Accuracy of ICD-9-CM Codes in Detecting Community-acquired Pneumococcal Pneumonia for Incidence and Vaccine Efficacy Studies

Ramon E. Guevara,1 Jay C. Butler,1 Barbara J. Marston,2 Joseph F. Plouffe,3 Thomas M. File, Jr.,4 and Robert F. Breiman1

Studies have used medical record discharge data as coded by the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) to estimate pneumococcal pneumonia incidence and vaccine efficacy. However, the accuracy of coding data to identify laboratory-confirmed pneumococcal pneumonia is not known. With the use of information collected in Ohio for a community-based pneumonia incidence study, the authors calculated the sensitivities, specificities, positive predictive values (PPV), and negative predictive values (NPV) of specific codes for pneumococcal pneumonia among hospitalized patients with community-acquired pneumonia. Sensitivities of the most common ICD-9-CM codes listed in the first five positions for patients with laboratory-confirmed pneumococcal pneumonia were 58.3% (code 481.0, pneumococcal pneumonia), 20.4% (38.0, streptococcal septicemia), 19.2% (38.2, pneumococcal septicemia), 15.0% (518.81, respiratory failure), 14.2% (486.0, pneumonia, organism unspecified), and 11.3% (482.3, streptococcal pneumonia). Using the first five listed ICD-9-CM codes rather than just the first listed code increased sensitivity without causing substantial change in specificity, PPV, and NPV. Sensitivity, PPV, and NPV of individual and groups of codes varied with different case definitions of pneumococcal pneumonia. Incidence and vaccine efficacy studies with the ability to validate diagnoses by medical chart review can use a combination of many ICD-9-CM codes to maximize sensitivity. However, without the ability to review medical charts, researchers must carefully decide which codes would best suit their studies.

Case-control and serotype prevalence studies have demonstrated that pneumococcal vaccine is effective for preventing invasive disease, such as bacteremia and meningitis (1–7). However, it has been much more difficult to determine the effectiveness of pneumococcal vaccine in preventing pneumococcal pneumonia or bronchitis without bacteremia (8). Whereas diagnosis of invasive infection is 100 percent specific on the basis of recovering Streptococcus pneumoniae from blood or cerebrospinal fluid (9), case ascertainment for pneumococcal pneumonia without bacteremia is much more challenging because of the limited sensitivity and specificity of sputum Gram’s stain and culture for diagnosis. This diagnostic uncertainty limits the validity of the case definitions used for pneumococcal pneumonia without bacteremia in epidemiologic studies designed to assess vaccine effectiveness. Moreover, the size of such studies is generally limited because case definitions based on laboratory test results require considerable resources for review of microbiologic records.

A more efficient method of case finding that has been utilized in studies with large numbers of patients is to identify patients with pneumococcal pneumonia by using computerized medical record discharge diagnostic data as coded according to the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) (10). However, the quality of data from this type of study depends on the accuracy of coding. Therefore, we evaluated the accuracy of ICD-9-CM codes for identifying pneumococcal pneumonia cases in persons hospitalized with community-acquired pneumonia.
During 1991-1992, 4,385 patients hospitalized with community-acquired pneumonia at all 15 acute-care hospitals in Summit and Franklin Counties, Ohio, were prospectively entered into a population-based study of the incidence, etiology, and epidemiology of pneumonia (11). Medical records were abstracted and patients were interviewed using standardized questionnaires. In addition to physical findings and symptoms, medical record abstracts included radiographic data and detailed microbiologic findings, e.g., results of sputum Gram’s stain, sputum cultures, and blood cultures. Serum specimens were obtained and evaluated systematically for a variety of pneumonia etiologies, including Legionella, Mycoplasma pneumoniae, Chlamydia pneumoniae, respiratory syncytial virus, and parainfluenza virus. We calculated the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of ICD-9-CM codes for patients with community-acquired pneumococcal pneumonia by comparing frequencies of ICD-9-CM codes among patients with and without evidence of pneumonia caused by S. pneumoniae.

MATERIALS AND METHODS

This study used the data from a community-based pneumonia incidence study in Ohio (11, 12). All adults hospitalized for pneumonia in all 15 acute-care hospitals within the two study counties, Franklin and Summit, between January 1991 and May 1992 were identified prospectively and were eligible for inclusion in the study. Franklin and Summit Counties have populations of noninstitutionalized adults (age ≥18 years) of 705,800 and 380,000, respectively (1990 census data).

