Validating linked hospital and laboratory data to investigate the aetiology of respiratory infections

Faye J Lim^a, Christopher Blyth^{a,b,c,d}, Nicholas de Klerk^a, Beverly Valenti^a, Oliver J Rouhiainen^b, Dominic Yu-An Wu^b, Christopher Jansz^b, Hannah C Moore^a

^a Wesfarmers Centre of Vaccines and Infectious Diseases, Telethon Kids Institute, The University of Western Australia, PO Box 855, West Perth, Western Australia, 6872 Australia

^b School of Paediatrics and Child Health, The University of Western Australia, GPO Box D184, Perth, Western Australia, 6840 Australia

^c Department of General Paediatrics, Princess Margaret Hospital for Children, GPO Box D184, Perth, Western Australia, 6840 Australia

^d PathWest Laboratory Medicine WA, Princess Margaret Hospital for Children, GPO Box D184, Perth, Western Australia, 6840 Australia

Authors email addresses

Faye J Lim – Janice.Lim@telethonkids.org.au

Christopher Blyth – christopher.blyth@uwa.edu.au

Nicholas de Klerk – Nick.deKlerk@telethonkids.org.au

Beverly Valenti – bev.valenti@gmail.com

Oliver J Rouhiainen – 20500745@student.uwa.edu.au

Dominic Yu-An Wu – 20748677@student.uwa.edu.au

Christopher Jansz – 20354233@student.uwa.edu.au

Hannah C Moore – Hannah.Moore@telethonkids.org.au

Corresponding author

Faye J Lim

Telethon Kids Institute, PO Box 855, West Perth, Western Australia, 6872 Australia

Email: Janice.Lim@telethonkids.org.au

Phone: +61 8 9489 7769

Conflicts of interest

No conflicts of interest to declare.

Author contributions

HCM and CB designed the study. BV, CJ, OR and DYW were involved in data collection. FJL, HCM, CB and NdK were involved in data analysis and interpretation of results. FJL wrote the first draft of the manuscript. All authors edited and approved the final version of the manuscript.

Abstract

Objective

Despite a recommendation for microbiological testing, only 45% of children hospitalised for respiratory infections in our previous data linkage study linked to a microbiological record.

We conducted a chart review to validate linked microbiological data.

Study Design and Setting

The chart review consisted of children aged <5 years admitted to 7 selected hospitals for respiratory infections in Western Australia, 2000-2011. We calculated the proportion of admissions where testing was performed and any pathogens detected. We compared these proportions between the chart review and our previous data linkage study. Poisson regression was used to identify factors predicting the likelihood of microbiological tests in the chart review cohort.

Results

From the chart review, 77% of 746 records had a microbiological test performed compared with 46% of 18,687 records from our previous data linkage study. Of those undergoing testing, 66% of the chart review and 64% of data linkage records had ≥ 1 respiratory pathogen(s) detected. In the chart review cohort, frequency of testing was highest in children admitted to metropolitan hospitals.

Conclusion

Validation studies are essential to ensure the quality of linked data. Our previous data linkage study failed to capture all relevant microbiological records. Findings will be used to optimise extraction protocols for future linkage studies.

Keywords

Validation, data linkage, chart review, respiratory infections, hospitalization, children

Running title

Validating administrative data

Word count

Abstract – 205 words

Manuscript – approximately 3,700 words

What's new?

- Despite a recommendation for microbiological testing, only 45% of paediatric hospital admissions for respiratory infections linked to a corresponding microbiological test record through data linkage
- Through the conduct of a medical chart review, we noted 77% of admissions had a microbiological test performed
- Frequency of pathogen detection was similar among those that had a test reported from the chart review and our previous data linkage study
- Lower levels of microbiological testing in our linked data were a result of missed records during the data extraction phase

- Data extraction protocols can now be optimised to ensure full capture of microbiological testing for future data linkage studies

1 Introduction

Acute lower respiratory infections (ALRI), such as pneumonia, bronchiolitis and whooping cough, are a significant cause of mortality and morbidity in children. In 2010, it was estimated that ALRI accounted for 20% of mortality among children aged 1-11 months worldwide [1]. In Western Australia (WA), Aboriginal and Torres Strait Islander children (hereafter referred to as Aboriginal) bear a greater burden of infection compared to non-Aboriginal children [2]. In 2000-2005, the rate of hospitalisations for pneumonia was 45 per 1000 child-years in Aboriginal children aged less than 6 months compared to 4 per 1000 child-years in similarly aged non-Aboriginal children [3].

