Identifying Pediatric Community-Acquired Pneumonia Hospitalizations
Accuracy of Administrative Billing Codes

Derek J. Williams, MD, MPH; Samir S. Shah, MD, MSCE; Angela Myers, MD, MPH; Matthew Hall, PhD; Katherine Auger, MD; Mary Ann Queen, MD; Karen E. Jerardi, MD, MEd; Lauren McClain, MD; Catherine Wiggleton, MD, ScM; Joel S. Tieder, MD, MPH

IMPORTANCE Community-acquired pneumonia (CAP) remains one of the most common indications for pediatric hospitalization in the United States, and it is frequently the focus of research and quality studies. Use of administrative data is increasingly common for these purposes, although proper validation is required to ensure valid study conclusions.

OBJECTIVE To validate administrative billing data for hospitalizations owing to childhood CAP.

DESIGN AND SETTING Case-control study of 4 tertiary care, freestanding children's hospitals in the United States.

PARTICIPANTS A total of 998 medical records of a 25% random sample of 3646 children discharged in 2010 with at least 1 International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) code representing possible pneumonia were reviewed. Discharges (matched on date of admission) without a pneumonia-related discharge code were also examined to identify potential missed pneumonia cases. Two reference standards, based on provider diagnosis alone (provider confirmed) or in combination with consistent clinical and radiographic evidence of pneumonia (definite), were used to identify CAP.

EXPOSURE Twelve ICD-9-CM-based coding strategies, each using a combination of primary or secondary codes representing pneumonia or pneumonia-related complications. Six algorithms excluded children with complex chronic conditions.

MAIN OUTCOMES AND MEASURES Sensitivity, specificity, and negative and positive predictive values (NPV and PPV, respectively) of the 12 identification strategies.

RESULTS For provider-confirmed CAP (n = 680), sensitivity ranged from 60.7% to 99.7%; specificity, 75.7% to 96.4%; PPV, 67.9% to 89.6%; and NPV, 82.6% to 99.8%. For definite CAP (n = 547), sensitivity ranged from 65.6% to 99.6%; specificity, 68.7% to 93.0%; PPV, 54.6% to 77.9%; and NPV, 87.8% to 99.8%. Unrestricted use of the pneumonia-related codes was inaccurate, although several strategies improved specificity to more than 90% with a variable effect on sensitivity. Excluding children with complex chronic conditions demonstrated the most favorable performance characteristics. Performance of the algorithms was similar across institutions.

CONCLUSIONS AND RELEVANCE Administrative data are valuable for studying pediatric CAP hospitalizations. The strategies presented here will aid in the accurate identification of relevant and comparable patient populations for research and performance improvement studies.
Community-acquired pneumonia (CAP) is one of the most common causes of childhood mortality worldwide, especially among those younger than 5 years. Although mortality is lower in developed countries, pediatric CAP is still associated with substantial morbidity and remains the most common indication for pediatric hospitalization outside the newborn period in the United States. It is also one of the most expensive, with hospital costs alone approaching $1 billion annually. As a result, pediatric CAP is often the focus of epidemiologic studies and outcomes research and serves as an ideal target for quality benchmarking and improvement efforts.

Administrative data are often used for such assessments because they are convenient and provide ready access to large study populations without the substantially higher costs of prospective studies. However, studies using administrative data often rely on discharge diagnosis codes, such as the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM), to identify study populations, and the assignment of these codes may be subject to error or influenced by reimbursement patterns and local coding practices. In the case of pneumonia, many ICD-9-CM discharge codes represent its many microbiologic causes (eg, influenza vs Streptococcus pneumoniae) and potential complications (eg, empyema). Furthermore, a discharge code indicating pneumonia does not readily differentiate CAP from pneumonia that occurs from a hospital-acquired infection, such as ventilator-associated pneumonia. As a result, use of administrative data to identify CAP hospitalizations without appropriate validation is not advisable because of the heterogeneous nature of the disease and the potential for misclassification and erroneous conclusions.

