Validation of an administrative claims-based diagnostic code for pneumonia in a US-based commercially insured COPD population

David M Kern1
Jill Davis2
Setareh A Williams3
Ozgur Tunceli1
Bingcao Wu1
Sally Hollis4
Charlie Strange5
Frank Trudo2

1HealthCore, Inc., Wilmington, DE, 2AstraZeneca Pharmaceuticals, Wilmington, DE, 3AstraZeneca Pharmaceuticals, Gaithersburg, MD, USA, 4AstraZeneca Pharmaceuticals, Cheshire, UK, 5Department of Medicine, Division of Pulmonary and Critical Care Medicine, Medical University of South Carolina, Charleston, SC, USA

Objective: To estimate the accuracy of claims-based pneumonia diagnoses in COPD patients using clinical information in medical records as the reference standard.

Methods: Selecting from a repository containing members’ data from 14 regional United States health plans, this validation study identified pneumonia diagnoses within a group of patients initiating treatment for COPD between March 1, 2009 and March 31, 2012. Patients with ≥1 claim for pneumonia (International Classification of Diseases Version 9-CM code 480.xx–486.xx) were identified during the 12 months following treatment initiation. A subset of 800 patients was randomly selected to abstract medical record data (paper based and electronic) for a target sample of 400 patients, to estimate validity within 5% margin of error. Positive predictive value (PPV) was calculated for the claims diagnosis of pneumonia relative to the reference standard, defined as a documented diagnosis in the medical record.

Results: A total of 388 records were reviewed; 311 included a documented pneumonia diagnosis, indicating 80.2% (95% confidence interval [CI]: 75.8% to 84.0%) of claims-identified pneumonia diagnoses were validated by the medical charts. Claims-based diagnoses in inpatient or emergency departments (n=185) had greater PPV versus outpatient settings (n=203), 87.6% (95% CI: 81.9%–92.0%) versus 73.4% (95% CI: 66.8%–79.3%), respectively. Claims-diagnostics verified with paper-based charts had similar PPV as the overall study sample, 80.2% (95% CI: 71.1%–87.5%), and higher PPV than those linked to electronic medical records, 73.3% (95% CI: 65.5%–80.2%). Combined paper-based and electronic records had a higher PPV, 87.6% (95% CI: 80.9%–92.6%).

Conclusion: Administrative claims data indicating a diagnosis of pneumonia in COPD patients are supported by medical records. The accuracy of a medical record diagnosis of pneumonia remains unknown. With increased use of claims data in medical research, COPD researchers can study pneumonia with confidence that claims data are a valid tool when studying the safety of COPD therapies that could potentially lead to increased pneumonia susceptibility or severity.

Keywords: positive predictive value, pneumonia, validation, claims data, medical record review

Introduction

Claims data are generated primarily from administrative transactions and are essential for the accounting and reimbursement purposes associated with health care delivery. Increasingly, however, administrative claims data are repurposed for real-world observational studies in the United States and other developed countries. Typically maintained in regularly updated repositories, claims data provide convenient and easy access to researchable details on demographics, clinical characteristics, health care utilization – frequency and type of prescription medication fills, inpatient...
hospitalization, and the use of outpatient, emergency, and physician offices – and the cost of care. Claims databases efficiently and comprehensively capture structured data such as diagnostic codes from the International Classification of Diseases Version 9 (ICD-9-CM) or generic product identifier medication codes. In general, these databases do not enable access to data on patients’ race or ethnicity, laboratory test values and other clinical measures or to unstructured data, including clinicians’ progress notes, all of which could be valuable in identifying disease and establishing patient status. It is essential, therefore, when assessing outcomes related to disease identified with claims data, to understand the degree of agreement between claims-based observations and actual physician diagnoses, and such validations are not available for several diseases.

