical Research Methodology*. 2010;10(1):26.

rates are presented for infants <5 years old (Table A2). Comparison of notification rates for invasive pneumococcal disease and pertussis, and vaccination coverage with the 3rd dose of pneumococcal conjugate vaccine are provided for illustration.

Abbreviations

ACIR	Australian Childhood Immunisation Register
AIHW	Australian Institute of Health and Welfare
AIR	Australian Immunisation Register
ARIA	Accessibility/Remoteness Index of Australia
CHeReL	Centre for Health Record Linkage
ICD	International Classification of Diseases
NSW	New South Wales
PIN	Person Identification Number
SEIFA	Socio- Economic Indexes for Areas score
SNOMED CT	Systematised Nomenclature of Medicine Clinical Terms
WA	Western Australia
WADLB	WA Data Linkage Branch

Appendix 1: Comparison of selected outcome data in assembled cohort to other data sources

To demonstrate the quality of the linked data, key indicators were compared with other available sources. Mortality rate comparisons (Table A1) are confined to children aged 1-4 years, as direct comparison with reported mortality in <1 year olds was not possible (due to differing methodologies), and 1-4 year olds represent the majority of person-time available in our cohort. For similar reasons (and because they have the highest disease rates) comparisons to unlinked disease notification



Table 1A: Mortality rates (per 100,000) for children 1-4 years of age by data source, state of birth and year

State of birth	Year ^a	ABS ^b	Linked data
New South Wales	2003	25	24
	2004	26	22
	2005	26	22
	2006	24	22
	2007	22	20
	2008	20	22
	2009	19	20
Western Australia	2003	31	30
	2004	27	26
	2005	28	23
	2006	25	22
	2007	23	28
	2008	20	23
	2009	20	29

Notes: ^a Year=Year of death (linked data), year of registration (unlinked data).

^b ABS=Australian Bureau of Statistics

The numerator for both rates is deaths included in the National Death Index. However, the denominator for rates=person time (linked data), estimated mid-year resident population (unlinked data). See <http://stat.data.abs.gov.au/> for more details.



Table A2: Vaccine preventable disease notification rates (per 100,000) for children 0-4 years of age by data source, state of birth and year

Disease	State of birth	Year of onset	NNDSS ^a	Linked data
Pertussis				
	New South Wales	2000	63	57
		2001	88	77
		2002	47	41
		2003	51	46
		2004	56	49
		2005	60	55
		2006	41	39
		2007	38	34
		2008	263	243
		2009	602	572
		2010	291	267
		2011	507	466
		2012	248	226
	Western Australia	2013	91	81
		2000	16	18
		2001	39	35
		2002	49	49
		2003	41	38
		2004	192	190
		2005	66	63
		2006	22	21
		2007	5	5
		2008	48	43
		2009	72	73
	2010	111	108	
	2011	319	305	
	2012	293	282	
	2013	91	76	



Disease	State of birth	Year of onset	NNDSS ^a	Linked data
Invasive pneumococcal disease				
	New South Wales	2005	32	32
		2006	14	11
		2007	19	16
		2008	21	20
		2009	16	15
		2010	20	18
		2011	15	13
		2012	14	13
	Western Australia	2013	12	10
		2005	17	19
		2006	14	11
		2007	21	18
		2008	21	17
		2009	20	16
		2010	30	26
		2011	35	27
		2012	12	9
		2013	16	14

Notes: ^a NNDSS=National Notifiable Diseases Surveillance System (unlinked disease notification rates)

The numerator for both rates is the number of disease-specific notifications in each state. However, unlinked disease notifications may include repeat notifications for the same person whereas the rates derived from the linked data are based only on the first recorded notification for each person.

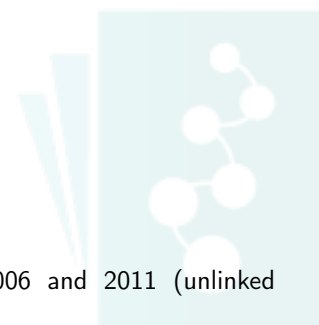
The denominator for rates=person time (linked data), estimated mid-year resident population (unlinked data).

Table A3: Percentage of children immunised by 12 months of age with the 3rd dose of pneumococcal conjugate vaccine by birth cohort, data source, and state of birth

State of birth	Year of birth	ACIR ^a	Linked data
New South Wales	2006	91	92
	2007	92	92
	2008	92	93
	2009	92	92
	2010	91	92
	2011	91	92
Western Australia	2006	89	89
	2007	88	90
	2008	89	90
	2009	90	90
	2010	89	90
	2011	89	90

Notes: ^a ACIR=Australian Immunisation Register (unlinked immunisation data).

The denominator is number of children registered on Medicare with a date of birth between 2006 and 2011 (unlinked immunisation data), children included in the study cohort (linked data).





Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Gidding, HF; McCallum, L; Fathima, P; Snelling, TL; Liu, B; de Klerk, N; Blyth, CC; Sheppard, V; Andrews, RM; Jorm, L; McIntyre, PB; Moore, HC

Title:

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.

Date:

2017-09-18

Citation:

Gidding, H. F., McCallum, L., Fathima, P., Snelling, T. L., Liu, B., de Klerk, N., Blyth, C. C., Sheppard, V., Andrews, R. M., Jorm, L., McIntyre, P. B. & Moore, H. C. (2017). 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*, 2 (1), pp.406-. <https://doi.org/10.23889/ijpds.v2i1.406>.

Persistent Link:

<http://hdl.handle.net/11343/271786>

File Description:

Published version

License:

CC BY-NC-ND