Studies in adults have assessed the validity of ICD-9-CM codes to identify hospitalizations owing to pneumonia, although with varying results depending on the population studied and the codes selected for identification (positive predictive value [PPV], 57%-96%; sensitivity, 48%-98%). The findings from these studies may be limited in their application since they relied on small sample sizes, were conducted at a single site, or were completed in health care systems outside the United States. Most important, none of these studies included children, and thus the results may not be readily generalizable to pediatric populations. Consequently, the primary objective of our study was to assess the performance of a variety of ICD-9-CM coding strategies to identify pediatric CAP hospitalizations using data from 4 tertiary care, freestanding children’s hospitals in the United States.

Methods

Design and Setting

For this multicenter retrospective study, the Pediatric Health Information System database from the Children’s Hospital Association (Overland Park, Kansas) was used to identify children from 4 tertiary care, freestanding children’s hospitals (The Monroe Carell Jr Children’s Hospital at Vanderbilt, Nashville, Tennessee; Children’s Mercy Hospitals & Clinics, Kansas City, Missouri; Seattle Children’s Hospital, Seattle, Washington; and Cincinnati Children’s Hospital Medical Center, Cincinnati, Ohio). The institutional review board at each hospital approved the study.

Study Population

A total of 5023 children aged 60 days to 18 years who were discharged from one of the participating hospitals between January 1, 2010, and December 31, 2010, with at least 1 ICD-9-CM discharge diagnosis code (primary or any secondary) indicating pneumonia, pleural effusion, or empyema were considered for inclusion. Two investigators (D.J.W. and J.S.T.) independently reviewed concomitant ICD-9-CM codes for each hospitalization (up to 21 codes), excluding 1377 patients with diagnoses that precluded CAP (eg, cystic fibrosis or immunodeficiency). A total of 998 medical records of a 25% random sample of the remaining 3646 discharges was then selected for review. To identify potential missed pneumonia cases, we also selected 1000 discharges without an ICD-9-CM pneumonia-related diagnosis code (matched by date of admission to account for seasonal trends in pneumonia admissions) for medical record review.

Data Sources

Two data sources were used for this study: the Pediatric Health Information System database and medical record review. The Pediatric Health Information System database, which contains clinical and billing data from 42 tertiary care, freestanding children’s hospitals, was used to identify all potential participants and define patient demographics. Data quality was ensured through a joint effort between the Children’s Hospital Association and participating hospitals, as described previously. Medical record data were extracted by trained investigators (D.J.W., A.M., K.A., M.A.Q., K.J., L.M., and C.W.) at each site using a central, web-based data collection system.

Reference Standards and ICD-9-CM CAP Identification Strategies

Two reference standards were used to represent CAP, one highly sensitive (provider diagnosis) and the other highly specific (provider diagnosis plus consistent clinical and radiographic evidence). We assessed each medical record for the presence of the following criteria: (1) provider diagnosis of pneumonia within the first 48 hours of hospitalization (mention of suspected CAP along with consistent management strategy), (2) abnormal temperature (≥38.0°C) or white blood cell count (<5 or >15 μL [to convert to ×10⁹/L, multiply by 0.001]), (3) evidence of a lower respiratory tract illness (eg, cough or increased work of breathing), and (4) chest radiograph indicating pneumonia (eg, infiltrate or consolidation). Any child with a condition precluding CAP (eg, cystic fibrosis or effusions following admission for cardiothoracic surgery) was not classified as having CAP regardless of the above-mentioned criteria. After applying this exclusion, we considered children with at least a provider diagnosis of pneumonia to have provider-confirmed CAP and those with all 4 criteria to have definite CAP.

For each reference standard, we assessed the performance of 12 ICD-9-CM coding algorithms to identify CAP hospitalizations among our study population. These strategies incorpo-
rate a number of previously described CAP algorithms as well as the classification of complex chronic conditions (CCCs) described by Feudtner et al. Strategies that included codes representative of pneumonia symptoms (eg, cough) were also considered; however, these algorithms did not identify additional CAP cases and were excluded from further analyses.