ALRI are caused by a range of pathogens including *Streptococcus pneumoniae*, *Bordetella pertussis*, *Mycoplasma pneumoniae*, respiratory syncytial virus, influenza viruses, parainfluenza viruses, adenoviruses and human metapneumovirus. The aetiology of ALRI can be highly variable and specific to geographic location [2, 4]. Accurate local epidemiological data are therefore essential to monitor the circulating pathogens and can aid in assessing vaccine effectiveness. Population-based linkage of administrative data is an efficient method by which to obtain such information.

Data linkage is a process of combining data from separate datasets that relate to the same person, place or time [5, 6]. The Western Australian Data Linkage System (WADLS) was formed in 1995 and facilitates the linkage of data from a wide range of administrative datasets. These include birth, hospital and microbiological records, many of which are routinely collected at a population-level [5, 7].

The WADLS follows current best practice guidelines for data linkage and uses the separation principal when performing linkages. This means that those who have access to clinical data, such as researchers, will not have access to identifiers and vice-versa [8]. As data extraction is usually performed by the data custodians on the researchers' behalf, researchers need detailed knowledge of the dataset and the scope of available data prior to submitting a request for data to be extracted for linkage. This can pose a challenge when linking to a dataset for the first time as there may not be any documentation for the types of data available and its nuances, particularly if the dataset has not been used for research purposes before. With assistance from staff at the WADLS and custodians of microbiological data, we were the first in Australia to use the WADLS to link microbiological and administrative data to investigate the aetiology of ALRI in a total population cohort of WA children [9].

In WA, it is recommended for children admitted to hospital with a respiratory infection to undergo microbiological testing to determine the causative pathogen. Despite this recommendation, we found only 45% of hospitalisations for ALRI corresponded to a microbiological test record; some remote areas had less than 5% of ALRI hospitalisations linking to a test record [9]. We considered the possibility that this may be due to mislinks (i.e. misidentification of individuals across datasets) even though the proportions of mislinks is very low and continues to fall [10, 11]. This possibility was ruled out as the linkage rate of demographic data between the microbiological data source and the other datasets used in this project was 96% (C Garfield, personal communication). As this was the first time microbiology data had been linked, we were unable to exclude the possibility that the lower than expected proportions of linked hospital records may be due to low levels of

microbiological testing, missed records during the data extraction process or a combination of both.

Quality checks of administrative linked data for infectious diseases research has been flagged as an area needing further attention [12, 13]. Medical chart reviews have previously been used as a tool for validating linked data [14, 15]. A prior study in WA has used this method to identify gaps in the recording of comorbidity data on administrative datasets [14]. Given the novelty of linked microbiological data as well as the increasing demand and complexity for linked data [16], it is imperative that we validate linked hospital and microbiological data for future studies.

Due to privacy constraints surrounding the approved use of de-identified linked data, we were unable to re-identify individuals from our previous data linkage study who did not have a corresponding microbiological test record. As a way of validation, we conducted a chart review to determine if the low proportion of linked hospital and microbiological test records in our previous data linkage study was due to records being missed during data extraction or due to low levels of testing. We hypothesized that the low proportions of linkage in the previous study was due to records being missed during the extraction process rather than low levels of testing. Secondary to this, we also identified the demographic predictors of pathogen testing using data collected from the medical chart review so we may identify any groups that may have been missed during data extraction.

2 Materials and methods

2.1 Study setting and population

WA spans 2.5 million square kilometres with a population of 2.3 million people, approximately 4% of whom identify as Aboriginal [17]. Approximately 80% of the non-Aboriginal population reside in the Perth metropolitan area, while over 60% of the Aboriginal population reside in the rural and remote regions of the state [17]. Australia has a publicly-funded healthcare system. Located in metropolitan Perth, Princess Margaret Hospital for Children (PMH) is a public hospital and is the only tertiary paediatric hospital in the state. PathWest Laboratory Medicine (hereafter referred to as PathWest) is the only public laboratory system in WA. PathWest processes all microbiological specimens collected through public hospitals and is a reference laboratory for virology tests performed in other laboratories in the state.

