Journal of Clinical Epidemiology 69 (2016) 23−31

Journal of Clinical Epidemiology

Optimization is required when using linked hospital and laboratory data to investigate respiratory infections

Faye J. Lim[a,*], Christopher C. Blyth[a,b,c,d], Nicholas de Klerk[a], Beverly Valenti[a], Oliver J. Rouhiainen[b], Dominic Yu-An Wu[b], Christopher S. 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*

Accepted 17 August 2015; Published online 21 August 2015

Abstract

Objective: Despite a recommendation for microbiological testing, only 45% of children hospitalized 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 seven 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 optimize extraction protocols for future linkage studies.

Keywords: Validation; Data linkage; Chart review; Respiratory infections; Hospitalization; Children

1. Introduction

Acute lower respiratory infections (ALRIs), 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 hospitalizations 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 [2].

ALRIs 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 etiology of ALRI can be highly variable and specific to geographic location [3,4]. Accurate local epidemiologic 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 data sets 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 data sets. These include birth, hospital, and microbiological records, many of which are routinely collected at a population level [5,7].

Conflicts of interest: There are no conflicts of interest to declare.

* Corresponding author. Tel.: +61 8 9489 7769; fax: +61 8 9489 7700.
E-mail address: Janice.Lim@telethonkids.org.au (F.J. Lim).

http://dx.doi.org/10.1016/j.jclinepi.2015.08.003
0895-4356/

What's new?

- Despite a recommendation for microbiological testing, only 45% of pediatric hospital admissions for respiratory infections are 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 who 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 optimized to ensure full capture of microbiological testing for future data-linkage studies.

The WADLS follows current best practice guidelines for data linkage and uses the separation principle 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 data set and the scope of available data before submitting a request for data to be extracted for linkage. This can pose a challenge when linking to a data set for the first time as there may not be any documentation for the types of data available and its nuances, particularly if the data set 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 etiology 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 hospitalizations for ALRI corresponded to a microbiological test record; some remote areas had less than 5% of ALRI hospitalizations linking to a test record [9]. We considered the possibility that this may be due to mislinks (ie, misidentification of individuals across data sets) even though the proportions of mislinks are very low and continue to fall [10,11]. This possibility was ruled out as the linkage rate of demographic data between the microbiological data source and the other data sets 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 have 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 data sets [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.

Because of privacy constraints surrounding the approved use of deidentified linked data, we were unable to reidentify 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 were 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 km^2 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, whereas over 60% of the Aboriginal population reside in the rural and remote regions of the state [17]. Australia has a publicly funded health-care system. Located in metropolitan Perth, Princess Margaret Hospital for Children (PMH) is a public hospital and is the only tertiary pediatric 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