Inclusion criteria for the analysis of community-acquired pneumonia requiring hospitalization included age ≥18 years, residence in either of the two study counties, no other hospitalization within the previous 30 days, chest radiograph taken within 48 hours of admission revealing a new density consistent with pneumonia in a patient with any one of the following: 1) fever (reported fever or chills, or documented temperature >100.8°F (>38.2°C) within 24 hours of admission), 2) abnormal white blood cell count (>11,000/mm³, <3,000/mm³, or with abnormal differential), 3) hypothermia (documented temperature of <96°F (<35.6°C) within 24 hours of admission), or 4) productive cough (reported by patient or patient proxy).

Patients were categorized as having definite, probable, possible, or no pneumococcal infection based on the following case definitions:

**Definite.** S. pneumoniae (Sp) isolated from blood or pleural fluid.

**Probable.** Isolation of Sp from purulent sputum (defined as sputum with moderate or large numbers of neutrophils seen on Gram’s stain) in which gram-positive diplococci were noted in moderate or large amounts on sputum Gram’s stain.

**Possible.** Isolation of Sp from purulent sputum in the presence of a compatible Gram’s stain, or a Gram’s stain of purulent sputum demonstrating a predominance of gram-positive diplococci without isolation of Sp from the sputum.

**Non-pneumococcal disease.** No evidence of Sp in blood, pleural fluid, sputum, or sputum Gram’s stain (blood, pleural fluid, or sputum was cultured for pyogenic bacteria in 89 percent of non-Sp cases; sputum Gram’s stains were done in 61 percent).

We calculated the frequencies with which each of the ICD-9-CM codes was listed in each of the first five positions for each category of pneumococcal disease.

We evaluated the sensitivities, specificities, PPVs, and NPVs of the most common ICD-9-CM codes for the definite, probable, and possible categories of Sp disease using the following proportions shown in figure 1.

When the ICD-9-CM codes are used to identify patients with Sp disease, those classified as having the disease will be the total of A + B, while those classified as not having the disease will be the total of C + D (the denominators of PPV and NPV, respectively). High PPV and NPV would indicate a high probability of correct classification.

To analyze how broadening the case definition for Sp pneumonia affected sensitivity, specificity, PPV, and NPV of the codes, we used the following classes of case definitions (based on diagnostic certainty):

- **Class 1:** definite Sp cases;
- **Class 2:** definite and probable Sp cases;
- **Class 3:** definite, probable, and possible Sp cases.

When the Class 1 case definition was used, probable and possible Sp cases were excluded from calculations of sensitivity, specificity, PPV, and NPV unless otherwise specified. Similarly, with the Class 2 case definition, possible Sp cases were excluded from the analysis unless otherwise specified.

RESULTS

Of the 4,385 patients hospitalized for pneumonia, 240 (5.5 percent) had definite Sp disease, 53 (1.5 percent) had probable, and 268 (6.1 percent) had possible Sp disease.

The most commonly cited ICD-9-CM codes for Sp disease

The most commonly cited ICD-9-CM codes for all patients in the study were those for streptococcal sep-
ticemia, pneumococcal septicemia, pneumococcal pneumonia (includes lobar pneumonia, organism unspecified), pneumonia due to *Streptococcus pneumoniae* (organism unspecified), and respiratory failure (table 1). When considering Class 2 and Class 3 case definitions, the most common codes and their proportions were similar to those of definite pneumococcal pneumonia cases (data not shown).

The four most common ICD-9-CM codes listed at position 1 among 53 probable *Sp* cases were those for pneumococcal pneumonia (*n* = 38), pneumonia due to unspecified organism (*n* = 6), streptococcal pneumonia (*n* = 4), and respiratory failure (*n* = 3). For 268 possible *Sp* cases, the six most common position 1 codes were those for pneumococcal pneumonia (*n* = 100), pneumonia due to unspecified organism (*n* = 44), streptococcal pneumonia (*n* = 43), pneumonia due to *Haemophilus influenzae* (482.20) (*n* = 12), respiratory failure (*n* = 11), and pneumonia due to staphylococcus (482.40) (*n* = 10).