### Statistical Analysis

Characteristics of the study population were summarized using proportions for categorical variables and median and interquartile range for continuous variables. Characteristics of children classified as having definite or provider-confirmed CAP were compared using \( \chi^2 \) tests for categorical variables and Wilcoxon rank sum for continuous variables.

To determine the performance of each identification strategy for CAP, we calculated the sensitivity, specificity, PPV, and negative predictive value (NPV) for each reference standard. We randomly sampled from all discharges with a pneumonia-related discharge code; thus, the PPV estimates are assumed to reflect the entire population under study, including disease prevalence within the population. To maximize the possibility of identifying false negatives (algorithm negative and reference standard positive), we selected a random sample of hospitalizations without a pneumonia-related discharge code matched on admission date to those hospitalizations with a pneumonia-related discharge code.

To describe interhospital variation in coding practices for CAP, we characterized the performance of each algorithm for provider-confirmed CAP for each hospital. To assess differences in patient characteristics and outcomes across identification strategies for children with provider-confirmed CAP, we summarized several patient-level variables, including concurrent lower respiratory tract diagnoses, care use measures, and severe outcomes (intensive care admission, mechanical ventilation, or empyema).

### Results

#### Study Population

We reviewed 998 medical records of a possible 3646 discharges with a pneumonia-related ICD-9-CM discharge code and 1000 discharges without a pneumonia-related discharge code (Figure 1). Among discharges with a pneumonia-related code, 677 (67.8%) were classified as provider-confirmed CAP; 545 (80.5%) were further classified as definite CAP. Children with definite CAP were less likely to have a concurrent discharge diagnosis of asthma but more likely to have a parapneumonic effusion or empyema compared with those without definite CAP (Table 1). Although hospital length of stay was slightly longer, children with definite CAP were less likely to have a complex chronic condition or require admission to intensive care.
Table 1. Characteristics of the Study Population With Community-Acquired Pneumonia: Comparison of 2 Reference Standards*  

| Characteristic                          | All Provider-Confirmed CAP (n = 680) | Definite CAP (n = 547) | Provider-Confirmed CAP Only (n = 133) | P Value  
|----------------------------------------|-------------------------------------|------------------------|---------------------------------------|---------  
| Age, y                                 |                                     |                        |                                       |          
| <1                                     | 116 (17.1)                          | 91 (16.6)              | 25 (18.8)                             | .31      
| 1-5                                    | 365 (53.7)                          | 302 (55.2)             | 63 (47.4)                             |          
| 6-10                                   | 135 (19.9)                          | 102 (18.6)             | 33 (24.8)                             |          
| ≥11                                    | 64 (9.4)                            | 52 (9.5)               | 12 (9.0)                              |          
| Male sex                               | 369 (54.5)                          | 301 (55.2)             | 68 (51.5)                             | .44      
| Race/ethnicity                         |                                     |                        |                                       |          
| Non-Hispanic white                     | 379 (55.7)                          | 317 (57.9)             | 62 (47.0)                             | .001     
| Non-Hispanic black                     | 118 (17.3)                          | 78 (14.3)              | 39 (29.5)                             |          
| Hispanic                               | 65 (9.6)                            | 55 (10.1)              | 10 (7.6)                              |          
| Asian                                  | 19 (2.8)                            | 17 (3.1)               | 2 (1.5)                               |          
| Other                                  | 99 (14.6)                           | 80 (14.7)              | 19 (14.4)                             |          
| Insurance                              |                                     |                        |                                       |          
| Public                                 | 363 (53.4)                          | 282 (51.6)             | 81 (60.9)                             | .16      
| Private                                | 316 (46.5)                          | 263 (48.3)             | 52 (39.4)                             |         1  
| Other                                  | 1 (0.1)                             | 1 (0.2)                | 0                                     |         2  
| Complex chronic condition*             | 62 (9.1)                            | 41 (7.5)               | 21 (15.8)                             | .003    3  
| Intensive care admission               | 75 (11.1)                           | 51 (9.4)               | 24 (18.0)                             | .004    4  
| Mechanical ventilation                 | 28 (4.1)                            | 24 (4.4)               | 4 (3.0)                               | .47     5  
| Empyema                                | 102 (15.0)                          | 94 (17.2)              | 8 (6.0)                               | .001    6  
| Concurrent diagnosis                   |                                     |                        |                                       |         7  
| Asthma                                 | 219 (32.2)                          | 159 (29.1)             | 60 (45.1)                             | <.001   8  
| Bronchiolitis                          | 53 (7.8)                            | 46 (8.4)               | 7 (5.3)                               | .23     9  
| Hospital length of stay, median (IQR), d | 2 (1-4)                             | 2 (2-4)                | 2 (1-3)                               | <.001   10  
| Hospital charges, median (IQR), $      | $11 755 (8000-19 840)               | $11 645 (8000-19 647) | $12 700 (8024-20 026)                 | <.001   11  