2.2 Data linkage study (Cohort 1)

Our previous data linkage study consisted of linked hospital and microbiological data on a birth cohort of WA children to investigate the aetiology of ALRI [9]. Information on the data extraction, cleaning and coding for this group are described in more detail elsewhere [3, 9, 18]. Briefly, hospital and microbiological data were extracted for all singleton live births in WA between 1996 and 2005. A hospital admission for ALRI was defined as having a principal or co-diagnosis code for ALRI using a pre-determined selection of 9th and 10th edition International Classification of Disease codes (ICD9 and ICD10) [18]. A child was coded as Aboriginal if at least one record indicated that the child was Aboriginal [9, 18].

Data were restricted to admissions between 2000 and 2005 as microbiological testing data were only available for this period [9]. Of 19,857 admissions, 45.2% had linked to a corresponding microbiological test record [9]. To enable comparisons to the chart review cohort, records of children aged 5 years or more were excluded, thereby leaving 18,687 linked data records available for comparison, hereafter referred to as Cohort 1.

2.3 Medical chart review (Cohort 2)

We selected 7 major hospitals throughout WA from which to conduct a chart review (in the analysis referred to as Cohort 2). These hospitals collectively represented 57% of all hospitalisations for ALRI throughout WA. These were 3 metropolitan hospitals (including PMH), 2 rural and 2 remote hospitals. Our intention was to review no more than 900 admission records (300 from PMH and 100 from each of the remaining hospitals). To ensure representativeness of the sample of admissions to be reviewed to the total population data linkage cohort [9, 19], we set a number of restrictions for the chart review from each hospital. These restrictions were a) the child's date of birth was between 1 January 1996 and 31 December 2011; b) hospital admissions between 2000 and 2011 with an equal number of records from the time periods 2000-2005 and 2006-2011; c) the child was aged less than 5 years at the time of admission with 50% of all records selected in the <12-month age group, 25% in the 12-23-month age group, 25% in the 2-4-year age group and d) hospital admissions with a diagnosis code of ALRI, according to ICD10 diagnoses [20].

ALRI was defined as a principal or co-diagnosis of influenza (J09-J11), pneumonia (J12-18), bronchiolitis (J21), whooping cough (A37) or unspecified ALRI (J22). We set the selection

criteria so that records coded as influenza, pneumonia or bronchiolitis would each comprise 25% of all records from each hospital, while whooping cough and unspecified ALRI records would each comprise 12.5% of all records from each hospital. Multiple admission records of the same child were eligible for selection provided all other selection criteria were met.

Based on these criteria, a maximum of 888 records could be selected from the 7 hospitals.

The data custodian of the Hospital Morbidity Database System randomly selected admission records from each hospital that met these pre-established criteria. Of the 888 records requested, only 761 records (85.7%) were selected by the data custodian for review. This was because there were fewer than expected records coded as whooping cough or influenza in selected rural and remote hospitals that met all other selection criteria.

Supplementary records were not selected if there were an insufficient number of records in a particular cell. Figure 1 displays a flow chart of records that were requested, selected and then reviewed.

Demographic, clinical and microbiological data were reviewed from the 761 medical records by a research nurse (BV) and trained medical students (OJR, DYW, CJ) and recorded on a standardised data collection form. Demographic data collected included date and country of birth, gender and Aboriginal status. A child was coded as Aboriginal if indicated as such in medical charts. Clinical data collected included dates of admission and hospital discharge; risk factors and comorbidities present such as preterm birth, immunosuppression and chronic respiratory disease; and principal and secondary diagnoses. Immunosuppression was defined as prednisolone use for 2 weeks or more, use of a disease-modifying agent (e.g. chemotherapy) or any congenital or acquired immunodeficiency. Moderate to severe

preterm birth was defined as less than 34 weeks gestation. Microbiological data collected included respiratory specimens collected, tests requested and test results.

Ethical approvals to conduct the chart review were received from the Department of Health WA Human Research Ethics Committee, the Western Australian Aboriginal Health Ethics Committee, WA Country Health Service Research Ethics Committee and the Joondalup Health Campus Human Research Ethics Committee.

2.4 Statistical analysis

For both cohorts, ALRI diagnoses were identified from principal and co-diagnosis codes and grouped according to a hierarchical diagnosis algorithm [9], which ranked admissions in the following order: whooping cough, pneumonia, bronchiolitis, influenza, unspecified ALRI and bronchitis. For Cohort 1, geographical region was coded according to the place of the child's birth whereas the location of the hospital where the admission took place was used for Cohort 2.