Pneumonia was associated with more than 1.1 million inpatient hospitalizations and 50,000 deaths in 2010, and is more common and more fatal in COPD patients than in those without COPD. In the first year after a COPD diagnosis, individuals are at 16 times the risk for pneumonia compared to those without COPD. In a recent study the incidence rate of community acquired pneumonia was 22.4 events per 1,000 person years in the 10 years following the diagnosis of COPD, and more than 50% higher in those categorized as having severe COPD. Furthermore, the economic impact of pneumonia is greater for those with COPD, illustrated by a doubling of direct medical costs following an inpatient hospitalization for pneumonia compared to those without COPD in a study of older individuals. Thus, pneumonia is especially of interest in individuals with COPD, and administrative claims databases can be a useful tool for studying this disease if the data are determined to be sufficiently valid.

In the past few years alone, pneumonia has been widely researched using claims data in patients with and without COPD, with studies focusing on the economic and clinical impact of the disease, the development of algorithms to predict inpatient readmission, and measuring the safety and efficacy of treatments. Although the accuracy of claims-based diagnoses is critical to the sensitivity and validity of observational research, this remains largely unexamined for pneumonia. To our knowledge, no study has validated claims-based pneumonia diagnoses across different points of service such as inpatient hospitalizations, emergency departments (EDs), and physicians’ offices.

The current study aimed to determine the validity of pneumonia diagnosis codes in a large US-based administrative claims database by comparing patients’ claims-based records and their medical records at the time of their observed pneumonia diagnosis. The objective was to confirm the accuracy of claims-based diagnoses of pneumonia by using the physicians’ diagnoses (medical record-based) as the reference standard.

Materials and methods

Data source

This retrospective cohort study utilized claims data for commercially insured individuals from the HealthCore Integrated Research Environment. The HealthCore Integrated Research Environment contains a diverse spectrum of longitudinal claims data for more than 32 million individuals, and has the capability of linking claims data to providers, which facilitates the selection of medical records.

All study materials were handled in compliance with the Health Insurance Portability and Accountability Act, and a limited dataset was used for all analyses, as defined by the Privacy Rule. A national, institutional review board (IRB), the New England IRB, reviewed the protocol and approved this study prior to the researchers obtaining patient medical records. HealthCore was granted a waiver of authorization to obtain the medical records without patient authorization of release after the IRB determined the use or disclosure of protected health information in this research project involved no more than minimal risk to the individuals.

Study population

This validation study was nested within a larger, retrospective cohort study (NCT01921127) comparing the effectiveness of two common inhaled corticosteroid/long-acting β2-agonist combination (ICS/LABA) medications in COPD patients (ICD-9-CM code 491.xx, 492.xx, or 496.xx). Pneumonia, a complication in many COPD patients, was analyzed as a secondary outcome, and the validation study was performed to assess the accuracy of the pneumonia diagnosis code (ICD-9-CM code 480.xx–486.xx) to reflect a documented diagnosis of pneumonia reported in the medical records. Patients initiating ICS/LABA between March 1, 2009 and March 31, 2012 were followed for 1 year following treatment initiation, during which patients diagnosed with pneumonia were identified. Only the first event was assessed for patients with more than one claim with a diagnosis of pneumonia.

Medical record abstraction

A professional medical record abstraction agency, employing trained and experienced medical data reviewers, obtained and abstracted the patient medical records using a study-specific instrument developed by the research team. The medical record
Statistical analysis

