

## Appropriateness of administrative data for vaccine impact evaluation: the case of pneumonia hospitalizations and pneumococcal vaccine in Brazil

S. SGAMBATTI<sup>1</sup>\*, R. MINAMISAVA<sup>2</sup>, E. T. AFONSO<sup>1,3</sup>, C. M. TOSCANO<sup>4</sup>,  
A. L. BIERRENBACH<sup>4</sup> AND A. L. ANDRADE<sup>4</sup>

<sup>1</sup> Department of Medicine, Pontifical Catholic University of Goiás, Goiânia, Goiás, Brazil

<sup>2</sup> Faculty of Nursing, Federal University of Goiás, Goiânia, Goiás, Brazil

<sup>3</sup> Department of Pediatrics, School of Medicine, Federal University of Goiás, Goiânia, Goiás, Brazil

<sup>4</sup> Department of Community Health, Federal University of Goiás, Goiânia, Goiás, Brazil

Received 13 November 2013; Final revision 2 March 2014; Accepted 22 March 2014;  
first published online 23 April 2014

### SUMMARY

Ten-valent pneumococcal conjugate vaccine (PCV10) was recently introduced into the Brazilian Immunization Programme. Secondary data are used as a measurement of community-acquired pneumonia (CAP) burden, but their completeness and reliability need to be ascertained. We performed probabilistic linkage between hospital primary data from active prospective population-based surveillance (APS) and hospital secondary data from the Hospital Information System administrative database of the National Unified Health System (SIH-SUS). Children aged 2–23 months hospitalized during January–December 2012 were identified. Incidence rates of hospitalized CAP were estimated. Agreement of case identification was measured by kappa index. A total of 1639 (26%) CAP cases were identified in APS and 1714 (35%) in SIH-SUS. Of these 3353 records, 1127 CAP cases were present in both databases. Kappa on CAP case identification was 0·72 (95% confidence interval 0·69–0·75). CAP hospitalization incidence using administrative (5285/100 000) and hospital (5054/100 000) primary data were similar ( $P=0\cdot184$ ). Our findings suggest that administrative databases of hospitalizations are reliable sources to assess PCV10 impact in time-series analyses.

**Key words:** Administrative data, children, community-acquired pneumonia, linkage, pneumococcal vaccine, pneumonia surveillance, vaccine impact.

### INTRODUCTION

Acute respiratory infections, in particular pneumonia, account for high morbidity and mortality in infants throughout the world [1, 2]. The proportion of pneumonia cases requiring hospitalization ranges from 5·9% to 16·8%, resulting in about 15 million

hospitalized pneumonia cases per year [3]. Bacterial pneumonia is responsible for 20–40% of hospitalizations of children aged <5 years in the Americas [4], with *Streptococcus pneumoniae* the most frequent aetiological agent, contributing 60–75% of all childhood cases with known aetiology [5].

In Brazil, pneumonia is the main cause of hospitalization in childhood [6]. A population-based active surveillance of community-acquired pneumonia (CAP) study conducted from May 2007 to June 2009 in the Central-Western Region of Brazil demonstrated that pneumonia hospitalizations were responsible for

\* Author for correspondence: Dr S. Sgambatti, Department of Medicine, Pontifical Catholic University of Goiás, Goiânia, Goiás, Brazil. Av. Universitária 1.440, Setor Universitário, Goiânia, Goiás, CEP 74605-010, Brasil.  
(Email: sasgambatti@gmail.com)

a significant annual burden in children aged <3 years (3840 cases/100 000 children) [7].

Ten-valent pneumococcal conjugate vaccine (PCV10) was introduced in 2010 into the Brazilian routine immunization programme [8]. It is thus imperative to plan for the evaluation of its impact. Potential sources of data for such evaluation include hospital primary data collected in observational studies and hospital secondary data from administrative data sources. Although hospital secondary data have been commonly used in long-term vaccine impact assessment studies as a measure of CAP burden [9, 10], their completeness and reliability need to be ascertained, especially in Brazil. This can be pursued through its comparisons with hospital primary data.

An active prospective population-based surveillance (APS) of hospitalized CAP in children has been conducted since 2011 in Goiânia, Brazil, with ongoing collection of hospital primary data. We therefore had the opportunity to ascertain the quality and appropriateness of hospital secondary hospitalization data from nationwide administrative databases compared to hospital primary data from the APS. In particular, we estimated the burden of hospitalized CAP cases in children considering both data sources, with a view of using hospital secondary data source in future vaccine impact assessment studies.

## METHODS

### Study location

The present investigation was conducted in Goiânia municipality, in the Central-Western Region of Brazil. The population of Goiânia is estimated at 1 333 767 inhabitants, of which 32 429 are children aged 2–23 months [11]. In 2011, infant mortality in the municipality was 14.3/1000 live births [12].