Any mention of a test in the medical charts in Cohort 2 was considered to be indicative of a completed microbiological test. Nasopharyngeal swabs, nasopharyngeal/pernasal aspirates and throat swabs were grouped together as nasopharyngeal specimens for analyses.

Respiratory specimens were tested using one or more of immunofluorescence, polymerase chain reaction or viral culture. A positive test was defined by detection of a respiratory pathogen on nasopharyngeal specimens, isolation of a respiratory pathogen from blood or

pleural fluid or detection of elevated antibodies or seroconversion to respiratory pathogens such as *B. pertussis* or *S. pneumoniae*.

We calculated the proportion of admission records with any microbiological test reported by age, diagnosis, Aboriginal status, location and prematurity in both cohorts. We compared the proportions with a microbiological test between Cohorts 1 and 2 in 2000-2005 and between 2000-2005 and 2006-2011 for Cohort 2. Among those with a microbiological test reported, we then calculated the proportion of admission records with at least one respiratory pathogen detected by age, diagnosis and Aboriginal status. The frequencies of positive pathogen detections were also compared between Cohorts 1 and 2 in 2000-2005 and between 2006-2011 and 2000-2005 in Cohort 2.

In order to investigate the predictors of microbiological testing using information collected during the chart review, we performed univariate and multivariable Poisson regression analyses using the presence of a microbiological test as the outcome variable. Records were clustered by patient to allow for children with more than one admission record included in the chart review and robust standard errors were estimated, to correct for the usual over-estimation of error when using Poisson regression with common outcomes [21]. Aboriginal status, hospital location, gender, prematurity, presence of comorbidities, age and year of admission were used as potential predictor variables and were included in the multivariable model. Records with missing data on the variables analysed were dropped from the regression models. Data cleaning and analyses were conducted in Excel, IBM SPSS version 22 and STATA version 13. Confidence intervals for the crude ratios of proportions were calculated using ESCI [22] and EpiBasic [23].

3 Results

A total of 746 of 761 (98.0%) selected hospital admission records were reviewed for Cohort 2. Remaining records were not reviewed as records were washed away by floods (n=3) or could not be found (n=12; Figure 1). Sixteen children had multiple admission records that were included in the chart review, affecting a total of 34 records (4.6%); all records were included in the following analyses.

3.1 Comparisons between data linkage (Cohort 1) and chart review cohorts (Cohort 2)

There was a higher proportion of males in Cohort 1 compared to Cohort 2, however, both cohorts were comparable for Aboriginal status (Table 1). As fewer than requested number of records with a diagnosis of whooping cough and influenza were selected for review, it was expected that the distribution of ALRI diagnoses would differ. Whooping cough and influenza accounted for a larger proportion of Cohort 2 compared to Cohort 1 (Table 1). Chronic respiratory disease was the most common comorbidity recorded in Cohort 2 but comparable data on comorbidities were not available for Cohort 1.

Overall, 571 admission records (76.5%) from Cohort 2 had evidence of a microbiological test. In 2000-2005, the frequency of microbiological testing was 1.7 times (95% CI=1.6, 1.8) higher in Cohort 2 compared to Cohort 1 (Table 2). Testing was more frequently recorded in Cohort 2 than in Cohort 1 in both Aboriginal and non-Aboriginal children and across all age groups, prematurity status and locations, although this was less marked in records from rural regions (Table 2). There was twice as much testing reported in records of children aged

12 months or more in Cohort 2 than in Cohort 1 and for admissions coded as pneumonia and unspecified ALRI (Table 2).

Of the records with a microbiological test reported in Cohort 2, overall, 65.5% (n=374) had at least one respiratory pathogen detected. Overall, there were no differences in the frequency of pathogen detection between Cohorts 1 and 2 in 2000-2005 (Table 3). Pathogen detection was less often reported in records of children hospitalised with pneumonia in Cohort 2 compared to Cohort 1 (ratio=0.8, 95%CI=0.6, 1.0). Conversely, pathogens were more frequently reported in records of children aged 2-4 years in Cohort 2 than in Cohort 1 (ratio=1.4, 95%CI=1.1, 1.6).

3.2 Chart review cohort (Cohort 2) descriptive analyses

Overall, between 2005-2010 and 2000-2005, microbiological testing remained steady over time (ratio=1.0, 95% CI=0.9, 1.1). Testing was reported more frequently in 2006-2011 compared to 2000-2005 in records of Aboriginal children and children admitted to rural or remote hospitals (Table 2). However, the increase in testing recorded in rural and remote hospitals was not uniform across hospitals. For example, in one remote hospital testing doubled (ratio=2.2, 95%CI=1.3, 3.6) in 2006-2011 compared to 2000-2005 while over the same time period, testing remained stable in the other remote hospital (ratio=1.0, 95%CI=0.6, 1.6).