**Sensitivity, specificity, PPV, and NPV: position 1 versus positions 1–5**

The sensitivity of each of the most common ICD-9-CM codes for *Sp* pneumonia increased when the analysis included them in any of the first five positions (rather than limiting to position 1) (table 2). Nevertheless, sensitivity was generally low (<25 percent, except for code 481). Code 481 (pneumococcal pneumonia) was the most sensitive ICD-9-CM code. Definite *Sp* cases that code 481 did not identify in positions 1–5 were identified by the codes for streptococcal septicemia (*n* = 14, 5.8 percent), pneumonia, organism unspecified (*n* = 30, 12.5 percent), pneumococcal pneumonia (*n* = 9, 3.4 percent), and respiratory failure (*n* = 12, 5.0 percent). Sensitivity, specificity, PPV, and PNV were similar during each of the 2 years of the study.

Except for codes 486 (pneumonia, organism unspecified) and 518.81 (respiratory failure), the specificity of each code differed by 0.3 percent or less when comparing specificity in position 1 with specificity in positions 1–5. The most specific codes were 38.20 (pneumococcal septicemia), 38.00 (streptococcal septicemia), and 482.30 (streptococcal pneumonia). The specificity of the most sensitive code (481) was only 2.4 percent less than that of the most specific codes.

Generally, PPV of the most common codes for *Sp* pneumonia slightly increased when the analysis included positions 1–5. Although the specificities of 38.00 (streptococcal septicemia), 482.30 (streptococcal pneu-
Accuracy of ICD-9-CM for Pneumococcal Pneumonia

**TABLE 1.** The most common International Classification of Diseases, 9th revision, Clinical Modification (ICD-9-CM) codes among all patients hospitalized with community-acquired Streptococcus pneumoniae (Sp) (n = 4,385) and among those with definite pneumococcal pneumonia (n = 240), Franklin and Summit Counties, Ohio, January 1991 to May 1992

<table>
<thead>
<tr>
<th>ICD-9 code</th>
<th>All patients</th>
<th>Definite Sp and non-Sp pneumonia*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Position 1</td>
<td>Positions 1–5</td>
</tr>
<tr>
<td></td>
<td>No.</td>
<td>%</td>
</tr>
<tr>
<td>38.00 Streptococcal septicemia</td>
<td>31</td>
<td>0.7</td>
</tr>
<tr>
<td>38.20 Pneumococcal septicemia</td>
<td>13</td>
<td>0.3</td>
</tr>
<tr>
<td>38.80 Other specified septicemias</td>
<td>11</td>
<td>0.3</td>
</tr>
<tr>
<td>481.00 Pneumococcal pneumonia (lobar pneumonia, organism unspecified)</td>
<td>330</td>
<td>7.5</td>
</tr>
<tr>
<td>482.30 Pneumonia due to Streptococcus (excludes 481)</td>
<td>93</td>
<td>2.1</td>
</tr>
<tr>
<td>486.00 Pneumonia, organism unspecified</td>
<td>1,803</td>
<td>41.1</td>
</tr>
<tr>
<td>518.81 Respiratory failure</td>
<td>115</td>
<td>2.6</td>
</tr>
</tbody>
</table>

* Excludes probable and possible pneumococcal cases; total definite Sp cases = 240; total non-Sp cases = 3,824.

