Use of Administrative Data for the Identification of Laboratory-Confirmed Influenza Infection: The Validity of Influenza-Specific ICD-9 Codes

Kristen A. Feemster,1,2 Kateri H. Leckerman,1 Maria Middleton,1 Danielle M. Zerr,3 Alexis M. Elward,4 Jason G. Newland,5 Lindsay Asti,6 Rebecca M. Guth,4 Rangaraj Selvarangan,5 and Susan E. Coffin1,2

1Children’s Hospital of Philadelphia, and 2Perelman School of Medicine at the University of Pennsylvania, Philadelphia; 3Seattle Children’s Hospital, Washington; 4St Louis Children’s Hospital, and 5Children’s Mercy Hospital and Clinics, Kansas City, Missouri; and 6Nationwide Children’s Hospital, Columbus, Ohio

Corresponding Author: Kristen A. Feemster, MD, MPH, MSHP, Assistant Professor, Perelman School of Medicine, University of Pennsylvania, Division of Infectious Diseases, The Children’s Hospital of Philadelphia, CHOP North, 3535 Market St, Rm 1511, Philadelphia, PA 19104. E-mail: feemster@email.chop.edu.

Received November 23, 2011; accepted March 15, 2012; electronically published July 2, 2012.

We used Pediatric Health Information System data and laboratory records from 3 children’s hospitals to determine whether administrative data accurately identify children with laboratory-confirmed influenza. Among 23 282 inpatients, diagnosis codes for influenza detected 73% of laboratory-confirmed influenza cases, whereas <1% of patients without a diagnosis code had laboratory-confirmed influenza.

Key words. Electronic Health Record; Influenza Surveillance; Pediatric Health Information System

Influenza is a significant source of morbidity and mortality among children, resulting in an estimated 3.5 to 12.0/10 000 children hospitalized with laboratory-confirmed influenza each year [1–4]. Although effective strategies to prevent influenza infection exist, high levels of transmission occur annually. The accurate identification of influenza cases is essential to better describe the epidemiology of infection and to evaluate disease mitigation strategies. One method for influenza surveillance includes the use of ICD-9 codes from administrative data. As more clinical care networks adopt the use of electronic health records (EHRs), EHR data can be leveraged to conduct surveillance across large populations with diverse patient, practice, and community characteristics. However, EHR data must be able to accurately and comprehensively capture cases. Studies have shown that discharge ICD-9 codes perform well for syndromic surveillance of influenza-like illness, but these studies have not included pediatric populations or focused on influenza-specific ICD-9 codes [5, 6]. The objective of this multicenter study was to validate influenza-specific ICD-9 codes for the identification of influenza cases among hospitalized children.

METHODS

We conducted a cross-sectional study using administrative data from the Pediatric Health Information System data (PHIS) and laboratory records from 3 US children’s hospitals. PHIS is a comprehensive administrative database that contains clinical and billing information for all pediatric discharges from the 43 Child Health Corporation of America member hospitals. Using PHIS, we identified all hospitalized children who were discharged with an influenza ICD-9 code (487.0: “Influenza with pneumonia”; 487.1: “Influenza with other respiratory manifestations”; and 487.8: “Influenza with other manifestations”) between October 2006 and May 2007. Laboratory data were
reviewed at each hospital to identify all laboratory-confirmed influenza cases during the study period. Laboratory-confirmed cases of influenza included all patients with a positive rapid test, polymerase chain reaction, or culture for an influenza virus.

From the total number of children discharged from the 3 study hospitals within the study period, patients were divided into 4 groups: (1) influenza ICD-9 code present with positive laboratory test (true positive), (2) influenza ICD-9 code present with no laboratory confirmation (false positive), (3) no influenza ICD-9 code with a positive laboratory test (false negative), and (4) no influenza ICD-9 code with no positive laboratory test (true negative). These groups were compared to calculate the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the administrative data for the identification of influenza cases. All statistical analyses were conducted using SAS version 9.2 (SAS Institute Inc). Confidence intervals (CIs) for each value were also calculated to provide a representation of the range of the true proportions that might exist in a larger population.