Healthcare in Brazil is provided by public and private sectors. Public healthcare service is provided by the National Unified Health System (SUS), which offers free, universal coverage to all the population. The SUS Hospital Information System (SIH-SUS) is an administrative database primarily used for SUS reimbursement purposes in Brazil.

### Study design

Hospitalization incidence rates of CAP in infants aged 2–23 months were estimated considering both SIH-SUS and APS databases. Record linkage of all

CAP hospitalizations recorded in the SIH-SUS and APS databases was conducted and agreement on diagnosis was assessed. The assessment of type of health insurance as a risk factor for CAP hospitalizations was conducted using data from the APS study.

## Sources of data

### *Hospital primary data: APS database*

The APS study has been ongoing since 2011 to evaluate the impact of PCV10 against hospitalized CAP in childhood. All-cause hospitalization in children aged <36 months was monitored in all 17 hospitals providing care for the paediatric population in the city. Four of the 17 hospitals are exclusively funded by SUS, whereas the remaining 13 hospitals provide care for patients through both SUS and private insurance. Thus, the APS study includes children covered by both public and private healthcare systems. Written informed consent was obtained from the parent(s)/legal guardian(s) of each study participant before revising the medical records. Medical charts of suspected CAP cases were fully reviewed in order to ascertain the cause of hospitalization and other underlying conditions of each hospitalized child. Socio-demographic data and information on type of health insurance were also obtained. Trained nurses were responsible for data extraction from the medical records. Data were entered onsite into an electronic system, using Smartphones and an in-house software (*e-Pneumo*), developed specifically for this study. Data were transmitted online to a server (icloud) and could be accessed in real time.

Ethical approval for this investigation was granted by the Ethics Committee, Federal University of Goiás, Goiânia, Brazil (protocol no. 100/11).

### *Hospital secondary data: SIH-SUS database*

SIH-SUS is a nationwide administrative database in which all hospitalizations paid by the SUS system in Brazil are recorded. These include hospitalizations occurring in SUS hospitals and those occurring in private or non-for-profit hospitals but paid by SUS; SUS payments for hospitalizations in public hospitals and reimbursement for hospitalizations in private hospitals outsourced by the SUS are generated by the SIH-SUS database. Admission diagnoses for each hospitalization are assigned by attending pediatricians, according to clinical findings at admission. After hospital discharge, medical archivists assign a diagnosis

and its corresponding International Classification of Diseases–10th revision (ICD-10) code [13], by reviewing medical charts, accepting or changing the admission diagnosis, according to the patient's information available from the hospitalization period. As such, the SIH-SUS database is able to generate hospital secondary data for epidemiological studies, and consequently contribute to planning and assessing public health policies in Brazil [14]. It is estimated that the SIH-SUS database covers about 70% of all hospitalizations in the country, recording about 12 million admissions annually [15, 16]. The remaining 30% of all hospitalizations, which take place in private hospitals not outsourced by SUS, are not included in the SIH-SUS database. Local health authorities are responsible for certifying the reliability of the information and later transmitting it to regional and national levels. Thus, the Brazilian Ministry of Health, at the national level, manages the SIH-SUS database, providing updated databases periodically [15, 17]. For this investigation, information from the SIH-SUS database was extracted in March 2013.

### Study population

The present investigation included children aged 2–23 months living in Goiânia, hospitalized due to all causes from January 2012 to December 2012, in all 17 paediatric hospitals participating in the APS study.

### Definition of pneumonia hospitalization

For both data sources (APS and SIH-SUS databases), we considered records in which discharge diagnosis was one of the following ICD-10 codes, chapter X: J12, viral pneumonia; J13, *Streptococcus pneumoniae*-caused pneumonia; J14, *Haemophilus influenzae*-caused pneumonia; J15, bacterial pneumonia not otherwise specified; J16, other infectious organism not elsewhere classified; J17, pneumonia in diseases classified elsewhere; and J18, unspecified organism [13]. We considered only the ICD-10 codes used as the principal diagnosis to detect CAP, since nosocomial pneumonia is more likely to be recorded as a secondary diagnosis. Two- and three-digit ICD codes were considered.

### Definition of other diagnoses

A similar methodology was used to assign ICD-10 codes for non-CAP cases for both APS and

SIH-SUS databases. The chapters and respective ICD-10 codes used are given in the Supplementary online material.

### Record linkage

A probabilistic record linkage of all-cause hospitalized cases recorded in the SIH-SUS and APS databases was performed. We used the Link Plus program, a free probabilistic record linkage software [18]. To assess the reliability of pneumonia diagnosis from SIH-SUS, we defined children aged 2–23 months as our target population, since PCV10 vaccination was introduced for children aged <2 years.