Of those records which indicated microbiological testing was conducted, pathogens were detected more frequently in children diagnosed with unspecified ALRI in 2006-2011 than in

2000-2005 (Table 3) although the difference was not statistically significant. Of those that had a pathogen detected in Cohort 2, the most common were respiratory syncytial virus (n=119, 31.8%), influenza viruses (n=111, 29.7%) and *B. pertussis* (n=61, 16.3%).

Univariate analysis of Cohort 2 showed that Aboriginal children and children presenting to rural and remote hospitals were least likely to have had a microbiological test performed (Table 4). Children older than 6 months were also less likely to have a microbiological test performed compared to children aged less than 6 months although this did not reach statistical significance (Table 4). After adjusting for all other covariates, hospital location was most predictive of having a microbiological test with children presenting to metropolitan hospitals most likely to have had a test (Table 4).

4 Discussion

We conducted a medical chart review to investigate the reason for the low proportion of hospital admission records that linked to a corresponding microbiological record in our previous data linkage study. We found that in 2000-2005, the proportion of hospital admissions for ALRI that documented a microbiological test was 1.7 times higher in the chart review cohort compared to our previous data linkage cohort. Of admissions that documented a microbiological test in 2000-2005, the frequency of pathogen detection was similar in both cohorts. Using data collected during the chart review and after adjusting for other factors, we found children admitted to metropolitan hospitals were more likely to undergo testing.

The higher proportion of testing observed in Cohort 2 compared to Cohort 1 suggests that the reason for low numbers of hospital records with a corresponding laboratory record that we observed in our previous linked data study was likely due to missed records. Extraction of PathWest data for linkage with hospital data for Cohort 1 was conducted as a proof-of-concept study as part of a previous program of work [9]. In this instance, there was limited documentation available on the codes used in the PathWest dataset, their changes over time and their utilisation. Therefore, even with assistance from data custodians, it is likely that all relevant records were not identified in the initial data request sent to the WADLS and as such, may not have been extracted for the previous data linkage study. These results demonstrate the power of this approach but highlight some of the complexities of data linkage and how important in-depth knowledge of the dataset and validation studies can be, particularly for newly linked datasets such as PathWest. With validation and better documentation of codes used in the PathWest dataset and how they have changed over time, PathWest remains a powerful source of linked pathology data and may still be used as a data source for future studies looking at pathogen-specific burden of respiratory and other infectious diseases or effects at the population-level [9].

To comply with best practise protocols for data linkage [8] and due to constraints surrounding the approved use of linked data, we were unable to re-identify individuals from Cohort 1 with hospital admissions that did not have a corresponding PathWest record to individually access their medical charts. This meant that we were unable to directly identify predictors of testing or conduct sensitivity analyses. We have attempted to compensate for this by imposing strict selection criteria for Cohort 2 in order to achieve a representative,

but small, sample of admissions which we could review across varying demographics such as age, location and ALRI diagnosis.

Of the admission records that had a microbiological test reported, overall, there was no difference in the frequency of pathogen detection between the two cohorts. However, the proportion of records from 2-4 year-olds with reported positive pathogen detections were 1.4 times higher in Cohort 2 compared to Cohort 1. This difference may be partially explained by the inadvertent omission of blood culture records during data extraction for Cohort 1, which has been highlighted previously [9]. Nonetheless, the high degree of similarity in the positivity rate overall suggests that our previous linked dataset did provide a reasonable snapshot of the aetiology of ALRI at a population level [9].

During the chart review, we observed an increase in the frequency of microbiological testing reported in 2006-2011 compared to 2000-2005 among rural and remote hospitals.

However, this was not uniform across all hospitals in these 2 regions. As such, inter-hospital variability in testing practices is likely to account for the differences in microbiological testing. If not accurately documented, these differences are likely to impact on data quality which in turn has the potential to introduce bias in studies using linked data [24, 25]. Future studies using microbiological data need to take into account the potential impact of differential testing practises on study findings, and validation studies may need to use a wider sample of hospitals to fully assess inter-hospital heterogeneity.