**Chart Review**

To investigate cases with a discrepancy between laboratory confirmation and assignment of an influenza ICD-9 code, a systematic chart review was performed at one of the participating institutions. If there was documentation of a positive influenza laboratory test from an outpatient or inpatient healthcare facility, or a positive laboratory test from a recent (within 7 days) hospital admission, then the discharge code was considered to be accurate even if there was no confirmatory laboratory test from the study institution for that particular admission. Cases in which there was laboratory confirmation of influenza infection but the patient did not have an influenza-specific discharge ICD-9 code were also investigated.

**RESULTS**

There were 23 282 patients discharged from the 3 participating hospitals during the study period, and a total of 333 patients were assigned 1 of the 3 ICD-9 discharge diagnosis codes for influenza (Table 1). Based upon laboratory data, there were a total of 273 laboratory-confirmed influenza cases during the same time intervals and 198 (72.5%) of these cases were also assigned an influenza ICD-9 discharge code. The sensitivity, specificity, PPV, and NPV of an influenza ICD-9 discharge code for the identification of laboratory-confirmed influenza cases are presented in Table 1. Sensitivity ranged from 65.8% to 82.8% at each site and was 72.5% (95% CI, 61.5–75.3) overall. Specificity ranged from 99.0% to 99.7% at each site and was 99.4% (95% CI, 99.3–99.5) overall. The PPV for an influenza ICD-9 discharge code was 59.5% overall (95% CI, 53.0–64.0) and ranged widely from 37.8% to 77.7% at each site.

There were a total of 210 patients with discrepancies between ICD-9 code assignment and laboratory confirmation. One hundred and thirty-five of these patients had a discharge ICD-9 code for influenza but no laboratory confirmation in the medical record, and 75 had a positive laboratory test for influenza but no influenza-specific discharge ICD-9 code. In this latter group, 46 had discharge codes for fever, pneumonia, upper respiratory infection, viral infection, or other respiratory conditions. We performed structured chart reviews of discrepant cases (n = 87) at one of the sites (Site 2). Among the 31 patients with an influenza-specific ICD-9 code but no laboratory test, 13 (42%) were positive for parainfluenza virus, 1 (3%) was positive for Haemophilus influenzae, and 13 (42%) had documentation of a positive test from an outside healthcare facility (data not shown). The 13 patients who had a positive test from another healthcare facility or from a previous recent admission were counted as an accurate ICD-9 code assignment. Among the 56 cases who had laboratory-confirmed influenza but were not assigned an ICD-9 code for influenza, 33 (59%) patients were hospitalized for <24 hours and 6 (11%) had test results that were still pending at the time of discharge.

**DISCUSSION**

In this multicenter validation study of the use of administrative data as a case-finding strategy for children hospitalized with laboratory-confirmed influenza, we found that influenza-specific ICD-9 codes were moderately sensitive and highly specific for the identification of laboratory-confirmed influenza among children admitted to tertiary care pediatric facilities. A discharge diagnosis code for influenza will detect approximately 73% of laboratory-confirmed influenza cases with a PPV of 60%, whereas less than 1% of patients without a diagnosis code will have had laboratory-confirmed influenza.

Our results show a higher sensitivity and lower PPV compared with a previous study that evaluated
sensitivity of influenza-specific ICD-9 codes at one of the participating institutions during the 2001 through 2004 influenza season [7]. There are several factors that may explain our results. First, it is known that current surveillance methods underestimate the number of true cases of influenza. Even if testing is performed, other discharge codes for acute respiratory diseases or pneumonia may be used. We found that the majority of discrepant cases with positive test results had discharge diagnoses for other respiratory symptoms and conditions. Current surveillance methods for influenza-like illness use many of these diagnostic codes to optimize case-capture—the addition of other codes may increase sensitivity for the diagnosis of influenza among hospitalized children, but this would likely be at the expense of specificity [1, 3, 8, 9]. Second, there may be misclassification related to inaccurate labeling of children diagnosed with other respiratory pathogens. We identified 14 cases with discharge codes for influenza who had positive tests for parainfluenza virus and H. influenza. We presume this misclassification arose because of similar names for these pathogens. In addition, because influenza symptoms can be similar to symptoms caused by other respiratory pathogens, a patient may be diagnosed clinically with influenza and coded as such but have negative test results. Finally, although laboratory confirmation is considered the gold standard for diagnosis, not all testing modalities are equally sensitive [10]. Thus, it is possible that some patients had a false-negative or false-positive test and would have therefore been misclassified. We did not have the necessary data to stratify our results based upon the testing modalities used for laboratory confirmation.