Prior to record linkage, both databases underwent a pre-processing stage of quality analysis to minimize errors and increase the likelihood of finding matched records [19]. These procedures included mainly standardization of dates (admission data, discharge data, and date of birth) and nominal variables (patient's name and patient's mother's name). Differently from APS, where each record represents a hospitalization episode of a single child, in the SIH-SUS database it is possible to find duplicated records concerning to the same child during the same hospitalization period. As there might be some variations in the identification details of such duplicated records, we decided not to remove them *a priori* from the linkage with the expectation that by leaving all duplicated records in the SIH-SUS database we would increase the likelihood of finding a match with an APS record.

The record linkage process consisted of the following steps:

- (1) The following matching variables were included in the probabilistic linkage: patient's name, date of birth, mother's name, admission date, and discharge date.
- (2) Gender was used as a blockage variable to increase linkage processing speed.
- (3) A score was computed for each linked pair by the Link Plus program to assess the agreement and disagreement of the variables selected for the linkage. The higher the score, the greater the probability of finding a true matched pair.
- (4) A score cut-off point equal to 3.0 was defined.
- (5) We used manual review for discriminating pairs of records found by the Link Plus software into matched and unmatched pairs, using information from the matching variables and also the principal and secondary diagnosis.

Table 1. Association between community-acquired pneumonia and type of health insurance in children aged 2–23 months admitted to hospital, considering patients enrolled in the APS study. Goiânia, Brazil, January–December 2012

Type of health insurance (APS)	Community-acquired pneumonia			RR	95% CI	P value
	Yes (%)	No (%)	Total			
Public insurance (SUS)	1366 (32.7)	2809 (67.3)	4175 (67.1)	2.45	2.18–2.76	0.000
Private insurance	273 (13.3)	1772 (86.7)	2045 (32.9)			
Total	1639 (26.3)	4581 (73.7)	6220 (100.0)			

APS, Active prospective population-based surveillance; SUS, National Unified Health System; RR, relative risk; CI, confidence interval.

Matched pairs were defined as those that, after manual review, had the same or very similar information on patient's name, mother's name and birthday and that corresponded to the same hospitalization period. As it was possible that the admission and discharge dates were not the same in APS and SIH-SUS records, we considered as a matched pair those that had up to 10 days' difference between the discharge data of the former record and the admittance of the later one.

#### Data analysis

The assessment of type of health insurance as a risk factor for CAP hospitalization was conducted using data from the APS study. The association between type of health insurance and CAP was assessed by relative risk (RR), and the 95% confidence interval (CI) was estimated.

Hospitalization incidence rates of CAP (all-cause pneumonia hospitalization) and their 95% Poisson CIs were estimated for infants aged 2–23 months. Numerator data were obtained from both primary (APS) and secondary (SIH-SUS) databases. Denominator data were population estimates for 2012 obtained from the Brazilian Institute of Geography and Statistics (IBGE) [20].

Record linkage was performed to identify potential pairs of cases from both databases. Matched pairs were separated into those with concordant and discordant diagnoses. Matched pairs with concordant diagnoses were those that had a CAP diagnosis in the APS database and a J12–J18 ICD-10 code in the SIH-SUS database, or those that did not have a CAP diagnosis in the APS database and that had another ICD-10 code rather than J12–J18 in the SIH-SUS database. All other matched pairs were considered discordant.

Agreement of matched pairs on CAP diagnosis (ICD-10: J12–J18) and non-CAP diagnosis (all

ICD-10 except for J12–J18) was measured by adjusted Kappa index and statistical significance was tested by MacNemar's test.

All *P* values <0.05 were considered statistically significant. All analyses were performed using SPSS v. 18.0 (SPSS Inc., USA).

#### RESULTS

During the study period, the APS database computed a larger number of all-cause hospitalized children ( $n=8573$ ) compared to the SIH-SUS database ( $n=6686$ ). This was an expected finding, since the APS included children that were hospitalized using both private and public health insurances, whereas SIH-SUS database included only children hospitalized through SUS.

A total of 6220 hospitalizations in children aged 2–23 months were recorded in the APS database, representing 1639 cases (26.3%) hospitalized due to CAP. More than two thirds of hospitalizations (4175/6220, 67.1%) from the APS were SUS hospitalizations. CAP accounted for 32.7% of all SUS and 13.3% of all private/insurance paid hospital admissions. This indicates that children admitted through SUS were more likely to have a diagnosis of CAP than another diagnosis (RR 2.45, 95% CI 2.18–2.76) (Table 1).

In the SIH-SUS database, 4903 hospitalizations in children aged 2–23 months were recorded, all of which were admitted through SUS. Of these, 1714 (35%) were hospitalized due to CAP.