**Abbreviations:** CAP, community-acquired pneumonia; IQR, interquartile range.  
* Provider-confirmed CAP includes discharges with a provider diagnosis of pneumonia within 48 hours of hospitalization and no exclusionary conditions (n = 680). Definite CAP includes discharges with a provider diagnosis of pneumonia within 48 hours of hospitalization, abnormal temperature or white blood cell count, evidence of a lower respiratory tract illness, radiographic evidence of pneumonia, and no exclusionary conditions (n = 547).  
1 Statistical comparisons are displayed between those classified as provider-confirmed CAP only (n = 133) and those classified as definite CAP (n = 547).  
2 See Feudtner et al.  
3 Statistical comparisons are displayed between those classified as provider-confirmed CAP only (n = 133) and those classified as definite CAP (n = 547).  
4 Provider-confirmed CAP includes discharges with a provider diagnosis of pneumonia within 48 hours of hospitalization and no exclusionary conditions (n = 680). Definite CAP includes discharges with a provider diagnosis of pneumonia within 48 hours of hospitalization, abnormal temperature or white blood cell count, evidence of a lower respiratory tract illness, radiographic evidence of pneumonia, and no exclusionary conditions (n = 547).  
5 Statistical comparisons are displayed between those classified as provider-confirmed CAP only (n = 133) and those classified as definite CAP (n = 547).  
6 See Feudtner et al.
few differences when comparing the individual algorithms across hospitals. Small differences in sensitivity for several algorithms occurred at 1 institution, but no other differences were noted.

Finally, we described patient characteristics, including other lower respiratory tract diagnoses, complications, and outcomes for CAP, for each algorithm to assess potential differences (Table 2). Several differences were noted according to the algorithm selected, including the proportion of children with a concurrent diagnosis of asthma or bronchiolitis and those experiencing a severe outcome (intensive care admission, mechanical ventilation, or empyema).

Discussion
This study determined the performance of 12 ICD-9-CM coding algorithms in identifying pediatric CAP hospitalizations at 4 tertiary care, freestanding children’s hospitals in the United States. Unrestricted application of the pneumonia-related codes does not accurately identify CAP in children, although several coding strategies offer improvements in specificity while also retaining good sensitivity. Administrative data can be used to study pediatric CAP hospitalizations, although the study population must be carefully defined and the specificity of the diagnosis understood to ensure valid study conclusions.

Use of the pneumonia and effusion or empyema codes in any position (algorithm 1) was not accurate, although this strategy identified nearly all children with CAP. This is an important attribute for studies aimed at identifying all episodes of CAP, such as assessments of disease frequency or monitoring for adverse drug effects. To avoid misclassification, application of this algorithm requires a multistep identification strategy, such as ICD-9-CM-based screening with medical record review to confirm the diagnosis.

