Epidemiol. Infect. (2011), **139**, 1296–1306. © Cambridge University Press 2010
doi:10.1017/S0950268810002529

Classification algorithms to improve the accuracy of identifying patients hospitalized with community-acquired pneumonia using administrative data

O. YU[1]*, J. C. NELSON[1,2], L. BOUNDS[3] AND L. A. JACKSON[3,4]

[1] *Biostatistics Unit, Group Health Research Institute, Seattle, WA, USA*
[2] *Department of Biostatistics, University of Washington, Seattle, WA, USA*
[3] *Group Health Research Institute, Seattle, WA, USA*
[4] *Department of Epidemiology, University of Washington, Seattle, WA, USA*

(*Accepted 13 October 2010; first published online 19 November 2010*)

SUMMARY

In epidemiological studies of community-acquired pneumonia (CAP) that utilize administrative data, cases are typically defined by the presence of a pneumonia hospital discharge diagnosis code. However, not all such hospitalizations represent true CAP cases. We identified 3991 hospitalizations during 1997–2005 in a managed care organization, and validated them as CAP or not by reviewing medical records. To improve the accuracy of CAP identification, classification algorithms that incorporated additional administrative information associated with the hospitalization were developed using the classification and regression tree analysis. We found that a pneumonia code designated as the primary discharge diagnosis and duration of hospital stay improved the classification of CAP hospitalizations. Compared to the commonly used method that is based on the presence of a primary discharge diagnosis code of pneumonia alone, these algorithms had higher sensitivity (81–98%) and positive predictive values (82–84%) with only modest decreases in specificity (48–82%) and negative predictive values (75–90%).

Key words: Accuracy, algorithm, classification and regression tree, community-acquired pneumonia, hospitalization.

INTRODUCTION

Pneumonia can cause substantial morbidity and mortality, particularly in older adults, but it can be a difficult outcome to define accurately for the purposes of epidemiological research. Administrative database studies often identify pneumonia hospitalizations using International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) diagnosis codes that are linked to hospitalization records. This method has been used to assess temporal trends [1–7], estimate disease burden [8], and identify risk factors [9] for pneumonia hospitalizations. It has also been used in evaluations of influenza vaccine effectiveness [10–12] and in pharmaco-epidemiological studies of the association of use of various medications with risk of pneumonia [13–16].

However, the assignment of a pneumonia hospital discharge diagnosis code lacks specificity for the true occurrence of community-acquired pneumonia (CAP), because pneumonia codes may be assigned to hospitalizations that are not truly associated with pneumonia. Furthermore, the coding system does not distinguish between CAP and nosocomial pneumonia. Some studies have employed additional methods, such

* Author for correspondence: Ms. O. Yu, Group Health Research Institute, 1730 Minor Ave, Suite 1600, Seattle, WA 98101, USA (Email: yu.o@ghc.org)

https://doi.org/10.1017/S0950268810002529 Published online by Cambridge University Press

as medical record review of hospitalizations assigned a pneumonia diagnosis code [17, 18], to validate the occurrence of CAP. However, chart review is very labour intensive and so is often not feasible. For this reason there is a need to develop methods that more accurately identify cases of CAP utilizing information readily available from administrative data sources.

In this study, we identified 3991 hospitalizations assigned a pneumonia discharge diagnosis code in persons of all ages from the administrative data systems within a managed care organization. For each hospitalization identified, we accessed from a prior study the true CAP status as determined by medical record review [19]. We then developed classification algorithms that used additional administrative information associated with the hospitalization to more accurately identify true CAP cases compared to the commonly used method of identifying cases presumptively based on a pneumonia diagnosis code alone. We evaluated the accuracy of the newly developed classification algorithms by comparing the cases classified as CAP or not CAP by the algorithms against the gold standard definition of CAP based on medical record review.

METHODS

Study population

The study was conducted among enrollees of Group Health Cooperative, a managed care organization in Washington State with administrative data systems that record information on ICD-9-CM codes assigned to in-patient and outpatient medical encounters. In a previous study of trends in pneumonia rates over time in this study population [19], these administrative data systems were used to identify outpatient medical encounters and hospitalizations with a pneumonia ICD-9-CM code (480–487.0 and 507.0) assigned to any diagnosis field from January 1997 to January 2005, among persons of all ages. Since the main focus was to compare the pneumonia rates before and after the introduction of the 7-valent pneumococcal conjugate vaccine (PCV7) in 2000, only the more recent and shorter pre-PCV7 time period since 1998 was used and reported in the previous study. In the current study, we used a slightly broader study period and included pneumonia-coded hospitalizations that occurred during January 1997 to January 2005. The study was approved by the Group Health Institutional Review Board.