Incidence rates of CAP hospitalizations were similar for the SIH-SUS database (5285/100 000 children, 95% CI 5046–5533) and APS database (5054/100 000 children, 95% CI 4820–5296), although SIH-SUS detected a higher number of CAP cases (Table 2).

Table 2. Incidence rates of hospitalization due to community-acquired pneumonia in infants, stratified by age and source of data. Goiânia, Brazil, January–December 2012

Age group (months)	Hospital primary data (APS)		Administrative database (SIH-SUS)	
	N/population*	Incidence/100 000	N/population*	Incidence/100 000
2–11	895/14 970	5978 (5593–6384)	981/14 970	6553 (6149–6976)
12–23	744/17 459	4261 (3961–4579)	733/17 459	4198 (3899–4514)
Total	1639/32 429	5054 (4820–5296)	1714/32 429	5285 (5046–5533)

Values in parentheses are 95% confidence intervals.

APS, Active prospective population-based surveillance; SIH-SUS, Hospital Information System of National Unified Health System.

\* Source: <http://www2.datasus.gov.br/DATASUS/index.php?area=0206&VObj=http://tabnet.datasus.gov.br/cgi/deftohtm.exe?ibge/cnv/pop>

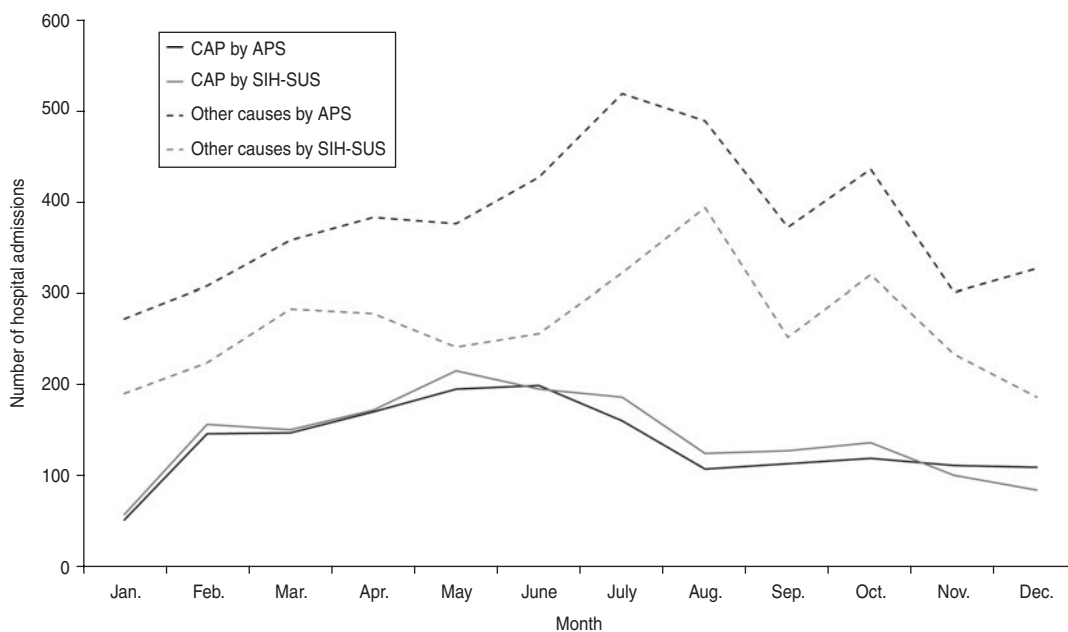


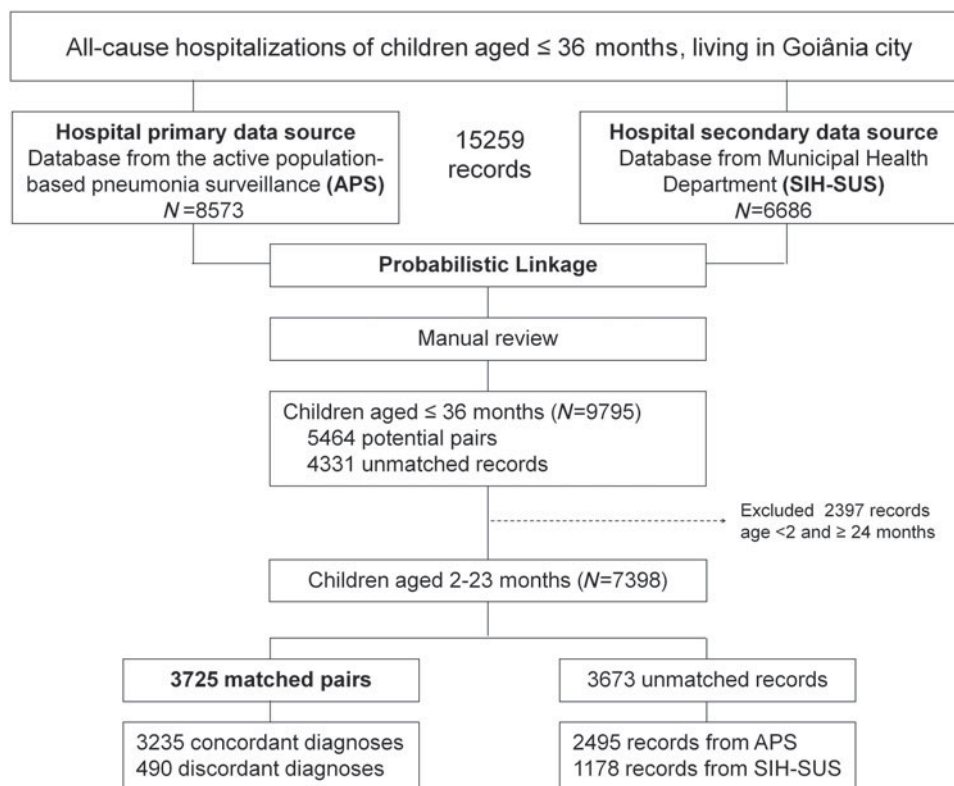
Fig. 1. Monthly number of children aged 2–23 months hospitalized by diagnosis and source of data. Goiânia, Brazil, January–December 2012. CAP, Community-acquired pneumonia; APS, active prospective population-based surveillance; SIH-SUS, Hospital Information System of National Unified Health System.

Trends in the number of cases of CAP and non-CAP hospitalizations in both SIH-SUS and APS databases are shown in Figure 1. While the number of CAP cases recorded in both APS and SIH-SUS databases was similar throughout the year, there was an excess number of non-CAP cases recorded in the APS database, which was fairly constant over the year. The parallelism of both CAP and non-CAP trend lines were only lost during November and December, with a decrease in cases recorded in the SIH-SUS database.

Results of record linkage of hospitalized children identified from hospital primary and secondary data

sources are presented in Figure 2. After running the Link Plus software and discarding children that were not aged between 2 months and 23 months, a total of 7398 hospitalized cases were identified in either of the databases, of which 3725 (50.4%) records were considered as matched pairs, while 3673 (49.6%) were unmatched pairs. Of the 3725 matched pairs, 1127 (30.2%) were concordant on CAP diagnosis, 2108 (56.6%) were concordant on non-CAP diagnosis, and 490 (13.2%) were discordant diagnoses. The Kappa agreement index on CAP diagnosis considering data from both databases was 0.72 (95% CI 0.69–0.75), with an agreement of 86.85% (95% CI