<table>
<thead>
<tr>
<th>Class and test</th>
<th>38.00</th>
<th>38.20</th>
<th>481.00</th>
<th>482.30</th>
<th>486.00</th>
<th>518.81</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1–5</td>
<td>1–5</td>
<td>1–5</td>
<td>1–5</td>
<td>1–5</td>
<td>1–5</td>
</tr>
<tr>
<td>Class 1†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>10.4</td>
<td>20.4</td>
<td>5.0</td>
<td>19.2</td>
<td>45.4</td>
<td>58.3</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>99.6</td>
<td>99.6</td>
<td>&gt;99.9</td>
<td>&gt;99.9</td>
<td>97.6</td>
<td>97.5</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>87.7</td>
<td>74.2</td>
<td>92.3</td>
<td>95.8</td>
<td>56.8</td>
<td>59.1</td>
</tr>
<tr>
<td>NPV (%)</td>
<td>94.7</td>
<td>95.2</td>
<td>94.4</td>
<td>95.2</td>
<td>96.6</td>
<td>97.4</td>
</tr>
<tr>
<td>Class 2†</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>8.5</td>
<td>17.1</td>
<td>4.1</td>
<td>15.7</td>
<td>50.2</td>
<td>61.4</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>99.8</td>
<td>99.6</td>
<td>&gt;99.9</td>
<td>&gt;99.9</td>
<td>97.8</td>
<td>97.5</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>80.7</td>
<td>74.6</td>
<td>92.3</td>
<td>95.8</td>
<td>63.9</td>
<td>65.0</td>
</tr>
<tr>
<td>NPV (%)</td>
<td>94.3</td>
<td>94.0</td>
<td>93.2</td>
<td>93.9</td>
<td>96.2</td>
<td>97.1</td>
</tr>
<tr>
<td>Class 3§</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>4.5</td>
<td>8.9</td>
<td>2.1</td>
<td>8.4</td>
<td>44.0</td>
<td>52.9</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>99.8</td>
<td>99.6</td>
<td>&gt;99.9</td>
<td>&gt;99.9</td>
<td>97.8</td>
<td>97.5</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>80.7</td>
<td>74.5</td>
<td>92.3</td>
<td>95.9</td>
<td>74.9</td>
<td>75.4</td>
</tr>
<tr>
<td>NPV (%)</td>
<td>87.7</td>
<td>88.2</td>
<td>87.4</td>
<td>88.2</td>
<td>92.3</td>
<td>93.4</td>
</tr>
</tbody>
</table>

* See table 1 for conditions included under codes.
† n = 4,064, probable and possible Sp cases excluded.
‡ n = 4,117, possible Sp cases excluded.
§ n = 4,385.

monia), and 481 (pneumococcal pneumonia) were high (>97.5 percent), PPVs of these codes were low relative to that of code 38.20 (pneumococcal septicemia).

The differences between NPVs of codes in position 1 and NPVs of codes in the first five positions did not exceed 1.6 percent. Having the least amount of false negatives (patients without the code but with definite Sp infection), code 481.00 (pneumococcal pneumonia) demonstrated the highest NPV.

Overall, using codes in any of the first five positions was preferable to using first position codes only, particularly because codes in positions 1–5 detected more pneumococcal pneumonia cases without substantial differences in specificity, PPV, and NPV.

**Comparison by case definition class**

Except for codes 38.00 (streptococcal septicemia) and 38.20 (pneumococcal septicemia), sensitivity remained stable (i.e., did not increase or decrease >10 percent) when the case definition expanded from Class 1 to Class 3 (table 2). PPV rose 10 percent or more for
codes 481 (pneumococcal pneumonia), 482.3 (streptococcal pneumonia), and 518.81 (respiratory failure) when case definition expanded from Class 1 to Class 3. However, as case definition expanded, NPV for each code dropped.

When probable and possible Sp cases are classified as non-Sp cases

The PPV and specificity of codes 481 (pneumococcal pneumonia), 38.20 (pneumococcal septicemia), and 482.30 (streptococcal pneumonia) decreased when probable and possible Sp cases were included as non-Sp cases in specificity, PPV, and NPV calculations for Class 1 cases (table 3); sensitivity did not change and NPV increased by 1.1 percent or less. Similarly, the PPV and specificity of codes 481, 38.20, and 482.30 fell when possible Sp cases were included in specificity, PPV, and NPV calculations for Class 2 cases; sensitivity remained the same and NPV increased by 1.1 percent or less.

Combining codes

To develop a strategy for maximizing the sensitivity and specificity of case ascertainment, we calculated the effects of different combinations of ICD-9-CM codes. We started with the most specific code, 38.20 (pneumococcal septicemia), and formed the following groups:

- Group 1 = 38.20;
- Group 2 = Group 1 + 481.00;
- Group 3 = Group 2 + 38.00;
- Group 4 = Group 3 + 482.30;
- Group 5 = Group 4 + 518.81;
- Group 6 = Group 5 + 486.00.

Using more ICD-9-CM codes to identify Sp pneumonia cases generally increased sensitivity and NPV, but lowered specificity and PPV (table 4 and figure 2). With a Class 1 case definition, Group 4 gave the highest set of values for sensitivity (81.25 percent) and specificity (96.05 percent), Group 1 provided the highest set of values for PPV (95.83 percent) and NPV (95.17 percent), and Group 3 had the highest set of values for sensitivity, specificity, PPV, and NPV (table 4). Group 3 and Group 4 provided the highest combined set of values for sensitivity, specificity, PPV, and NPV for Class 2 (table 4) and Class 3 (figure 2) case definitions, respectively.