In contrast to algorithms 1 through 3, which allowed the effusion or empyema codes to represent pneumonia without the requirement of an additional pneumonia code, algorithms 4 through 6 required a more explicit pneumonia code, with only algorithm 6 including the effusion or empyema codes (as a pneumonia-related complication). This resulted in an 11% increase in specificity for pneumonia codes in any position (algorithm 4) and identified nearly all CAP cases. Predictive performance was further improved when restricting the population to those with a primary diagnosis of pneumonia (algorithm 5) or a primary diagnosis of a pneumonia-related complication plus a secondary diagnosis of pneumonia (algorithm 6), similar to validation studies in other populations.22,23 However, these algorithms identified only approximately 70% of provider-confirmed CAP, and specificity was only marginally improved over the analogous algorithms 2 and 3. Nonetheless, these identification strategies may be a better choice for assessments that seek to maximize specificity over sensitivity and are unable to include confirmation using medical records. For example, these algorithms could be used to improve the validity of outcomes data in a study comparing the effectiveness of antimicrobial treatment choices for CAP.

The algorithms with the best performance characteristics (ie, both highly sensitive and specific) were 1b and 4b. These identification strategies, which do not restrict use of the pneumonia or empyema codes in any position, were similar in performance to those using the additional pneumonia code (algorithm 6).

The algorithms that utilize a primary diagnosis of pneumonia in any position (algorithm 4) generally offer higher sensitivity, with the potential for improved specificity over the algorithms that require the additional pneumonia code. The algorithm that required the most explicit identification of pneumonia was the most specific for CAP, but also identified only approximately 70% of CAP cases. These algorithms may be useful for studies seeking to improve the validity of CAP outcomes data, but they may not be as appropriate for identifying all episodes of pneumonia-related complications with CAP.

The use of ICD-9-CM codes for pneumonia and related conditions in pediatric hospital settings is a common practice, but the accuracy of these codes in identifying pneumonia and related complications can vary widely. This study demonstrates the importance of using multiple identification strategies to improve the specificity and sensitivity of pneumonia codes in administrative data, as well as the need for validation studies to ensure accurate identification of CAP.

Figure 2. Sensitivity and Specificity of 12 ICD-9-CM Code Identification Algorithms for CAP Using 2 Reference Standards

Provider-confirmed CAP (n = 680) (A) and definite CAP (n = 547) (B). Horizontal and vertical bars represent calculated 95% confidence intervals. CAP indicates community-acquired pneumonia; ICD-9-CM, International Classification of Diseases, 9th Revision, Clinical Modification. Algorithm definitions: 1 = primary or any secondary diagnosis of pneumonia or effusion/empyema (1b excludes complex chronic conditions [CCCs]); 2 = primary diagnosis of pneumonia or effusion/empyema (2b excludes CCCs); 3 = primary diagnosis of pneumonia or effusion/empyema or primary diagnosis of pneumonia-related complication plus any secondary diagnosis of pneumonia or effusion/empyema (3b excludes CCCs); 4 = primary diagnosis of pneumonia (algorithm 4b excludes CCCs); 5 = primary diagnosis of pneumonia (5b excludes CCCs); and 6 = primary diagnosis of pneumonia or primary diagnosis of pneumonia-related complication or effusion/empyema plus any secondary diagnosis of pneumonia (6b excludes CCCs). The ICD-9-CM codes used in the study are as follows: pneumonia, 480.0 to 480.2, 480.8 to 480.9, 481, 482.0, 482.10 to 482.12, 482.41 to 482.42, 482.83, 482.9 to 482.90, 483.8, 484.3, 485, 486, and 4870; effusion/empyema, 510.0, 510.9, 511.0 to 511.1, 511.8 to 511.9, and 513; and pneumonia-related complication, 38.9, 458.9, 518.81, 790.7, 799.3, 995.91 to 995.92, and 997.3.
monia codes, excluded children with CCCs identified with a previously reported classification scheme. The sensitivity and specificity of both algorithms were approximately 90% for provider-confirmed CAP. Overall, application of the CCC restriction yielded meaningful improvements in specificity for all algorithms studied without large reductions in sensitivity. Likely, restriction of these codes excludes children with a high probability of complicating factors, which may alter the true risk of CAP (eg, frequent hospitalizations or technology dependence) while minimally affecting children without these risk factors. As a result, these strategies are ideal for identifying and studying pediatric CAP hospitalizations, especially if the presence of complex comorbid conditions is thought to confound hypothesized relationships.