A target final sample size of 400 patients was chosen a priori to allow for the estimation of validity with a maximum margin of error <5%. A threshold of 5% margin of error was selected to achieve a confidence interval (CI) no more than 10% wide. Additionally, robust sample sizes, ideally of ≥100 records, were wanted for each of the subgroup analyses by place of service (inpatient/ED versus outpatient) and chart type (paper, electronic, or a combination of both paper and electronic records [hybrid]), with the assumption that no subgroup would account for less than one quarter of all records based on the claims results and prior experience. The primary outcome of the validation analysis was calculated using the positive predictive value (PPV) of the claims diagnosis for pneumonia relative to the medical record review, the reference standard. PPV was calculated as the number of pneumonia cases with a documented pneumonia diagnosis in the medical records (“diagnosed pneumonia”) (true positives) divided by the total number of cases identified from the claims data (true positives plus false positives). Results were reported overall and stratified by place of service (inpatient or ED vs outpatient) and medical record type (paper, electronic, or hybrid). A 95% CI for the estimated PPV was constructed using Clopper-Pearson (exact) confidence limits. Additional descriptive statistics are reported for each of the demographic characteristics, symptoms, diagnostic tests, and antibiotic use. All analyses were performed using SAS version 9.2 (SAS Institute Inc., Cary, NC, USA).

Results

Medical record collection

A total of 1,345 patients had ≥1 claim with a pneumonia diagnosis during an inpatient hospitalization, ED visit, or outpatient visit over the 12-month follow-up period; 1,022 of the diagnoses were from claims containing provider contact information, and were deemed eligible for medical record abstraction. A random subset of 800 patients was selected with the goal of abstracting information on 400 patients, allowing for the possibility that up to 50% of targeted records may be unobtainable. There were 730 record requests to providers because an initial review showed 70 patients with insufficient contact information. Medical records were obtained for 407 patients; however 19 were unusable due to missing information on birth dates or sex, or had unmatched names, or incorrect time periods, which resulted in 388 abstracted records (Figure 1).

Patient characteristics

The average age of patients with abstracted records was 67.8 (±11.6) years old, and 53.9% were female (Table 1). Of the 47.9% of patients with information on race noted in the record, 92.5% were white. Patients’ smoking status was reported as active (23.2%), former (36.3%), never (10.3%), and status unknown for the remaining 30.2%. Patients whose records were abstracted (n=388) and pneumonia patients identified with claims data (n=1,345) had similar mean age (67.8 and 67.4 years), sex (53.9% and 52.9% female), and health plan type (60.3% and 60.6%, enrolled in preferred provider organizations), respectively. A slightly greater proportion of patients were from the Midwest in the record-reviewed vs the claims group (51.8% vs 47.1%) and a slightly smaller proportion from the Northeast (12.1% vs 17.5%).

Primary analysis

A documented pneumonia diagnosis was found in 311 out of the total 388 available medical records (Table 2), indicating that 80.2% (PPV 95% CI: 75.8% to 84.0%) of pneumonia positive patients identified with the claim code algorithm were diagnosed with pneumonia according to the medical records. Stratification by health care setting showed that...
Table 1 Patient demographics

<table>
<thead>
<tr>
<th>All pneumonia patients identified in claims (N=1,345)</th>
<th>Medical record status</th>
<th>Documented pneumonia diagnosis (N=311)</th>
<th>No documented pneumonia diagnosis (N=77)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>Patients with provider contact information</td>
<td>Requested</td>
<td>Obtained</td>
</tr>
<tr>
<td>Patients with pneumonia diagnosis in claims data</td>
<td>1,345</td>
<td>1,022</td>
<td>800</td>
</tr>
<tr>
<td>Randomly selected patient list sent for abstraction</td>
<td>800</td>
<td>730</td>
<td>407</td>
</tr>
</tbody>
</table>

Notes: *Could not locate provider (n=36), no record of patient (n=44), no record of patient in timeframe requested (n=32), provider refuses to participate (n=52), provider requires patient consent or IRB approval (n=41), chart lost/destroyed/archived (n=11), no response from office (n=63), payment denied/record not needed (n=3), requests suspended (met quota) (n=41).

Abbreviations: DOB, date of birth; IRB, institutional review board.

Notes: Age, geographic region, and healthplan type were captured via claims data at the index date, while race and smoking status were identified from the medical records.

*Not applicable, data are only available in medical records, not the claims data.