Combinations of ICD-9-CM codes that other incidence and vaccine efficacy studies (10, 13) have used to identify Sp pneumonia patients were also evaluated. The combination of codes 481, 482.9 (bacterial pneumonia unspecified), 484.0–484.8 (pneumonia in infectious diseases classified elsewhere), 485 (bronchopneumonia, organism unspecified), 486, and 487 (influenza with pneumonia) in positions 1–5 had 71.3 percent sensitivity, 42.9 percent specificity, 7.3 percent PPV, and 96.0 percent NPV with a Class 1 case definition (probable and possible Sp patients excluded), and 68.3 percent sensitivity, 42.9 percent specificity, 14.9 percent PPV, and 90.2 percent NPV with a Class 3 case definition. With a Class 3 case definition, the combination of all pneumonia codes 480–487 in position 1 had 76.3 percent sensitivity, 31.1 percent specificity, 14.0 percent PPV, and 89.9 percent NPV, while in positions 1–5 such a definition had 90.0 percent sensitivity, 21.3 percent specificity, 14.4 percent PPV, and 93.6 percent NPV.

**TABLE 3.** Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the International Classification of Diseases, 9th revision, Clinical Modification (ICD-9-CM) codes when probable and possible Streptococcus pneumoniae (Sp) cases were classified as non-Sp cases in a population of hospitalized pneumonia patients in Franklin and Summit counties, Ohio, January 1991 to May 1992: position 1 vs. positions 1–5

<table>
<thead>
<tr>
<th>Class and test</th>
<th>ICD-9-CM code* and position no(s.)</th>
<th>38.00</th>
<th>38.20</th>
<th>481.00</th>
<th>482.30</th>
<th>486.00</th>
<th>518.81</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>1–5</td>
<td>1–5</td>
<td>1–5</td>
<td>1–5</td>
<td>1–5</td>
<td>1–5</td>
</tr>
<tr>
<td>Class 1†</td>
<td>Sensitivity (%)</td>
<td>10.4</td>
<td>20.4</td>
<td>5.0</td>
<td>19.2</td>
<td>45.4</td>
<td>58.3</td>
</tr>
<tr>
<td></td>
<td>Specificity (%)</td>
<td>99.9</td>
<td>99.6</td>
<td>&gt;99.9</td>
<td>99.9</td>
<td>94.7</td>
<td>93.9</td>
</tr>
<tr>
<td></td>
<td>PPV (%)</td>
<td>80.7</td>
<td>73.1</td>
<td>92.3</td>
<td>93.9</td>
<td>33.0</td>
<td>35.5</td>
</tr>
<tr>
<td></td>
<td>NPV (%)</td>
<td>95.1</td>
<td>95.6</td>
<td>94.8</td>
<td>95.5</td>
<td>96.8</td>
<td>97.5</td>
</tr>
<tr>
<td>Class 2‡</td>
<td>Sensitivity (%)</td>
<td>8.5</td>
<td>17.1</td>
<td>4.1</td>
<td>15.7</td>
<td>50.2</td>
<td>61.4</td>
</tr>
<tr>
<td></td>
<td>Specificity (%)</td>
<td>99.9</td>
<td>99.6</td>
<td>&gt;99.9</td>
<td>99.9</td>
<td>95.5</td>
<td>94.8</td>
</tr>
<tr>
<td></td>
<td>PPV (%)</td>
<td>80.7</td>
<td>74.6</td>
<td>92.3</td>
<td>93.9</td>
<td>44.6</td>
<td>45.7</td>
</tr>
<tr>
<td></td>
<td>NPV (%)</td>
<td>93.8</td>
<td>94.4</td>
<td>93.6</td>
<td>94.3</td>
<td>96.1</td>
<td>97.2</td>
</tr>
</tbody>
</table>

* See table 1 for conditions included under codes.
† n = 4,385, definite Sp cases.
‡ n = 4,385, definite and probable Sp cases.
TABLE 4. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the International Classification of Diseases, 9th revision, Clinical Modification (ICD-9-CM)* group codes for Streptococcus pneumoniae (Sp) pneumonia listed in the first five positions for hospitalized patients with community-acquired pneumonia, Franklin and Summit Counties, Ohio, January 1991 to May 1992: class 1 (definite Sp) and class 2 (definite and probable Sp) case definitions