Chart review to determine true cases of CAP

In a previous study [19], after presumptive pneumonia hospitalizations were identified using ICD-9-CM codes, medical record review was conducted to determine if such hospitalizations were true cases of CAP. Specifically, two chart abstractors (one was a nurse and the other was specifically trained for the abstraction in a previous study [19]) were trained extensively by an expert epidemiologist to review the hospitalization records forwarded by the treating non-Group Health facility and available in the Group Health outpatient medical record, which typically involved review of the hospital discharge summary. Chart abstractors were not given the information regarding the pneumonia ICD-9-CM codes recorded in the administrative database, and they were not trained to look for specific ICD-9-CM codes in the medical record. Instead, they were trained to look for the clinical diagnosis from the treating physician. To assess abstractor reliability and to ensure data quality, all the chart reviews conducted in the first several months of the study were reviewed by both abstractors, and a 10% random sample of the remaining charts were reviewed by both thereafter.

A hospitalization was defined as 'definite CAP' if there was documentation in the records that the treating physician considered pneumonia as the most likely cause of the illness present at admission. A hospitalization was defined as 'probable CAP' if the physician considered pneumonia as a possible cause of the illness present at admission. Hospitalizations due to causes other than pneumonia, including nosocomial pneumonia with pneumonia symptoms onset after hospital admission, were included as 'not CAP'. In the current study, the 'gold standard' definition of true CAP was a pneumonia-coded hospitalization that was determined to be definite or probable CAP by medical record review. Furthermore, those hospitalizations with insufficient information available for review or no physician assessment record available in the medical record were excluded. If an individual had more than one pneumonia-coded hospitalization during the study period, only the first such hospitalization was included in the current study.

Administrative predictors of CAP

We identified several additional potential predictors of CAP from administrative data sources that were linked to the pneumonia-coded hospitalizations, and we evaluated them for possible inclusion in the

https://doi.org/10.1017/S0950268810002529 Published online by Cambridge University Press

classification algorithms (Table 1). These variables included the presence of specific individual diagnostic pneumonia codes (*vs*. the presence of any code within the 480–487·0 and 507·0 code group), whether they were assigned as the primary discharge diagnosis or as any discharge diagnosis, other illness diagnosis codes, whether they were assigned as the primary or any discharge diagnosis, procedure codes, age, and length of hospital stay.

Statistical analysis

Our primary aim was to develop a method that could more accurately identify true cases of CAP using administrative information compared to the commonly used approach of identifying cases simply based on the presence of a pneumonia discharge diagnosis code alone. To accomplish this, we considered a wider array of administrative information associated with the apparent pneumonia hospitalization and potentially predictive of true CAP. We then developed an algorithm to identify true CAP cases using this additional information in a classification and regression tree (CART) analysis [20]. CART is a binary recursive partitioning method that builds a decision tree that classifies individuals as having or not having a particular outcome on the basis of a set of predictors. In our analysis, the outcome of interest is true CAP, where CAP was defined using the gold standard definition based on medical chart review, and the candidate predictor variables are the administrative variables described in the previous section. A separate classification algorithm was developed for each of the three age groups (0–17, 18–64, $\geqslant$65 years). The CART analysis was performed using 'rpart' in the R package [21], which includes as part of its standard implementation a tenfold cross-validation process for internal validation of the final tree.

To evaluate the performance of each newly developed classification algorithm with respect to its ability to identify true cases of CAP from all pneumonia-coded hospitalizations, we compared the cases classified as CAP or not CAP by the algorithms against the gold standard definition of CAP based on medical record review. We calculated sensitivity (percentage of true CAP hospitalizations as determined from medical record review that were correctly classified as CAP by our algorithm), specificity (percentage of hospitalizations determined not to be true CAP cases by medical record review that were correctly classified as not CAP by our algorithm), positive predictive value (PPV) (percentage of hospitalizations classified as CAP by our algorithm that were determined as true CAP by medical record review), negative predictive value (NPV) (percentage of hospitalizations classified as not CAP by our algorithm that were determined as not true CAP cases by medical record review) and their 95% confidence intervals (CI) based on a binomial distribution. For further comparison, we also calculated these performance measures for a simpler method that classified a hospitalization as CAP if it was assigned a pneumonia code (480–487.0 or 507.0) as the primary discharge diagnosis.

RESULTS

Over 10 000 hospitalizations that were assigned a pneumonia ICD-9-CM code during the study period were identified for chart review. Chart review was complete on 6938 hospitalizations and 2947 of these hospitalizations were excluded from the current study because there was no physician assessment record available in the medical record or they were recurrent events. As a result, a total of 3991 hospitalizations that were assigned a pneumonia ICD-9-CM code during the study period and had a gold standard CAP status (true CAP cases *vs*. not CAP) determined based on medical record review were included in the current study. Of those, 2491 (62%) were determined to be true CAP cases by medical record review (Table 2). The proportion of pneumonia-coded hospitalizations determined to be true CAP by medical record review varied by age group: 74% in persons aged 0–17 years, 54% in persons aged 18–64 years, and 65% in persons aged $\geqslant$65 years. Of the true CAP cases, 80% were defined as definite CAP and 20% as probable CAP.