Abbreviations: HMO, health maintenance organization; PPO, preferred provider organization; CDHP, consumer-driven health plan; SD, standard deviation.
claims originating from an inpatient or ED (n=185) had a higher PPV than those from outpatient settings (n=203) (87.6% vs 73.4%, respectively). Claims that were linked to paper charts had the same PPV as the overall study population, 80.2% (95% CI: 71.1%–87.5%), while those linked to electronic medical records (EMR) had a lower PPV, 73.3% (95% CI: 65.5%–80.2%), and those with data from both paper charts and EMR had a higher PPV, 87.6% (95% CI: 80.9%–92.6%).

Diagnosed pneumonia cases (n=311) vs those without a documented diagnosis in the medical records (n=77) included fewer females (51.4% vs 63.6%) and a higher proportion of smokers (26.0% active and 39.2% former smokers vs 11.7% and 24.7%, respectively).

### Table 2: Primary outcome – positive predictive value for pneumonia in claims

<table>
<thead>
<tr>
<th></th>
<th>Number abstracted</th>
<th>Number of documented pneumonia diagnosis</th>
<th>PPV (%) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All pneumonia patients</td>
<td>388</td>
<td>311</td>
<td>80.2 (75.8–84.0)</td>
</tr>
<tr>
<td>By place of service</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inpatient hospitalization/emergency department visit</td>
<td>185</td>
<td>162</td>
<td>87.6 (81.9–92.0)</td>
</tr>
<tr>
<td>Outpatient/office visit</td>
<td>203</td>
<td>149</td>
<td>73.4 (66.8–79.3)</td>
</tr>
<tr>
<td>By medical record type</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paper</td>
<td>101</td>
<td>81</td>
<td>80.2 (71.1–87.5)</td>
</tr>
<tr>
<td>Electronic medical records (EMR)</td>
<td>150</td>
<td>110</td>
<td>73.3 (65.5–80.2)</td>
</tr>
<tr>
<td>Hybrid of paper and EMR</td>
<td>137</td>
<td>120</td>
<td>87.6 (80.9–92.6)</td>
</tr>
</tbody>
</table>

Abbreviation: PPV, positive predictive value; CI, confidence interval.

### Pneumonia symptoms

Within all examined medical records, coughing (61.3%), was the most prevalent symptom reported, followed by dyspnea (53.9%), wheezing (42.5%), distant breath sounds (31.2%), rhonchi (28.9%), fever (26.0%), and rales (25.5%) (Table 3). There was a clear discrepancy in the frequency of reported symptoms between records with and without a pneumonia diagnosis; just 15.8% of patients with a pneumonia diagnosis documented in the medical records had none of the symptoms listed, while more than half (53.2%) of those without a diagnosis had no symptoms mentioned. Furthermore, individual symptoms were 2–5 times more prevalent in diagnosed cases compared with those without a diagnosis documented in the medical records.

### Table 3: Pneumonia symptoms captured from medical charts

<table>
<thead>
<tr>
<th>Symptoms (n, %)</th>
<th>All medical records (N=388)</th>
<th>Medical record status</th>
<th>Documented pneumonia diagnosis (N=311)</th>
<th>No documented pneumonia diagnosis (N=77)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No symptoms reported</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>90</td>
<td></td>
<td>49</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>1 symptom</td>
<td></td>
<td>29</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>2 symptoms</td>
<td></td>
<td>29</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>3 symptoms</td>
<td></td>
<td>37</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4+ symptoms</td>
<td></td>
<td>167</td>
<td>13</td>
</tr>
</tbody>
</table>

Pneumonia symptoms

Within all examined medical records, coughing (61.3%), was the most prevalent symptom reported, followed by dyspnea (53.9%), wheezing (42.5%), distant breath sounds (31.2%), rhonchi (28.9%), fever (26.0%), and rales (25.5%) (Table 3). There was a clear discrepancy in the frequency of reported symptoms between records with and without a pneumonia diagnosis; just 15.8% of patients with a pneumonia diagnosis documented in the medical records had none of the symptoms listed, while more than half (53.2%) of those without a diagnosis had no symptoms mentioned. Furthermore, individual symptoms were 2–5 times more prevalent in diagnosed cases compared with those without a diagnosis documented in the medical records.
Diagnostic testing and antibiotic use