<table>
<thead>
<tr>
<th>Class and test</th>
<th>ICD-9-CM code group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group 1</td>
</tr>
<tr>
<td>Class 1†</td>
<td></td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>19.17</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>99.95</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>95.83</td>
</tr>
<tr>
<td>NPV (%)</td>
<td>95.17</td>
</tr>
<tr>
<td>Class 2‡</td>
<td></td>
</tr>
<tr>
<td>Sensitivity (%)</td>
<td>15.70</td>
</tr>
<tr>
<td>Specificity (%)</td>
<td>99.95</td>
</tr>
<tr>
<td>PPV (%)</td>
<td>95.83</td>
</tr>
<tr>
<td>NPV (%)</td>
<td>93.93</td>
</tr>
</tbody>
</table>

* See table 1 for conditions included under codes and text for conditions included in groups.
† n = 4,064, probable and possible Sp cases excluded.
‡ n = 4,117, possible Sp cases excluded.

FIGURE 2. Groups of first five position ICD-9-CM codes for Streptococcus pneumoniae (Sp) using a Class 3 (definite, probable, and possible Sp) case definition among hospitalized community-acquired pneumonia patients of Franklin and Summit Counties, Ohio, January 1, 1991 to May 31, 1992. Group 1 = 38.20; Group 2 = Group 1 + 481.00; Group 3 = Group 2 + 38.00; Group 4 = Group 3 + 482.30; Group 5 = Group 4 + 518.81; and Group 6 = Group 5 + 486.00.

DISCUSSION

For hospitalized patients with community-acquired pneumococcal pneumonia, ICD-9-CM code 481 (pneumococcal pneumonia) is the most sensitive code regardless of position (position 1 vs. positions 1-5) or category of case definition (definite, probable, or possible Sp). In positions 1-5, code 481 detected 58 percent of patients with bacteremic pneumococcal pneumonia and 53 percent of patients with any microbiologic evidence of pneumococcal pneumonia. Despite instructions to coders to assign code 481 to patients diagnosed with lobar pneumonia due to an unspecified organism (14), the specificity of the code was fairly high. Only 2.5 percent of pneumonia patients with code 481 had no microbiologic evidence...
of \textit{Sp} disease. Code 481 appears to be useful in detecting cases of pneumococcal pneumonia when using large discharge diagnostic data sets; however, assuming vaccination status does not influence coding, the relatively low PPV of the code would lead to misclassification during vaccine efficacy studies and cause underestimation of vaccine efficacy.

The use of additional ICD-9-CM codes in positions 1–5 may further improve the sensitivity of code 481. Maximization of sensitivity would be the goal when the study population is small and medical charts can be reviewed to confirm pneumococcal etiology. Utilization of the ICD-9-CM codes as screening tools followed by medical chart review would be an accurate case ascertainment method that would lead to the least biased estimate of vaccine efficacy. Furthermore, having the ability to review medical charts allows the researcher to define the outcome of interest by the class of case definition and exclude probable and/or possible \textit{Sp} cases when necessary. When medical chart review is possible, Group 3 (codes 38.2, 481, and 38.0) with a Class 2 case definition (table 4) provides the optimal set of combined values for sensitivity (76.5 percent), specificity (97.0 percent), PPV (65.9 percent), and NPV (98.2 percent).

When review of medical charts is not feasible, there are at least two options to minimize misclassification. One option would involve selection of the most specific ICD-9-CM code or group of ICD-9-CM codes to identify pneumococcal pneumonia cases. For example, using code 38.20 in positions 1–5 to identify \textit{Sp} pneumonia patients would lead to a reasonable estimate of vaccine efficacy, because the PPV and specificity would be 95.9 percent and >99.9 percent, respectively (figure 2). However, only 8.4 percent of \textit{Sp} pneumonia cases would be detected. Thus, when only the most specific ICD-9-CM code is used to identify \textit{Sp} pneumonia patients, a very large population would be required to obtain an adequate number of \textit{Sp} pneumonia cases. Another option would be to use Group 4 (again restricted to the Class 3 case definition for the previously stated reason) to identify \textit{Sp} pneumonia patients. Of the six combination groups, Group 4 (codes 38.2, 481, 38.0, and 482.3) achieves the highest NPV and the highest combined set of values of sensitivity, specificity, PPV, and NPV.