While both cohorts are broadly comparable, some variables were coded differently which is a limitation of our validation analyses. One of these is Aboriginal status, where variability in

coding has been shown to result in over- or under-estimations of gains in health outcomes such as mortality and infection [26, 27]. In addition, Cohort 2 only includes records from a small subset of hospitals in the state, almost all of which are public hospitals (one hospital accepts both publicly- and privately-funded patients). Collectively, these hospitals still accounted for over half of all paediatric admissions for ALRI, notwithstanding the omission of records from private hospitals, which accounted for approximately 12% of admissions in Cohort 1 (data not shown). While our validation exercise may give a crude measure of the validity of our previous data linkage study, it is one of the few options available without re-identifying children in our linked dataset [8].

To our knowledge, this is the first study to validate linked hospital and microbiological data in Australia. Prior studies have highlighted potential biases and errors in these data that could easily be missed without validation studies [14, 25, 28]. For example, an earlier validation study in WA has found that comorbidity data are better recorded in medical charts compared to administrative datasets but further studies have concluded that this may be due to insufficient follow-up time [14, 29]. WA is one of the few jurisdictions that is currently able to link microbiological data of routine testing of infectious disease pathogens and there is increasing interest in locating and possibly linking microbiological data across jurisdictions in Australia and between countries [30]. Therefore, this study represents an essential step in the continued use of administrative data, especially routinely collected microbiological data. Findings from this study will be used to optimise the data extraction protocol for further data linkage studies investigating the pathogen-specific burden of ALRI and the clinical impact of co-infection.

5 Acknowledgements

This study was conducted with funds from the Telethon Kids Institute Small Grants scheme. We would like to acknowledge the Data Linkage Branch, in particular Jessica Lee and the In-patient Data Manager, Paul Stevens, for their assistance in selecting the hospital records for review. We would also like to acknowledge members of the Infectious Diseases Community Reference Group for supporting this study. We would like to thank staff at all the participating hospitals for their assistance in facilitating data collection for this study. This study was conducted as part of FJL's studies towards a Doctor of Philosophy, which is funded by a University Postgraduate Award at the University of Western Australia. FJL is also supported by a National Health and Medical Research Council project grant (APP1045668). HCM is supported by National Health and Medical Research Council Fellowship (APP1034254).

6 Figure and Table legends

Figure 1. Flow chart of medical chart review records requested, received and analysed (Cohort 2)

Table 1. Comparisons between the data linkage (Cohort 1) and chart review cohorts (Cohort 2)

Table 2. Frequency of microbiological testing in data linkage (Cohort 1) and chart review cohorts (Cohort 2)

Table 3. Positivity rates of respiratory pathogen detection in data linkage (Cohort 1) and chart review cohorts (Cohort 2)

Table 4. Modified Poisson regression of demographic predictors of having a microbiological test in the chart review (Cohort 2)

7 References

- [1] Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2013;380:2095-128.
- [2] Torzillo P, Dixon J, Manning K, Hutton S, Gratten M, Hueston L et al. Etiology of acute lower respiratory tract infection in Central Australian Aboriginal children. *Pediatr Infect Dis J* 1999;18:714-21.
- [3] Moore HC, Lehmann D, de Klerk N, Jacoby P, Richmond PC. Reduction in disparity for pneumonia hospitalisations between Australian Indigenous and non-Indigenous children. *J Epidemiol Community Health* 2012;66:489-94.
- [4] Kusel MMH, de Klerk NH, Holt PG, Keadze T, Johnston SL, Sly PD. Role of respiratory viruses in acute upper and lower respiratory tract illness in the first year of life: a birth cohort study. *Pediatr Infect Dis J* 2006;25:680-6.
- [5] Holman CDAJ, Bass JA, Rosman DL, Smith MB, Semmens JB, Glasson EJ et al. A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system. *Aust Health Rev* 2008;32:766-77.
- [6] Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public health research. *Annu Rev Public Health* 2011;32:91-108.
- [7] Data Linkage WA. Western Australian Data Linkage Information 2010 [accessed 15 June 2012]. Available from: <http://www.datalinkage-wa.org/about-us/faq#45>.
- [8] Kelman CW, Bass AJ, Holman CDJ. Research use of linked health data — a best practice protocol. *Aust N Z J Public Health* 2002;26:251-5.