Diagnostic testing was more common in diagnosed pneumonia cases vs those without a documented diagnosis of pneumonia in the medical record (Table 4). Respiratory rate testing was performed in 74.3% of diagnosed pneumonia cases compared with 42.9% of records that had no pneumonia diagnosis. Within those that had a test, abnormal respiratory rate results were reported for 55.0% of those with a documented diagnosis of pneumonia and 45.5% of those without. Similarly, chest X-ray and CAT scans were performed more often in those with vs without a documented diagnosis (77.8% and 24.8% vs 51.9% and 11.7%, respectively), and abnormal results in those with diagnosed pneumonia were found for 89.6% of X-rays and 90.5% of CAT scans compared to 61.5% and 77.8% of X-rays and CAT scans within patients without a diagnosis documented in the charts. Results of other diagnostic tests showed similar differences.

Medical record reviews indicated that antibiotics were administered or prescribed to 79.7% of the diagnosed pneumonia cases compared with 35.1% that had no documented diagnosis (Table 5). Levofloxacin (25.8%), azithromycin (25.0%), and ceftriaxone (24.2%) were the three most commonly used antibiotics in this patient sample; all three are commonly used to treat pneumonia. The lack of evidence of antibiotic use in 20% of confirmed cases may suggest that many of those cases were suspected to be viral in origin.

Discussion

This study tested the validity of pneumonia diagnoses among COPD patients identified in a large US administrative claims database using medical records as a reference standard. This study differs from prior investigations on this subject in a number of important ways. Prior studies have examined the validity of claims data in diagnosing pneumonia.25,26 However, those studies were either limited to a single hospital or ED,25,26 to children in pediatric hospitals,28 or to only hospitalizations due to pneumonia.29

We used ICD-9-CM codes to identify pneumonia diagnoses in COPD patients from outpatient, ED, and inpatient settings using a claims database that covers more than 30 million managed care patients across the US. Although there are major differences in the design of our study compared to others, the PPV calculated for claims-based pneumonia diagnoses in the current study was similar to those reported by Aronsky et al (79.4%–80.8% PPV based on the algorithm that used different combinations of ICD-9-CM codes)23 and Whittle et al (89%).26 Both of these PPVs were similar to the 87.6% PPV found in inpatient or ED cases in our study.

Table 4 Patients with available results for diagnostic tests recorded in medical charts and the proportion with abnormal results

<table>
<thead>
<tr>
<th>Diagnostic test</th>
<th>Medical record status</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All medical records (N=388)</td>
</tr>
<tr>
<td>Respiratory rate test</td>
<td>264</td>
</tr>
<tr>
<td>Results available in chart</td>
<td>264</td>
</tr>
<tr>
<td>Abnormal (&lt;12 or &gt;20 breaths per minute)</td>
<td>142</td>
</tr>
<tr>
<td>Heart rate test</td>
<td>295</td>
</tr>
<tr>
<td>Results available in chart</td>
<td>294</td>
</tr>
<tr>
<td>Abnormal (&lt;60 or &gt;100 bpm)</td>
<td>131</td>
</tr>
<tr>
<td>Chest X-ray test</td>
<td>282</td>
</tr>
<tr>
<td>Results available in chart</td>
<td>270</td>
</tr>
<tr>
<td>Abnormal*</td>
<td>231</td>
</tr>
<tr>
<td>Chest CAT scan test</td>
<td>86</td>
</tr>
<tr>
<td>Results available in chart</td>
<td>83</td>
</tr>
<tr>
<td>Abnormal*</td>
<td>74</td>
</tr>
<tr>
<td>Sputum Gram stain test</td>
<td>71</td>
</tr>
<tr>
<td>Results available in chart</td>
<td>58</td>
</tr>
<tr>
<td>Abnormal (positive)</td>
<td>40</td>
</tr>
<tr>
<td>Sputum culture test</td>
<td>78</td>
</tr>
<tr>
<td>Results available in chart</td>
<td>62</td>
</tr>
<tr>
<td>Abnormal (positive, isolated organism)</td>
<td>27</td>
</tr>
<tr>
<td>Blood culture test</td>
<td>144</td>
</tr>
<tr>
<td>Results available in chart</td>
<td>133</td>
</tr>
<tr>
<td>Abnormal (positive, isolated organism)</td>
<td>12</td>
</tr>
</tbody>
</table>