Requiring the CAP reference standard definition to include a provider diagnosis of pneumonia coupled with objective clinical evidence of a lower respiratory tract infection with radiographic confirmation (ie, definite CAP) resulted in expected reductions in specificity for all algorithms studied. Several differences were noted between patients classified as having definite CAP and those having provider-confirmed CAP only, underscoring the clinical challenge of diagnosing CAP in children. For instance, nearly half of those with provider-confirmed CAP who were not also classified as having definite CAP had a concurrent discharge diagnosis of asthma. Although asthma is a risk factor for CAP, in the clinical setting, it is often difficult to distinguish a superimposed pneumonia from atelectasis in an acutely wheezing asthmatic child when a chest radiograph reveals an opacity. In contrast, the presence of a parapneumonic effusion in a child with presumed pneumonia offers additional certainty of the CAP diagnosis. Indeed, the presence of effusion or empyema was more common among those diagnosed with definite CAP. Thus, the clinical uncertainty inherent in diagnosing CAP is reflected in the differences in predictive performance for the 2 reference standards presented here. This provides a more objective assessment of the degree of uncertainty likely to exist in a population of children with CAP identified using administrative data. To this end, it is important to note that 80.4% of the children with provider-confirmed CAP were also classified as having definite CAP.

Coding variation across institutions may limit the validity of administrative data. In our study, specificity across all 4 institutions and sensitivity at 3 of the 4 institutions studied were similar. Sensitivity was lower for several algorithms at 1 institution, although differences were small and likely reflect local variation in coding practices. Nonetheless, our findings demonstrate that coding practices for pediatric CAP are generally consistent across hospitals, suggesting that these identification strategies can be used to accurately benchmark performance and compare patients across institutions.

We also observed that identification strategies that considered either a primary or a secondary diagnosis of pneumonia (algorithms 1, 1b, 4, and 4b) identified a higher proportion of children with a concurrent diagnosis of asthma or bronchiolitis. In contrast, algorithms restricted to a primary diagnosis of pneumonia (algorithms 5 and 5b) included fewer children with these diagnoses but also identified the lowest proportion of children requiring intensive care unit admission or mechanical ventilation. This is an important consideration when designing studies to examine severe pneumonia outcomes. Moreover, strategies that considered a secondary diagnosis of pneumonia only if it was coupled with a primary diagnosis of a pneumonia-related complication (algorithms 6 and 6b) identified a high proportion of children with severe outcomes while minimizing those with concurrent diagnoses of asthma or bronchiolitis. This suggests that these latter algorithms may be particularly relevant for identifying severe CAP while also enriching the study population for bacterial pneumonia.

Strengths of our study include the multicenter design and large sample of children, the use of standardized data collection methods and CAP reference standards, a wide range of CAP identification algorithms, and the exclusive focus on pediat-

### Table 2. Outcomes and Concurrent Diagnoses for Provider-Confirmed Community-Acquired Pneumonia According to Identification Algorithm