**Fig. 2.** Linkage of hospitalized children identified from hospital primary and secondary data sources. Goiânia, Brazil, January–December 2012.

85.72–87.89), although in the records that could be matched SIH-SUS had a significantly higher number of CAP cases (305 only by SIH-SUS vs. 185 only by APS;  $P < 0.001$ , McNemar's test) (Table 3).

Table 4 shows the ICD-10 codes found in both APS and SIH databases and the disagreement on CAP diagnosis between them. The majority (89.5%) of discordant CAP diagnoses was classified as having other respiratory diseases, such as influenza, bronchiolitis, bronchitis and asthma in the APS database.

## DISCUSSION

In this investigation we found that the burden of CAP estimated by considering administrative and hospital primary data was similar, indicating the feasibility of using the SIH-SUS database as a source of data to assess pneumonia interventions, such as PCV10 vaccination. Furthermore, CAP diagnosis in children aged <2 years was concordant when comparing administrative and hospital primary databases.

Hospital primary data remains the gold standard data source for estimating disease burden, and therefore for use in vaccine impact evaluation studies. Nonetheless, use of hospital primary data is

dependent on the existence of ongoing population-based surveillance studies, which are usually expensive, and are difficult to perform due to complex logistics. Hospital secondary data, on the other hand, are available in most countries, are usually representative, and are important sources of data for vaccine impact evaluation studies [10, 21–23]. Despite its low cost and easy web-based accessibility, administrative databases in Brazil have been underutilized in vaccine impact evaluation studies, probably due to challenges in ascertaining their quality and coverage.

In developed countries, administrative data have traditionally been used in epidemiological studies to estimate the burden of CAP and to assess the impact of vaccination on childhood pneumonia [9, 24, 25]. High concordance between all-cause diagnosis obtained from medical record review data and administrative databases has also been reported by other studies [26–29]. This is the first study conducted in Brazil to focus on agreement of CAP diagnosis in childhood by comparing hospital primary and administrative databases.