Nosocomial pneumonia events were more common in adults aged $\geqslant$18 years compared to children. In all pneumonia-coded hospitalizations, the percentages determined as nosocomial pneumonia by medical record review were 6% in persons aged 0–17 years, 21% in persons aged 18–64 years, and 17% in persons aged $\geqslant$65 years.

In a previous study [19], it was determined that the agreement on the true CAP status between chart reviews by the two abstractors was high with 88% agreement and a kappa of 0·75 (95% CI 0·70–0·80).

Classification algorithms

A separate classification algorithm to identify true CAP hospitalizations based on administrative information

https://doi.org/10.1017/S0950268810002529 Published online by Cambridge University Press

https://doi.org/10.1017/S0950268810002529 Published online by Cambridge University Press

Table 1. *Selected variables associated with pneumonia-coded hospitalizations that were evaluated as potential predictors of CAP in the CART analyses, by age group and the true CAP status from medical record review. All pneumonia-coded hospitalizations had a discharge diagnosis code of 480–487.0 or 507.0 during January 1997 to January 2005*

Variables associated with pneumonia-coded hospitalization*	Position discharge diagnosis assigned	Hospitalizations with variable (%)					
		0–17 years		18–64 years		⩾65 years	
		CAP ($n=112$)	Not CAP ($n=40$)	CAP ($n=514$)	Not CAP ($n=440$)	CAP ($n=1865$)	Not CAP ($n=1020$)
Pneumonia diagnosis (ICD-9-CM code)							
Pneumonia diagnosis code (480 to 487.0 or 507.0)	Primary	66	30	63	8	65	15
Pneumonia diagnosis code (480 to 487.0)	Primary	63	30	59	5	51	8
	Any	96	70	91	69	81	58
Pneumonia, organism unspecified (486)	Primary	40	8	44	4	43	7
	Any	69	25	69	51	70	49
Pneumonitis due to inhalation of food or vomitus (507.0)	Primary	4	0	5	2	14	6
	Any	5	30	10	33	20	44
Viral pneumonia (480.x)	Primary	13	18	2	0	1	0
	Any	14	20	2	0	1	0
Other bacterial pneumonia (482.x)	Primary	4	5	6	1	4	1
	Any	8	23	11	14	7	8
Other diagnosis (ICD-9-CM code)							
Acute respiratory failure (518.81)	Primary	1	3	6	3	5	5
	Any	4	13	11	19	9	16
Asthma (493)	Primary	15	5	1	1	0	0
	Any	46	23	15	6	5	4
Chronic lung disease (491–492)	Primary	0	0	1	1	2	1
	Any	0	0	13	4	13	8
Congestive heart failure (428)	Primary	0	0	2	1	3	4
	Any	1	3	13	15	35	34
Cancer (140–172 or 174–208)	Primary	0	0	3	12	1	7
	Any	1	0	13	19	10	14
Injury (800–959)	Primary	0	20	0	8	1	10
	Any	0	23	2	9	4	13
External causes of injury and poisoning (any E code)		1	15	4	4	2	5

https://doi.org/10.1017/S0950268810002529 Published online by Cambridge University Press

Table 1 (*cont.*)

Variables associated with pneumonia-coded hospitalization*	Position discharge diagnosis assigned	Hospitalizations with variable (%)					
		0–17 years		18–64 years		⩾65 years	
		CAP ($n=112$)	Not CAP ($n=40$)	CAP ($n=514$)	Not CAP ($n=440$)	CAP ($n=1865$)	Not CAP ($n=1020$)
Procedure codes							
Any procedure code (01–86)		19	58	48	85	37	74
Mechanical ventilation (96.7x)		4	38	13	32	9	24
Any operation (01–86)		10	45	32	70	20	57
Chest operations and procedures (32–34)		2	8	14	17	6	10
Demographics							
Female		45	25	52	43	50	51
Length of hospital stay†							
0–2 days		54	20	29	12	30	12
3–4 days		28	20	30	15	33	21
⩾5 days		19	60	41	73	37	67
Season of hospital admission							
Winter (Jan.–Mar.)		43	38	33	27	34	26
Spring (Apr.–June)		23	28	24	19	21	26
Summer (July–Sep.)		15	23	18	26	18	21
Autumn (Oct.–Dec.)		19	13	25	28	27	27
Death from the admitting illness		0	3	8	15	16	25

CAP, Community-acquired pneumonia; CART, classification and regression tree.