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Asthma (%)</th>
<th>Bronchiolitis (%)</th>
<th>Intensive Care Admission (%)</th>
<th>Mechanical Ventilation (%)</th>
<th>Empyema (%)</th>
<th>Hospital Length of Stay, Median (IQR), d</th>
<th>Hospital Charges, Median (IQR), $</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>212 (31.9)</td>
<td>53 (8.0)</td>
<td>71 (10.7)</td>
<td>28 (4.2)</td>
<td>102 (15.3)</td>
<td>2 (1-4)</td>
<td>11 662 (8000-19 858)</td>
</tr>
<tr>
<td>1b</td>
<td>200 (33.1)</td>
<td>50 (8.3)</td>
<td>57 (9.4)</td>
<td>21 (3.5)</td>
<td>93 (15.4)</td>
<td>2 (1-3)</td>
<td>11 373 (7900-18 990)</td>
</tr>
<tr>
<td>2</td>
<td>120 (25.1)</td>
<td>12 (2.5)</td>
<td>26 (5.4)</td>
<td>9 (1.9)</td>
<td>88 (18.4)</td>
<td>2 (1-3)</td>
<td>10 608 (7648-17 169)</td>
</tr>
<tr>
<td>2b</td>
<td>111 (25.8)</td>
<td>11 (2.6)</td>
<td>19 (4.4)</td>
<td>6 (1.4)</td>
<td>81 (18.8)</td>
<td>2 (1-3)</td>
<td>10 246 (7583-15 885)</td>
</tr>
<tr>
<td>3</td>
<td>122 (24.9)</td>
<td>13 (2.7)</td>
<td>35 (7.2)</td>
<td>18 (3.7)</td>
<td>91 (18.6)</td>
<td>2 (1-3)</td>
<td>10 784 (7679-17 782)</td>
</tr>
<tr>
<td>3b</td>
<td>113 (25.9)</td>
<td>12 (2.7)</td>
<td>26 (5.9)</td>
<td>13 (3.0)</td>
<td>83 (19.0)</td>
<td>2 (1-3)</td>
<td>10 324 (7612-16 346)</td>
</tr>
<tr>
<td>4</td>
<td>212 (32.3)</td>
<td>53 (8.1)</td>
<td>70 (10.7)</td>
<td>28 (4.3)</td>
<td>93 (14.2)</td>
<td>2 (1-4)</td>
<td>11 523 (7931-19 551)</td>
</tr>
<tr>
<td>4b</td>
<td>200 (33.6)</td>
<td>50 (8.4)</td>
<td>56 (9.4)</td>
<td>21 (3.5)</td>
<td>85 (14.3)</td>
<td>2 (1-3)</td>
<td>11 217 (7870-18 413)</td>
</tr>
<tr>
<td>5</td>
<td>118 (25.9)</td>
<td>12 (2.6)</td>
<td>23 (5.0)</td>
<td>8 (1.8)</td>
<td>65 (14.3)</td>
<td>2 (1-3)</td>
<td>10 247 (7538-15 713)</td>
</tr>
<tr>
<td>5b</td>
<td>109 (26.7)</td>
<td>11 (2.7)</td>
<td>16 (3.9)</td>
<td>5 (1.2)</td>
<td>60 (14.7)</td>
<td>2 (1-3)</td>
<td>10 073 (7522-15 081)</td>
</tr>
<tr>
<td>6</td>
<td>122 (25.3)</td>
<td>13 (2.7)</td>
<td>35 (7.3)</td>
<td>18 (3.7)</td>
<td>84 (17.4)</td>
<td>2 (1-3)</td>
<td>10 611 (7653-17 338)</td>
</tr>
<tr>
<td>6b</td>
<td>113 (26.2)</td>
<td>12 (2.8)</td>
<td>26 (6.0)</td>
<td>13 (3.0)</td>
<td>77 (17.9)</td>
<td>2 (1-3)</td>
<td>10 276 (7583-16 100)</td>
</tr>
</tbody>
</table>

Abbreviation: IQR, interquartile range.