The evaluation of ICD-9-CM codes in this study was limited to hospitalized patients who had illness that met a screening case definition and who were enrolled in a prospective study of community-acquired pneumonia. To evaluate the sensitivity of the screening process, a convenience sample of patients discharged with a diagnosis of pneumonia (ICD-9-CM codes 480–487 in any position) was taken from two study hospitals in Summit County and three study hospitals in Franklin County soon after patient enrollment was initiated. Of 258 patients identified by these codes, 159 were ineligible for the study (14 lived outside the study counties, 67 lived in institutions, 17 had no chest radiographs or no abnormal chest radiographs, and 61 had nosocomial pneumonia (symptom onset >48 hours after admission to hospital)). Of the 99 patients eligible for the study, 72 were entered, 18 had pneumonia but were not entered because of research technician or admission log error, seven were not included because they presented with symptoms that were not part of the screening definition of pneumonia (e.g., abdominal pain), and two were not included because they had pneumonia as an incidental finding on a routine chest radiograph obtained on admission for a non-pneumonic process or an unrelated procedure (e.g., elective cardiac catheterization). Thus, 97 patients had community-acquired pneumonia that required hospitalization, and the screening process demonstrated a potential sensitivity of 93 percent (90/97). The actual 74 percent (72/97) sensitivity of the screen prompted retraining of technicians; however, a follow-up sample was not obtained to evaluate the effect of retraining.

An important question raised by these data is how generalizable are the findings of this study to other areas of the country. While we have no basis to assume there were any geographic differences in coding practices, the results of this study may need confirmation in other sites. Because classification of case status occurred following retrospective chart review and the data collected in the community-acquired pneumonia study were not available to medical records staff, we do not believe that the study itself had any effect on coding practices. Coding practices could have differed among area hospitals; however, because of insufficient sample sizes at individual hospitals, we could not compare consistency of coding from one hospital to another.

Sputum specimens for culture and Gram's stain were collected at the discretion of the attending physician. The proportion of patients with bacteremia (43 percent) among those with any evidence of \textit{Sp} infection was higher than the proportion of 15–30 percent generally reported (15), which suggests that sputum culture and Gram's stain were underutilized among study patients.

The appearance of codes for pneumonia due to \textit{Streptococcus} (482.30), pneumonia, organism unspecified (486.00), and respiratory failure (518.81) among the most commonly used ICD-9-CM codes for definite pneumococcal pneumonia demonstrated the limitations of the coding system. Because the code for pneumonia due to \textit{Streptococcus} is supposed to include patients with non-pneumococcal streptococcal infec-

\textbf{Am J Epidemiol} Vol. 149, No. 3, 1999
tions and exclude patients with pneumococcal pneumonia, the frequency of this code among the patients with pneumococcal bacteremia is a measure of mis-coding. In addition, 34 patients with definite evidence of pneumococcal disease (culture from blood or pleural fluid) were coded with pneumonia, organism unspecified (table 1). Finding the code for streptococcal septicemia (instead of the code for pneumococcal septicemia) among patients with bacteremic pneumococcal pneumonia also demonstrates inaccuracy of the coding system (tables 1 and 2).

Marrie et al. (16) found inaccuracy with the ICD-9-CM codes in a small number of hospitalized patients and proposed alternative coding practices to improve the accuracy and precision of the coding system. One suggestion involved changing the coding format so that coders would be able to indicate whether blood culture was positive, negative, or not done when pneumonia is diagnosed. A second suggestion involved the development of an algorithm by which coders could indicate the absence or results of specific diagnostic tests. This could be done by excluding "lobar pneumonia, organism unspecified" from code 481 and classified these cases as code 486.0, "pneumonia, organism unspecified." Code 481.0 could be used to designate "S. pneumoniae pneumonia, blood culture positive" and code 481.1 for "S. pneumoniae pneumonia, sputum culture positive, blood culture negative or not done."

In addition to improved coding procedures, the need for better diagnostic tests for pneumococcal pneumonia is obvious. Accurate and economical diagnostic tests are required to optimize clinical therapy. Improved diagnostic testing will also form the foundation of epidemiologic studies to elucidate the incidence of community-acquired pneumococcal pneumonia and to evaluate the efficacy of pneumococcal vaccine for prevention of pneumococcal pneumonia.

REFERENCES