Notes: The proportion of charts with abnormal results are based on the denominator of charts with available results. *Abnormal included any of the following notes: “consolidation”, “pleural effusion”, “fluid”, “infiltrate”, “haze”, or “consistent with pneumonia”.

Abbreviations: bpm, beats per minute; CAT, computerized axial tomography.
Compared to outpatient claims, pneumonia diagnoses in inpatient or ED claims were found to be more accurate (87.6% vs 73.4% PPV). Furthermore, the presence of a chest X-ray or chest CAT scan was more common in inpatient and ED cases, of which, all but one individual (99.4%) with documented pneumonia had evidence of either procedure; conversely, chest imaging was present in just 61.1% of cases originating in an outpatient setting with documented pneumonia. Further research is needed to investigate differences in validity by place of service since there is more detailed information in inpatient medical records compared to those in physicians’ offices. Our study suggests that differences in pneumonia outcomes may be more difficult to study in the outpatient setting; since neither medical records nor claims data strongly support the accuracy of the medical diagnosis of pneumonia as evidenced by the lack of imaging in significant numbers of individuals.

One goal of this research was to confirm that a patient was diagnosed with pneumonia during the same health care encounter in which we observed a diagnosis of pneumonia in the claims data. The goal was not to determine whether the diagnosis given by the physician was accurate; however, there was evidence that pneumonia diagnoses that were simultaneously documented in the medical records were usually accurately diagnosed by the physician. Many claims-only diagnoses may not have truly been pneumonia. For example, patients with a pneumonia diagnosis documented in the medical records had abnormal diagnostic test results and symptoms of pneumonia at least twice as often compared with those who did not have a diagnosis in the medical records. Additionally, the higher smoking rates in those with a documented diagnosis is consistent with research that has shown smoking to be the strongest independent risk factor of pneumonia in immunocompetent non-elderly adults.\textsuperscript{30,31} Conversely, although nearly 20% of patients in the claims data did not have a pneumonia diagnosis documented in the medical records, there was evidence that some of those patients may have had pneumonia. The prevalence of individual symptoms being less common in unconfirmed cases may be a function of the pneumonia being truly absent, but also may be a result of additional omissions in medical records (ie, physicians who omitted a note of pneumonia in the record may also be more likely to leave out notes of specific symptoms). Furthermore, more than 20% of patients with an unconfirmed diagnosis had three or more symptoms that are associated with pneumonia; of the 40 patients who had a chest X-ray, 61.5% had an abnormal result, and 35.1% unconfirmed cases were given antibiotics during their visit. Thus, although the medical record was considered the reference standard in this analysis, they might not be 100% accurate, and the results could be an underestimate of the true PPV of the claims diagnosis.

### Strengths
This study was able to directly link patients’ records from an administrative claims database to their medical records for the exact time they had an insurance claim for a diagnosis of pneumonia in outpatient, inpatient, and EDs. This allowed the confirmation that pneumonia diagnoses observed in the claims database were present in the medical records, where
the medical record was considered the reference standard in the validation analysis. Along with capturing the presence of a pneumonia diagnosis, we were also able to obtain information on the presence of pneumonia symptoms, diagnostic lab testing, and administration of antibiotics during the visit of interest. The high rates of symptoms and abnormal test results in the medical record documented cases provided further evidence that most were true pneumonia patients.