* For algorithm definitions, see the caption to Figure 2.
ric patients hospitalized with CAP, a population that, to our knowledge, has not been validated using administrative data. There are also several limitations. Discharges with and without a pneumonia-related discharge code were reviewed separately, although reviewers were masked to individual billing codes. A comparison of rater agreement also was not performed. There is no universally accepted reference standard for defining CAP, and thus, as with all studies of pneumonia, there is at least some level of uncertainty in classifying the disease. 30 We attempted to minimize misclassification by using 2 reference standards that rely on provider assessment and objectively measured criteria, investigator training and piloting of data collection procedures, and use of a common web-based data collection system. With more than 40,000 discharges in 2010 among the 4 study hospitals, it was not feasible to review all hospitalizations during the study period. However, we selected a random sample of discharges using a wide range of possible pneumonia-related diagnosis codes and reviewed a similar number of discharges without a pneumonia-related diagnosis code. Our matching strategy also maximizes the possibility of identifying missed pneumonia cases, resulting in an accurate estimation of PPV and a most conservative estimate of NPV. However, it is likely that this strategy overestimates sensitivity and underestimates specificity. As an example, applying the PPV and NPV estimates from algorithm 1 for provider-confirmed CAP to the entire study population (using actual numbers of algorithm-positive and total hospital discharges in 2010) results in an estimated sensitivity of 96.1% (reported 99.7%) and specificity of 97.7% (reported 75.7%). Although this exercise is an oversimplification, the true sensitivity and specificity can be expected to fall within this range. Finally, since this study was conducted at tertiary care, freestanding children’s hospitals, our results may not be generalizable to community hospitals. Nonetheless, Children’s Hospital Association-affiliated hospitals account for nearly 20% of pediatric hospitalizations each year, making the Pediatric Health Information System one of the largest and most frequently used administrative data sources for quality and research studies of pediatric hospitalizations in the United States. Moreover, given the high degree of agreement between hospitals, it is likely that our results would be applicable to other pediatric administrative data sets.

In conclusion, to our knowledge, this is the first study to validate ICD-9-CM discharge diagnosis codes for children hospitalized with CAP. We have demonstrated that administrative data are a valuable tool for studying pediatric CAP and provide several strategies that reliably identify this population. Application of these validated algorithms allows for a more accurate identification of relevant and comparable patient populations for research and performance improvement purposes. Understanding the strengths and limitations of these data, as well as the uncertainty associated with diagnosing CAP, will help ensure valid study conclusions.

ARTICLE INFORMATION
Accepted for Publication: January 17, 2013.
Published Online: July 29, 2013.

Author Affiliations:
Division of Hospital Medicine, The Monroe Carell Jr Children’s Hospital at Vanderbilt and Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, Tennessee (Williams, McClain, Wigginton); Division of Infectious Diseases, Cincinnati Children’s Hospital Medical Center and Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio (Shah, Jerardi); Division of Hospital Medicine, Cincinnati Children’s Hospital Medical Center and Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, Ohio (Shah, Jerardi); Division of Infectious Diseases, Children’s Mercy Hospitals & Clinics and University of Missouri–Kansas City School of Medicine, Kansas City, Missouri (Myers); Children’s Hospital Association, Overland Park, Kansas (Hall); Division of General Pediatrics, University of Michigan, Ann Arbor (Auger); Division of Hospital Medicine, Children’s Mercy Hospital and Clinics and University of Missouri School of Medicine, Kansas City, Missouri (Queen); Division of Hospital Medicine, Seattle Children’s Hospital and the Department of Pediatrics, University of Washington School of Medicine, Seattle, Washington (Tieder).

Author Contributions: Drs Williams, Shah, Myers, Hall, Auger, Jerardi, Tieder. Acquisition of data: Myers, Hall, Queen, Jerardi, McClain, Wigginton, Tieder. Analysis and interpretation of data: Williams, Shah, Myers, Hall, McClain, Wigginton, Tieder. Drafting of the manuscript: Williams, Myers, Hall, McClain, Tieder. Critical revision of the manuscript for important intellectual content: Williams, Shah, Auger, Queen, Jerardi, Tieder. Statistical analysis: Myers, Hall, McClain, Wigginton. Administrative, technical, and material support: Williams, Jerardi, Tieder. Study supervision: Williams, Shah, Tieder.

Conflict of Interest Disclosures: None reported.

Funding/Sponsorship: This study was supported by grant KL2 RR24977 from the National Institutes of Health through the Vanderbilt Clinical and Translational Research Scholars Program (Dr Williams) and the Robert Wood Johnson Foundation under its Clinical Scholars Program (Dr Auger).

Additional Contributions: Ross Newman, MD, Jennifer Soper, MD, Angela Statille, MD, and Connie Yau, BA, assisted with data collection.

REFERENCES