* Other variables evaluated as potential predictors included year of age and year of hospital admission. Other discharge diagnosis or procedure codes also evaluated in the CART analyses but not shown in the table due to low overall prevalence (<5%) included viral pneumonia (480.x), pneumonia due to respiratory syncytial virus (480.1), pneumococcal pneumonia (481), pneumonia due to other specified bacteria (483.x), pneumonia in infectious diseases classified elsewhere (484.x), bronchopneumonia (485), influenza with pneumonia (487.0), croup (464.4), acute bronchitis (490.x), bronchoscopy (33.2x), thoracentesis (34.91), chest radiograph (87.44 or 87.49), and thoracic CAT scan (87.41).

† Length of hospital stay was evaluated as a continuous variable in the CART analyses.

Table 2. *Results from medical record review of the pneumonia-coded hospitalizations during January 1997 to January 2005*

	Age group (years)							
	0–5	6–17	18–49	50–64	65–74	75–84	⩾85	All
No. of pneumonia-coded hospitalizations	92	60	325	629	668	1227	990	3991
Results from medical record review (%)								
True CAP cases	80	63	54	54	59	63	70	62
Definite CAP	73	50	48	44	45	50	55	50
Probable CAP	7	13	6	10	14	13	15	12
Not CAP	20	37	46	46	41	37	30	38
Nosocomial pneumonia	3	10	17	24	19	17	14	17
Other	17	27	29	22	22	20	16	21

CAP, Community-acquired pneumonia.

was developed for each of the three age groups (0–17, 18–64, ⩾65 years). A wide variety of variables was evaluated as potential predictors of CAP in the CART analyses (Table 1). As shown in Table 1, across all three age groups, the presence of a pneumonia code assigned as the primary discharge diagnosis and the presence of specific pneumonia discharge diagnosis codes of 486 (pneumonia, organism unspecified) or 480–487 were more prevalent in true CAP hospitalizations compared to those that were not determined to be true CAP from medical record review. In contrast, hospitalizations that were not determined to be true CAP by medical record review were more likely to have a code for pneumonitis due to inhalation of food or vomitus (507.0), an injury code, or a procedure code associated with the pneumonia-coded hospitalization. The length of hospital stay was shorter in true CAP hospitalizations than those that were not CAP.

The predictors ultimately selected for the classification algorithm and the order in which those predictors appear in the decision tree varied by age group. In the 0–17 years age group, the strongest predictor of CAP was length of hospital stay of no more than 3 days (Fig. 1). Of the 152 pneumonia-coded hospitalizations that occurred in that age group, 89 had a length of hospital stay of ⩽3 days and they were classified as CAP by the algorithm. Of those 89, only eight (9%) were not true CAP cases according to the gold standard medical record review and so were misclassified by the algorithm. The remaining 63 hospitalizations of >3 days' duration were further classified as CAP or not CAP by the following additional variables: an injury discharge diagnosis code assigned to the hospitalization in any position, a pneumonia code of 486 assigned in any position, and any pneumonia code (480–487·0 or 507·0) assigned as the primary discharge diagnosis.

For the other two age groups (18–64 and ⩾65 years), the strongest predictor of CAP was the presence of any pneumonia code (480–487·0 or 507·0) as the primary discharge diagnosis (Figs 2 and 3). Only 9% of hospitalizations with this designation in persons aged 18–64 years and 11% in persons aged ⩾65 years were not truly CAP as determined by medical record review and were thus misclassified by the algorithm. In both age groups, the other variables found to be predictive of CAP and therefore included in the classification algorithms were the length of hospital stay and the absence of a discharge diagnosis code of pneumonia due to inhalation of food or vomitus (507·0) in any position. A primary discharge diagnosis code of acute respiratory failure (518.81) was also found to be predictive of CAP for persons aged 18–64 years, whereas the presence of any operation procedure code assigned to the hospitalization was included in the algorithm for persons aged ⩾65 years.

Performance of the classification algorithms

The performance of the classification algorithms was evaluated by comparing the classification of CAP by the algorithms to the gold standard definition of CAP according to medical record review (Table 3). Across the algorithms for the three age groups, sensitivity was 81–98%, specificity was 48–82%, PPV was 82–84% and NPV was 75–90%.

Relative to the classification algorithms we derived using CART, a simpler approach for identifying a

https://doi.org/10.1017/S0950268810002529 Published online by Cambridge University Press