Limitations
Medical records, though used as the reference standard here, are not 100% accurate. For example, it was possible for records to be at multiple locations for a single patient during the same time period, due to different providers having separate paper charts and/or EMR systems; from which a record from one provider may have had a pneumonia diagnosis while another did not. Each event had a medical record pulled from a single location and, thus, it was possible to miss some records with a pneumonia diagnosis. Further, human error plays a role in accurate recording and abstracting of diagnoses. It is possible, that a patient was diagnosed with pneumonia but it was not noted in the record, or that part of the medical record was missing, or that a diagnosis for pneumonia was in the record but missed by the abstractor. However, all medical record abstractors were trained and experienced in the health care field and participated in project-specific data collection training that utilized the data collection manual and materials specifically developed for this study. Less than 80% of patients with pneumonia documented in the medical record had a chest X-ray noted, a procedure that should be performed prior to diagnosing pneumonia. It may be that imaging was performed but not included in the chart, performed at an alternative medical site that was not captured, or that the provider felt there was enough evidence to make a diagnosis of pneumonia without an imaging test. While a clinical diagnosis of pneumonia without a chest X-ray may not be clinically sound, the goal of this study was not to validate whether or not the clinician made a correct diagnosis, instead it was to determine whether or not the administrative claims accurately reflected the diagnosis made by the clinician.

The reference standard of medical record abstraction has limitations. Some of the comorbidities of COPD that include congestive heart failure, pulmonary emboli, and mucus plugging during a COPD exacerbation may be diagnosed as pneumonia with an abnormal radiograph and pulmonary symptoms, treated with antibiotics, and yet not be pneumonia at all. Biomarkers such as brain natriuretic peptide for congestive heart failure and D-dimer for thromboembolic disease should be captured in future prospective studies. However, medical records were considered the gold standard as they are the repository for the clinical information documented by the medical provider and other health care professionals and utilized for the care of the patient. Furthermore, medical records have been used as the gold standard in a number of other studies to determine the validity of pneumonia diagnoses.27-29,33

Because this study examined only patients with a pneumonia diagnosis observed in claims, the negative predictive value, sensitivity, and specificity of pneumonia identified via administrative claims could not be estimated from this study.

Lastly, all patients in the larger study design were required to have 12 months of follow-up and all pneumonia diagnoses were captured during those 12 months. Thus, only pneumonia patients who survived until the end of the follow-up period were included, and so the most severe cases of pneumonia resulting in death were not represented in this validation study. The validity recorded in this study might have been higher if cases of pneumonia that resulted in death were included, since these cases in an inpatient setting would be expected to have higher validity compared to outpatient diagnoses.

Conclusion
This study demonstrated that administrative claims data provide an accurate basis for the diagnosis of pneumonia in COPD patients across multiple service settings. Diagnostic accuracy varies by treatment setting and tends to be better in inpatient and EDs vs outpatient settings. Additional research will help to validate whether the service setting has a bearing on the accuracy of pneumonia diagnosis and whether differences were a consequence of data type or the actual content of the data inherent in claims and medical records. With the increasing amount of research being performed utilizing administrative claims it is important to understand the validity of such data. Confirming the validity of pneumonia diagnoses in claims data allows researchers to confidently study COPD associated pneumonia and investigate the safety of COPD therapies utilizing longitudinal observational data.

Acknowledgment
Bernard Tulsi provided editorial support for this manuscript. AstraZeneca Pharmaceuticals provided funding.

Disclosure
JD, FT, SAW, and SH have disclosed that they are employed by AstraZeneca. DMK, OT, and BW have disclosed that they are employed by HealthCore, Inc., which received funding from AstraZeneca for this study. CS has been a consultant for AstraZeneca LP, CSL Behring, Grifols, PneumRx, Pulmonx,
and Uptake Medical on the topic of COPD. The authors report no other conflicts of interest in this work.

References