ICD-10 codes J12–18 and their corresponding codes in ICD-09 are usually used as a 'proxy' for

Table 3. Agreement between CAP diagnoses in hospitalized children aged 2–23 months, by source of data. Goiânia, Brazil, January–December 2012

Administrative data (SIH-SUS)	Hospital primary data (APS)		Total
	CAP cases	Non-CAP cases	
CAP cases	1127	305	1432
Non-CAP cases	185	2108	2293
Total	1312	2413	3725

CAP, Community acquired pneumonia; APS, active prospective population-based surveillance; SIH-SUS, Hospital Information System of National Unified Health System; Kappa index (pneumonia vs. non-pneumonia) = 0.72 (0.69–0.75).

Table 4. Causes of hospitalization in discordant pairs of cases identified as hospitalized CAP in any data source in children aged 2–23 months, Goiânia, Brazil, January–December 2012

Group of causes (ICD-10)	CAP diagnosis by SIH-SUS and non-CAP by APS (N = 305) n (%)	CAP diagnosis by APS and non-CAP by SIH-SUS (N = 185) n (%)
I. Certain infectious and parasitic diseases (A00–B99)	6 (2.0)	29 (15.7)
IV. Endocrine, nutritional and metabolic diseases (E00–E90)	4 (1.3)	7 (3.8)
X. Diseases of respiratory system (J00–J99)*	273 (89.5)	125 (67.6)
XIV. Diseases of genitourinary system (N00–N99)	4 (1.3)	3 (1.6)
IX. Diseases of circulatory system (I00–I99)	3 (1.0)	1 (0.5)
Other causes†	15 (4.9)	20 (10.8)

ICD-10, International Classification of Diseases, 10th Revision; CAP, community-acquired pneumonia; SIH-SUS, Hospital Information System of National Unified Health System; APS, active prospective population-based surveillance.

\* Except for CAP (ICD-10: J12–18).

† Diseases of digestive system (ICD-10: K00–K93); diseases of nervous system (ICD-10 G00–G99); diseases of skin and subcutaneous tissue (ICD-10: L00–L99); congenital malformation and chromosomal abnormalities (ICD-10: Q00–Q99); injury, poisoning, and certain other consequences of external causes (ICD-10: S00–T98); external causes of morbidity and mortality (ICD-10: V01–Y98); factors influencing health status and contact with health services (ICD-10: Z00–Z99).

pneumonia diagnosis in hospitalization databases. The use of such codes improves the overall sensitivity of pneumonia detection [23, 24, 30]. Similarly to that observed in other countries, in Brazil specific codes for pneumococcal pneumonia (ICD-10: J13) are rarely reported in administrative hospitalization information systems [30]. This is probably due to the fact that aetiological confirmation of pneumonia in children is rarely performed routinely in healthcare services. To improve specificity of pneumonia coding in hospitalization databases, further training of hospital medical archivists (who are in charge of assigning discharge diagnosis) focusing on the importance of using more accurate ICD-10 codes might be helpful, as these information systems are being more frequently used for epidemiological assessment, not only for administrative and financial purposes.

In recent years, improvements on data quality and accessibility of hospital information systems in Brazil [16] have prompted investigators to conduct time-series analysis on the impact of vaccine introduction [21, 22]. Soon after the introduction of PCV10 in Brazil we evaluated the impact of vaccination on CAP hospitalization trends in children aged <2 years using data from the SIH-SUS for five metropolises in the country; a significant decline in pneumonia hospitalizations 1 year after introduction of PCV10 vaccination was found in three of them [22].

Assessing PCV impact requires observing trends over time. It is interesting to note that the administrative data tracks the hospital primary data for both CAP and other causes of hospitalization. Even when considering SUS and private insurance, peaks of seasonality are very similar between APS and

SIH-SUS, which is a good example of how good impact evaluations can be made with administrative data.

In the Brazilian SUS, the government pays hospitals for hospitalized patients by diagnosis-related groups [16]. Payment by diagnosis-related groups is not dependent on length of hospital stay, but rather on the patient's discharge diagnosis. Moreover, diagnosis-related groups' reimbursement method simplifies the payment process, promotes administrative efficacy, and improves equity. Of the discordant diagnoses in both database, 90% had a CAP diagnosis in SIH-SUS and another respiratory tract disease in APS. The reason why CAP diagnoses were preferred in SIH-SUS may be related to the fact that information on diagnosis-related groups in SIH-SUS records are used to determine the amount to be reimbursed by the government to the hospitals for the services provided. Studies on reliability of administrative data, especially the SIH-SUS database, have showed that diagnostics that allow higher payment are favoured in the database as opposed to other diagnosis-related groups which are reimbursed at a lower level [26, 29–32]. Thus, since pneumonia is the most well paid respiratory event per admission by SUS [16, 33], we cannot rule out over-reporting for CAP diagnosis. Despite these limitations, data quality of the SIH-SUS database is assuming increasing importance on epidemiological research. However, it is important to monitor and incorporate data quality continuously over time, especially in studies on vaccine impact evaluation.