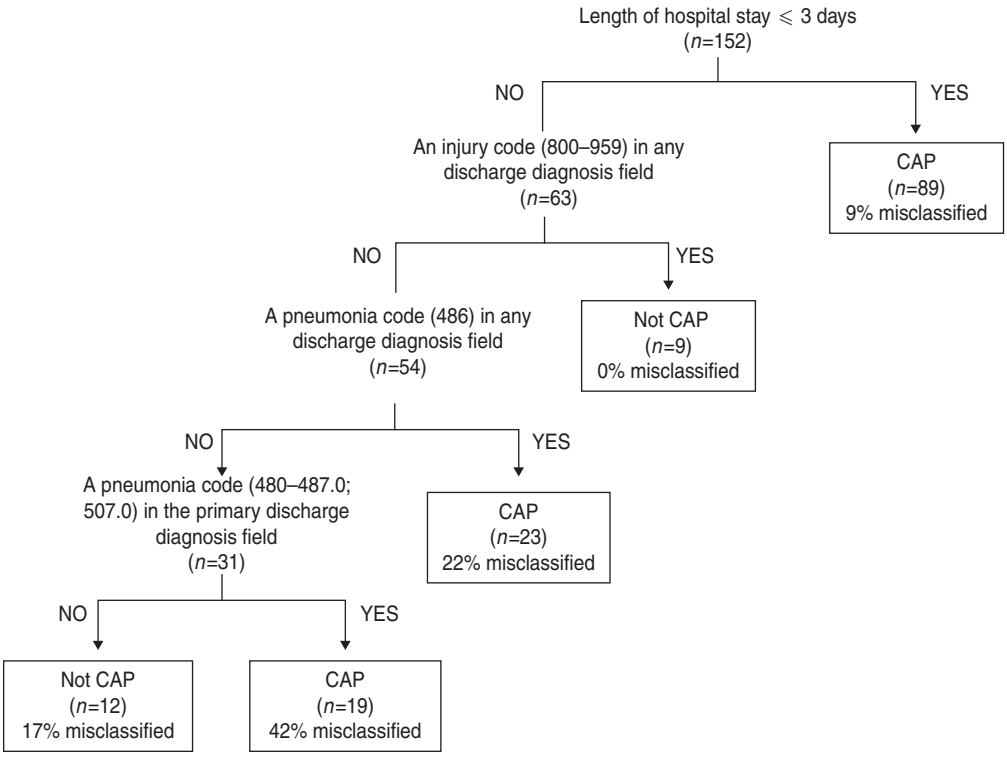

Fig. 1. Classification algorithm for community-acquired pneumonia (CAP) hospitalization in persons aged 0–17 years of age during January 1997 to January 2005.

CAP hospitalization is to use the presence of a discharge diagnosis code of pneumonia (480–487.0 or 507.0) in the primary position. This approach resulted in a sensitivity of 63–66%, specificity of 70–93%, PPV of 86–91% and NPV of 42–68% across the three age groups.

DISCUSSION

We included both children and adults in our study and developed three age group-specific algorithms for classifying hospitalized patients who were presumptively identified by ICD-9-CM pneumonia codes as truly having or not having CAP. Compared to a commonly used approach of identifying a CAP hospitalization based on the presence of a primary discharge diagnosis code of pneumonia alone, these algorithms that incorporated additional and readily available administrative information, such as the length of hospital stay and discharge codes for diagnoses other than pneumonia, increased sensitivity by 18–32% and NPV by 10–48% but reduced specificity by a modest 11–22% and PPV by a small 2–7% depending on age group. These algorithms were highly sensitive which may be useful in situations in which the primary objective is to identify true cases of CAP. For example, in a case-control study of seniors hospitalized with CAP, a highly sensitive algorithm could be used as a first step in identifying presumptive CAP hospitalizations that have a high likelihood of being true cases for further validation by chart review or other methods.

Several previous studies have examined the accuracy of pneumonia diagnosis codes for identifying patients with pneumonia [22–27]. In comparison to our findings when CAP was defined by the presence of a primary discharge diagnosis code of pneumonia alone, three studies found a similar sensitivity [23–25] and another three studies found a higher sensitivity of 76–98% [22, 26, 27]. Two studies also found a higher specificity (97% and 99%) [24, 27]. Our PPV of 86–91% was similar to or higher than that in previous studies. However, our NPV was relatively lower (42–68% *vs.* 74–98·2%) [22, 24, 26, 27].

Using CART analysis to develop algorithms for identifying true cases of CAP has several advantages. First, CART is a non-parametric method that makes no assumptions about the underlying distributions of the predictors or the relationships between the predictors and CAP. In addition, CART analysis

https://doi.org/10.1017/S0950268810002529 Published online by Cambridge University Press