- [9] Moore HC, de Klerk N, Keil AD, Smith DW, Blyth CC, Richmond P et al. Use of data linkage to investigate the aetiology of acute lower respiratory infection hospitalisations in children. *J Paediatr Child Health* 2012;48:520-28.
- [10] Holman CDAJ, Bass AJ, Rouse IL, Hobbs MST. Population-based linkage of health records in Western Australia: development of a health services research linked database. *Aust N Z J Public Health* 1999;23:453-9.
- [11] Rosman D, Garfield C, Fuller S, Stoney A, Owen T, Gawthorne G, editors. Measuring data and link quality in a dynamic multi-set linkage system. Symposium on Health Data Linkage; 20-21 March 2002; Sydney, New South Wales, Australia.
- [12] Leal JR, Laupland KB. Validity of ascertainment of co-morbid illness using administrative databases: a systematic review. *Clin Microbiol Infect* 2010;16:715-21.
- [13] Harron K, Wade A, Muller-Pebody B, Goldstein H, Gilbert R. Opening the black box of record linkage. *J Epidemiol Community Health* 2012;66:1198.
- [14] Preen DB, Holman CDAJ, Lawrence DM, Baynham NJ, Semmens JB. Hospital chart review provided more accurate comorbidity information than data from a general practitioner survey or an administrative database. *J Clin Epidemiol* 2004;57:1295-304.
- [15] Benchimol EI, Guttman A, Mack DR, Nguyen GC, Marshall JK, Gregor JC et al. Validation of international algorithms to identify adults with inflammatory bowel disease in health administrative data from Ontario, Canada. *J Clin Epidemiol* 2014;67:887-96.
- [16] Data Linkage WA. Project complexity 2014 [updated 9 June 2014; accessed 2 July 2014]. Available from: <http://www.datalinkage-wa.org/projects/project-complexity>.
- [17] Codde J. Rates Calculator. 9.5.5 ed. Perth, Western Australia: Health Information Centre, Department of Health; 2013.

- [18] Moore H, de Klerk N, Richmond P, Lehmann D. A retrospective population-based cohort study identifying target areas for prevention of acute lower respiratory infections in children. *BMC Public Health* 2010;10:757.
- [19] Moore H, de Klerk N, Jacoby P, Richmond P, Lehmann D. Can linked emergency department data help assess the out-of-hospital burden of acute lower respiratory infections? A population-based cohort study. *BMC Public Health* 2012;12:703.
- [20] Australian Institute of Health and Welfare. The International Statistical Classification of Diseases and Related Health Problems. 2nd ed. Sydney, Australia: National Centre for Classification in Health; 2000.
- [21] Zou G. A modified Poisson regression approach to prospective studies with binary data. *Am J Epidemiol* 2004;159:702-6.
- [22] Cumming G. Exploratory software for confidence intervals. Melbourne, Australia: La Trobe University; 2011.
- [23] Juul S. EpiBasic. 1.0 ed.
- [24] Bohensky M, Jolley D, Sundararajan V, Evans S, Pilcher D, Scott I et al. Data linkage: a powerful research tool with potential problems. *BMC Health Serv Res* 2010;10:346.
- [25] Harron K, Goldstein H, Wade A, Muller-Pebody B, Parslow R, Gilbert R. Linkage, evaluation and analysis of national electronic healthcare data: application to providing enhanced blood-stream infection surveillance in paediatric intensive care. *PLoS ONE* 2013;8:e85278.
- [26] Draper GK, Somerford PJ, Pilkington AAG, Thompson SC. What is the impact of missing Indigenous status on mortality estimates? An assessment using record linkage in Western Australia. *Aust N Z J Public Health* 2009;33:325-31.

- [27] Thompson S, Woods J, Katzenellenbogen J. The quality of Indigenous identification in administrative health data in Australia: insights from studies using data linkage. *BMC Med Inform Decis Mak* 2012;12:133.
- [28] Harron K, Wade A, Gilbert R, Muller-Pebody B, Goldstein H. Evaluating bias due to data linkage error in electronic healthcare records. *BMC Med Res Methodol* 2014;14:36.
- [29] Mnatzaganian G, Ryan P, Norman PE, Hiller JE. Accuracy of hospital morbidity data and the performance of comorbidity scores as predictors of mortality. *J Clin Epidemiol* 2012;65:107-15.
- [30] Boyd J, Ferrante A, O'Keefe C, Bass A, Randall S, Semmens J. Data linkage infrastructure for cross-jurisdictional health-related research in Australia. *BMC Health Serv Res* 2012;12:480.