Ensuring data quality of hospitalization information systems is essential for the analysis of the healthcare situation, especially for vaccine impact evaluation purposes. Having an ongoing population-based study on CAP in the municipality of Goiânia represented a unique opportunity to evaluate the reliability of the administrative data system through the comparison between hospital primary (APS) and secondary (SIH-SUS) databases. Nevertheless, some limitations of this investigation should be mentioned. The SIH database was extracted only 3 months after December 2012 (our endpoint); in this way, it was possible to have some notification delays on CAP, since there was a fall in the number of cases included in the last 2 months. Hence, the current CAP cases computed by the SIH could be underestimated. On the other hand, duplicated records in administrative databases are likely to occur, especially due to input errors [34]; we found 10% of duplicated records (exactly the same record) in the SIH database.

Although administrative data slightly overestimated the burden of CAP, we did not observe a clear trend of increase, nor a reduction in pneumonia overestimation during a period of 1 year of the investigation.

In conclusion, administrative data proved to be a feasible source of data to estimate the burden of CAP in infants. Considering the low cost of using SIH-SUS, similar studies in other municipalities of the country should be encouraged to confirm the appropriateness of using administrative databases in time-series analysis to evaluate the impact of interventions on CAP hospitalizations in infants.

## SUPPLEMENTARY MATERIAL

For supplementary material accompanying this paper visit <http://dx.doi.org/10.1017/S0950268814000922>.

## ACKNOWLEDGEMENTS

We thank Tatiana Sugita for revising the ICD-10 codes of all-cause hospitalizations. This investigation was supported by GlaxoSmithKline. A.L.A. (grant nos. 306096/2010-2) is a Fellow of the National Council for Scientific and Technological Development/CNPq. S.S. receives a scholarship from Foundation of Research Support of Goiás State/FAPEG.

## DECLARATION OF INTEREST

A.L.A. has received research and travel grants from GlaxoSmithKline and Pfizer.

## REFERENCES

1. **Scott JA.** The preventable burden of pneumococcal disease in the developing world. *Vaccine* 2007; **25**: 2398–2405.
2. **Nair H, et al.** Global and regional burden of hospital admissions for severe acute lower respiratory infections in young children in 2010: a systematic analysis. *Lancet* 2013; **381**: 1380–1390.
3. **Rudan I, et al.** Global estimate of the incidence of clinical pneumonia among children under five years of age. *Bulletin of the World Health Organization* 2004; **82**: 895–903.
4. **Mulholland K.** Magnitude of the problem of childhood pneumonia. *Lancet* 1999; **354**: 590–592.
5. **Andrade AL, et al.** Bacteriology of community-acquired invasive disease found in a multicountry prospective, population-based, epidemiologic surveillance for *Pneumococcus* in children in Latin America. *Pediatric Infectious Disease Journal* 2012; **31**: 1312–1314.