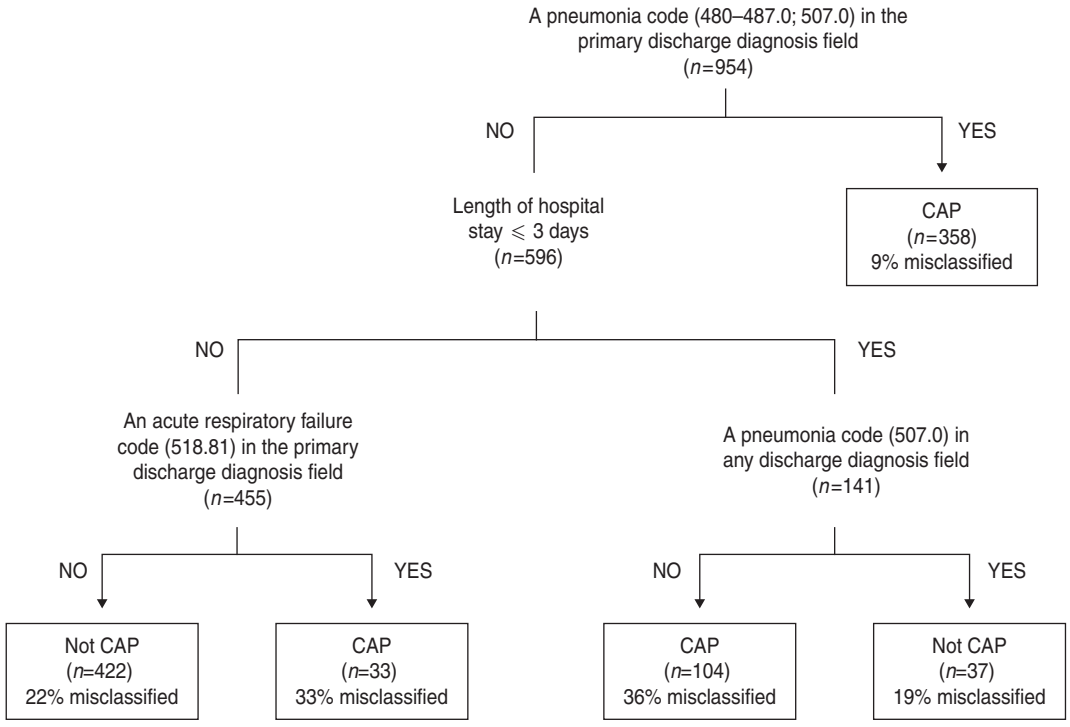

Fig. 2. Classification algorithm for community-acquired pneumonia (CAP) hospitalization in persons aged 18–64 years during January 1997 to January 2005.

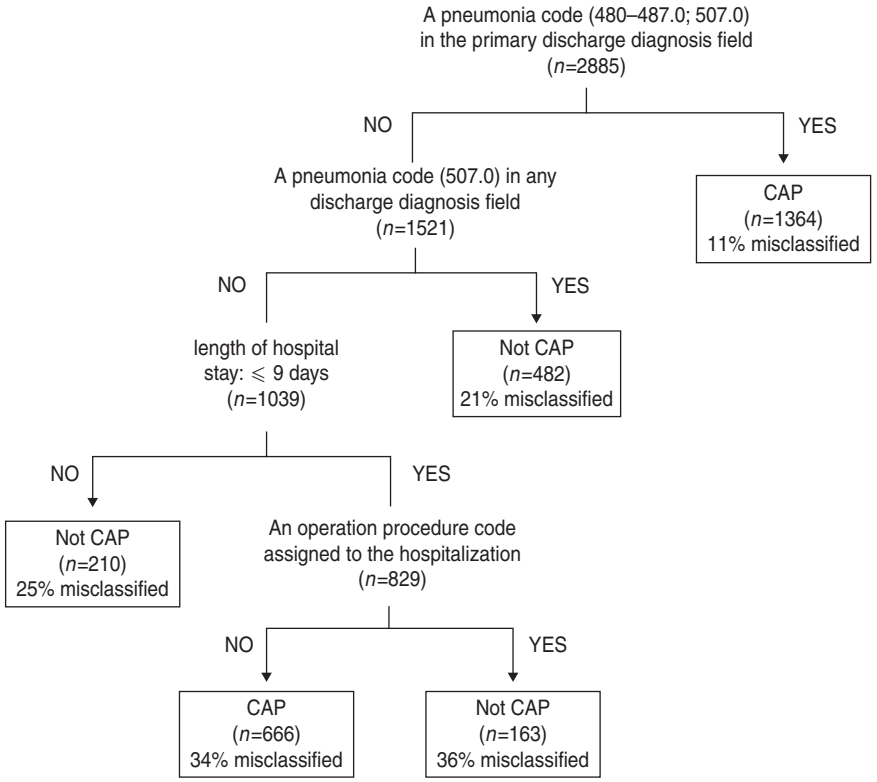

Fig. 3. Classification algorithm for community-acquired pneumonia (CAP) hospitalization among persons aged ⩾65 years during January 1997 to January 2005.

https://doi.org/10.1017/S0950268810002529 Published online by Cambridge University Press

https://doi.org/10.1017/S0950268810002529 Published online by Cambridge University Press

Table 3. *Performance of classification methods for CAP hospitalizations during January 1997 to January 2005, by age group*