Chart review of discrepant cases identified other issues that may affect the sensitivity and PPV of this case-finding strategy. Of the 56 laboratory-confirmed cases without an influenza-specific ICD-9 code investigated in our review, 33 children had a short-stay admission (<24 hours). In these cases, patients may have presented with influenza-like symptoms that prompted testing. However, results may not have been available before discharge, and such cases would not be assigned an influenza discharge code. This suggests that administrative data may not be the best way to identify cases among short-stay pediatric patients.

Our work has important limitations that may impact the potential use of administrative data for influenza surveillance. We only captured patients who presented for medical care and required hospital admission. Therefore, our results are unlikely to apply to administrative data from ambulatory networks where viral testing is performed less frequently. In addition, sensitivity of discharge ICD9 codes may be dependent upon the severity of the influenza season, because such sensitivity measures will not be reliable from year to year. Finally, conversion to an ICD-10 coding system is scheduled to occur in 2013. Whereas existing ICD-9 codes can be mapped to the new coding schema [11], future validation studies may be required to ensure accurate longitudinal surveillance.

In summary, our results indicate that using administrative data to detect influenza cases is moderately sensitive but highly specific. Relying on the presence of an influenza-specific ICD-9 code may miss up to 30% of influenza cases among hospitalized pediatric patients, but the lack of a discharge code for influenza can reliably indicate the absence of laboratory-confirmed infection. Given the rising number of networks that use EHRs, this case-finding strategy may be a useful epidemiologic tool. The availability of accurate rapid diagnostic testing may improve the sensitivity of this approach as will efforts to minimize misclassification.

<table>
<thead>
<tr>
<th>Site</th>
<th>ICD-9 Cases</th>
<th>Laboratory Confirmed Cases</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PPV (95% CI)</th>
<th>NPV (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>67</td>
<td>51</td>
<td>82.3% (69.1, 91.6)</td>
<td>99.5% (99.3, 99.7)</td>
<td>62.7% (50.0, 74.2)</td>
<td>99.8% (99.7, 99.9)</td>
</tr>
<tr>
<td>Site 2</td>
<td>139</td>
<td>164</td>
<td>65.8% (58.0, 73.0)</td>
<td>99.7% (99.5, 99.8)</td>
<td>77.7% (69.9, 84.3)</td>
<td>99.4% (99.3, 99.6)</td>
</tr>
<tr>
<td>Site 3</td>
<td>127</td>
<td>58</td>
<td>82.8% (71.1, 90.4)</td>
<td>99.0% (98.8, 99.3)</td>
<td>37.8% (29.8, 46.5)</td>
<td>99.9% (99.7, 99.9)</td>
</tr>
<tr>
<td>All sites</td>
<td>333</td>
<td>273</td>
<td>72.5% (66.9, 77.5)</td>
<td>99.4% (99.3, 99.5)</td>
<td>59.5% (54.1, 64.6)</td>
<td>99.7% (99.5, 99.7)</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

ICD-9 codes 487.0, 487.1, and 487.8.
Acknowledgments

We acknowledge the 4 study hospitals (The Children’s Hospital of Philadelphia, Seattle Children’s Hospital, St. Louis Children’s Hospital, and Children’s Mercy Hospitals and Clinics) for participation in this study.

Potential conflicts of interest. D. M. Z. received funding support from Sage Products, Inc and Vioguard Ltd. All other authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References