6. **Ministry of Health (Brazil).** DATASUS. Health Information. Epidemiological and Morbidity, 2013 (<http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sih/cnv/nruf.def>). Accessed 23 October 2013.
7. **Andrade AL, et al.** Population-based surveillance for invasive pneumococcal disease and pneumonia in infants and young children in Goiania, Brazil. *Vaccine* 2012; **30**: 1901–1909.
8. **Ministry of Health (Brazil).** Secretariat of Health Surveillance, Department of Epidemiological Surveillance, General Coordination of the National Immunization Program. Proposal for introduction of 10-valent pneumococcal vaccine into the National Immunization Program: March 2010, 2010 ([http://portal.saude.gov.br/portal/arquivos/pdf/intro\\_pneumococica10\\_val\\_04\\_02\\_10\\_ver\\_final.pdf](http://portal.saude.gov.br/portal/arquivos/pdf/intro_pneumococica10_val_04_02_10_ver_final.pdf)). Accessed 20 July 2013.
9. **Zhou F, et al.** Health care utilization for pneumonia in young children after routine pneumococcal conjugate vaccine use in the United States. *Archives of Pediatrics & Adolescent Medicine* 2007; **161**: 1162–1168.
10. **Griffin MR, et al.** U.S. hospitalizations for pneumonia after a decade of pneumococcal vaccination. *New England Journal of Medicine* 2013; **369**: 155–163.
11. **Ministry of Health (Brazil).** DATASUS. Health Information. Demographic and Socioeconomic; 2012 (<http://tabnet.datasus.gov.br/cgi/deftohtm.exe?ibge/cnv/popuf.def>). Accessed 23 October 2013.
12. **Ministry of Health (Brazil).** DATASUS. Health Information. Vital Statistics, 2011 (<http://www2.datasus.gov.br/DATASUS/index.php?area=0205&VObj=http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sim/cnv/inf10>). Accessed 23 July 2013.
13. **World Health Organization.** International Statistical Classification of Diseases and Related Health Problems, 10th Revision; 2010 ([http://www.who.int/classifications/icd/ICD10Volume2\\_en\\_2010.pdf](http://www.who.int/classifications/icd/ICD10Volume2_en_2010.pdf)). Accessed 15 February 2013.
14. **Bittencourt SA, Camacho LA, Leal Mdo C.** Hospital information systems and their application in public health. *Cadernos de Saude Publica* 2006; **22**: 19–30.
15. **Paim J, et al.** The Brazilian health system: history, advances, and challenges. *Lancet* 2011; **377**: 1778–1797.
16. **Ministry of Health (Brazil), Pan American Health Organization, Oswaldo Cruz Foundation.** *The Brazilian experience in health information systems. Talking about health information systems in Brazil.* Brasília, DF: The Ministry of Health's Publisher, 2009, Report No. 2.
17. **Ministry of Health (Brazil), Pan American Health Organization, Oswaldo Cruz Foundation.** *The Brazilian experience in health information systems. Production and propagation of health information in Brazil.* Brasília, DF: The Ministry of Health's Publisher, 2009, Report No. 1.
18. **Centers for Disease Control and Prevention.** Link Plus v. 2. Atlanta: CDC, 2007.
19. **Friedl JEF.** *Mastering Regular Expressions*, 3rd edn. Sebastopol, CA: O'Reilly Media, 2006.
20. **Instituto Brasileiro de Geografia e Estatística.** Population census, 2010 (<http://www.ibge.gov.br/home/estatistica/populacao/estimativa2012/default.shtm>). Accessed 23 October 2013.
21. **do Carmo GM, et al.** Decline in diarrhea mortality and admissions after routine childhood rotavirus immunization in Brazil: a time-series analysis. *PLoS Medicine* 2011; **8**: e1001024.
22. **Afonso ET, et al.** Effect of 10-valent pneumococcal vaccine on pneumonia among children, Brazil. *Emerging Infectious Diseases* 2013; **19**: 589–597.
23. **Grijalva CG, et al.** Decline in pneumonia admissions after routine childhood immunisation with pneumococcal conjugate vaccine in the USA: a time-series analysis. *Lancet* 2007; **369**: 1179–1186.
24. **De Wals P, et al.** Pneumonia after implementation of the pneumococcal conjugate vaccine program in the province of Quebec, Canada. *Pediatric Infectious Disease Journal* 2008; **27**: 963–968.
25. **Yu O, et al.** Classification algorithms to improve the accuracy of identifying patients hospitalized with community-acquired pneumonia using administrative data. *Epidemiology and Infection* 2011; **139**: 1296–1306.
26. **Veras CM, Martins MS.** Reliability of data from authorization forms for hospital admittance, Rio de Janeiro, Brazil. *Cadernos de Saude Publica* 1994; **10**: 339–555.
27. **Mathias TA, Soboll ML.** Reliability of diagnoses on authorization forms for hospital admission. *Revista de Saude Publica* 1998; **32**: 526–532.
28. **Escosteguy CC, et al.** Hospital admissions forms versus medical records to assess risk of in-hospital death from acute myocardial infarction in Rio de Janeiro, Brazil. *Cadernos de Saude Publica* 2005; **21**: 1065–1076.
29. **Simborg DW.** DRG creep: a new hospital-acquired disease. *New England Journal of Medicine* 1981; **304**: 1602–1604.
30. **Berezin EN, et al.** Pneumonia hospitalization in Brazil from 2003 to 2007. *International Journal of Infectious Diseases* 2012; **16**: e583–590.
31. **Noronha MF, et al.** The development of diagnosis-related groups–DRGs. Methodology for classifying hospital patients. *Revista de Saude Publica* 1991; **25**: 198–208.
32. **Hsia DC, et al.** Accuracy of diagnostic coding for Medicare patients under the prospective-payment system. *New England Journal of Medicine* 1988; **318**: 352–355.
33. **Ministry of Health (Brazil).** Management System of Procedures, Medicines and OPM of SUS–SIGTAP, 2013 (<http://sigtap.datasus.gov.br>). Accessed 8 October 2013.
34. **Carvalho CN, Dourado I, Bierrenbach AL.** Under-reporting of the tuberculosis and AIDS comorbidity: an application of the linkage method. *Revista de Saude Publica* 2011; **45**: 548–555.