Age group (yr)	Classification based on primary discharge diagnosis code of pneumonia (480–487.0 or 507.0)					Classification based on algorithms from CART analyses			
		Gold standard					Gold standard		
(*a*) 0–17 years		CAP	Not CAP	Total			CAP	Not CAP	Total
Discharge diagnosis code	CAP	74	12	86	Algorithm	CAP	110	21	131
	Not CAP	38	28	66		Not CAP	2	19	21
	Total	112	40	152		Total	112	40	152
	Sensitivity (95% CI) = 66% (57–75) Specificity (95% CI) = 70% (53–83) PPV (95% CI) = 86% (77–93) NPV (95% CI) = 42% (30–55)					Sensitivity (95% CI) = 98% (94–100) Specificity (95% CI) = 48% (32–64) PPV (95% CI) = 84% (77–90) NPV (95% CI) = 90% (70–99)			
		Gold standard					Gold standard		
(*b*) 18–64 years		CAP	Not CAP	Total			CAP	Not CAP	Total
Discharge diagnosis code	CAP	325	33	358	Algorithm	CAP	414	81	495
	Not CAP	189	407	596		Not CAP	100	359	459
	Total	514	440	954		Total	514	440	954
	Sensitivity (95% CI) = 63% (59–67) Specificity (95% CI) = 93% (90–95) PPV (95% CI) = 91% (87–94) NPV (95% CI) = 68% (64–72)					Sensitivity (95% CI) = 81% (77–84) Specificity (95% CI) = 82% (78–85) PPV (95% CI) = 84% (80–87) NPV (95% CI) = 78% (74–82)			
		Gold standard					Gold standard		
(*c*) ⩾65 years		CAP	Not CAP	Total			CAP	Not CAP	Total
Discharge diagnosis code	CAP	1214	150	1364	Algorithm	CAP	1654	376	2030
	Not CAP	651	870	1521		Not CAP	211	644	855
	Total	1865	1020	2885		Total	1865	1020	2885
	Sensitivity (95% CI) = 65% (63–67) Specificity (95% CI) = 85% (83–87) PPV (95% CI) = 89% (87–91) NPV (95% CI) = 57% (55–60)					Sensitivity (95% CI) = 89% (87–90) Specificity (95% CI) = 63% (60–66) PPV (95% CI) = 82% (80–83) NPV (95% CI) = 75% (72–78)			

CAP, Community-acquired pneumonia; CART, classification and regression tree; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

implicitly considers all possible interactions between potential predictors as well as all possible dichotomous cut-off points for categorical or continuous predictors when building the algorithm. A tenfold cross-validation provides for efficient internal validation of the algorithm. Finally, the classification algorithms can be depicted visually in a simple fashion which allows for straightforward interpretation of results and provides for easy application in other settings.

Our study has several limitations that are important to note. First, we observed relatively few CAP hospitalizations in children and thus cannot draw strong conclusions about the performance of our algorithm in that population. Second, since the algorithms were developed in a population of insured Group Health enrollees, they may not be readily generalizable to population in other settings. However, enrollees were hospitalized at a large number of hospitals across western Washington and as a result our study was not confined to hospitalizations occurring in a single institution. Third, our gold-standard CAP definition was only based on physician's clinical diagnosis reported in medical records and not chest radiographs or laboratory-confirmed results since such records were not forwarded to Group Health by the treating non-Group Health facility. Therefore, this method may not have captured all the true CAP cases, which could bias our results. Specifically, if the algorithms we developed also failed to identify these same true CAP cases, the sensitivity of the algorithms would be overestimated. Moreover, if the algorithms picked up the missed true CAP cases, the specificity would be underestimated. Last, we did not ascertain cases of CAP with a pneumonia diagnosis documented in medical record but not in the administrative database. Thus, CAP hospitalizations that were not identified by a pneumonia discharge diagnosis code remained undetected. In conclusion, we developed age group-specific classification algorithms for identifying CAP hospitalizations using information such as the length of the hospital stay and presence of a primary pneumonia discharge diagnosis code from administrative data sources. These algorithms had higher sensitivity and NPV but a relatively modest decrease in specificity and PPV compared to a more commonly used method based on the presence of a primary discharge diagnosis code of pneumonia alone. These algorithms could be readily implemented in future epidemiological studies with similar available administrative data and would result in more accurate identification of CAP hospitalizations.

ACKNOWLEDGEMENTS

This study was financially supported by Group Health Research Institute internal fund.

DECLARATION OF INTEREST

All authors were employees of Group Health Research Institute at the time of the study. The authors and Group Health Research Institute have no financial interests that are affected by the material in this paper.

REFERENCES

1. **Fry AM, *et al.*** Trends in hospitalizations for pneumonia among persons aged 65 years or older in the United States, 1988–2002. *Journal of the American Medical Association* 2005; **294**: 2712–2719.
2. **Poehling KA, *et al.*** Population-based impact of pneumococcal conjugate vaccine in young children. *Pediatrics* 2004; **114**: 755–761.
3. **De Wals P, *et al.*** Pneumonia after implementation of the pneumococcal conjugate vaccine program in the province of Quebec, Canada. *Pediatric Infectious Disease Journal* 2008; **27**: 963–968.
4. **Grijalva CG, *et al.*** Decline in pneumonia admissions after routine childhood immunisation with pneumococcal conjugate vaccine in the USA: a time-series analysis. *Lancet* 2007; **369**: 1179–1186.
5. **Zhou F, *et al.*** Health care utilization for pneumonia in young children after routine pneumococcal conjugate vaccine use in the United States. *Archives of Pediatrics & Adolescent Medicine* 2007; **161**: 1162–1168.
6. **Thomsen RW, *et al.*** Rising incidence and persistently high mortality of hospitalized pneumonia: a 10-year population-based study in Denmark. *Journal of Internal Medicine* 2006; **259**: 410–417.
7. **Baine WB, Yu W, Summe JP.** Epidemiologic trends in the hospitalization of elderly Medicare patients for pneumonia, 1991–1998. *American Journal of Public Health* 2001; **91**: 1121–1123.
8. **Ellis SE, *et al.*** Influenza- and respiratory syncytial virus-associated morbidity and mortality in the nursing home population. *Journal of the American Geriatrics Society* 2003; **51**: 761–767.
9. **Kornum JB, *et al.*** Diabetes, glycemic control, and risk of hospitalization with pneumonia: a population-based case-control study. *Diabetes Care* 2008; **31**: 1541–1545.
10. **Nichol KL, *et al.*** Effectiveness of influenza vaccine in the community-dwelling elderly. *New England Journal of Medicine* 2007; **357**: 1373–1381.

https://doi.org/10.1017/S0950268810002529 Published online by Cambridge University Press

11. **Nordin J, *et al.*** Influenza vaccine effectiveness in preventing hospitalizations and deaths in persons 65 years or older in Minnesota, New York, and Oregon: data from 3 health plans. *Journal of Infectious Diseases* 2001; **184**: 665–670.
12. **Jackson LA, *et al.*** Evidence of bias in estimates of influenza vaccine effectiveness in seniors. *International Journal of Epidemiology* 2006; **35**: 337–344.
13. **Thomsen RW, *et al.*** Preadmission use of statins and outcomes after hospitalization with pneumonia: population-based cohort study of 29,900 patients. *Archives of Internal Medicine* 2008; **168**: 2081–2087.
14. **Schlienger RG, *et al.*** Statins and the risk of pneumonia: a population-based, nested case-control study. *Pharmacotherapy* 2007; **27**: 325–332.
15. **Sarkar M, Hennessy S, Yang YX.** Proton-pump inhibitor use and the risk for community-acquired pneumonia. *Annals of Internal Medicine* 2008; **149**: 391–398.
16. **Hennessy S, *et al.*** Observed association between antidepressant use and pneumonia risk was confounded by comorbidity measures. *Journal of Clinical Epidemiology* 2007; **60**: 911–918.
17. **Jackson ML, *et al.*** Influenza vaccination and risk of community-acquired pneumonia in immunocompetent elderly people: a population-based, nested case-control study. *Lancet* 2008; **372**: 398–405.
18. **Jackson LA, *et al.*** Effectiveness of pneumococcal polysaccharide vaccine in older adults. *New England Journal of Medicine* 2003; **348**: 1747–1755.
19. **Nelson JC, *et al.*** Impact of the introduction of pneumococcal conjugate vaccine on rates of community acquired pneumonia in children and adults. *Vaccine* 2008; **26**: 4947–4954.
20. **Breiman L, *et al.*** *Classification and Regression Trees*. Boca Raton, FL: Chapman & Hall/CRC, 1984.
21. **R Development Core Team.** R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2008 (www.R-project.org).
22. **Guevara RE, *et al.*** Accuracy of ICD-9-CM codes in detecting community-acquired pneumococcal pneumonia for incidence and vaccine efficacy studies. *American Journal of Epidemiology* 1999; **149**: 282–289.
23. **Marrie TJ, Durant H, Sealy E.** Pneumonia – the quality of medical records data. *Medical Care* 1987; **25**: 20–24.
24. **Aronsky D, *et al.*** Accuracy of administrative data for identifying patients with pneumonia. *American Journal of Medical Quality* 2005; **20**: 319–328.
25. **van de Garde EM, *et al.*** International classification of diseases codes showed modest sensitivity for detecting community-acquired pneumonia. *Journal of Clinical Epidemiology* 2007; **60**: 834–838.
26. **Whittle J, *et al.*** Community-acquired pneumonia: can it be defined with claims data? *American Journal of Medical Quality* 1997; **12**: 187–193.
27. **Skull SA, *et al.*** ICD-10 codes are a valid tool for identification of pneumonia in hospitalized patients aged $\geqslant$65 years. *Epidemiology and Infection* 2008; **136**: 232–240.

https://doi.org/10.1017/S0950268810002529 Published online by Cambridge University Press