Methods Articles

Development and Validation of a Predictive Algorithm to Identify Adult Asthmatics from Medical Services and Pharmacy Claims Databases

Yuko Kawasumi, Michal Abrahamowicz, Pierre Ernst, and Robyn Tamblyn

Objective. To develop and validate the accuracy of a predictive model to identify adult asthmatics from administrative health care databases.

Study Setting. An existing electronic medical record project in Montreal, Quebec.

Study Design. One thousand four hundred and thirty-one patients with confirmed asthma status were identified from primary care physician's electronic medical record.

Data Collection/Extraction Methods. Therapeutic indication of asthma in an electronic prescription and/or confirmed asthma from an automated problem list were used as the gold standard. Five groups of asthma-specific markers were identified from administrative health care databases to estimate the probability of the presence of asthma. Cross-validation evaluated the diagnostic ability of each predictive model using 50 percent of sample.

Principal Findings. The best performance in discriminating between the patients with asthma and those without it included indicators from medical service and prescription claims databases. The best-fitting algorithm had a sensitivity of 70 percent, a specificity of 94 percent, and positive predictive value of 65 percent. The prescriptions claims-specific algorithm demonstrated a nearly equal performance to the model with medical services and prescription claims combined.

Conclusions. Our algorithm using asthma-specific markers from administrative claims databases provided moderate sensitivity and high specificity.

Key Words. Asthma/epidemiology, algorithms, databases, health services, medical record system, computerized

Asthma is a chronic health condition, which results in avoidable exacerbations of disease and unnecessary health care expenditures (Seung and Mittmann 2005). Despite the development and dissemination of asthma clinical guidelines for over a decade, optimal care has not been achieved (Boulet et al. 2001).

While hospital discharge databases have been frequently used for epidemiological research on asthma, this necessarily only provides information on the more severe and most poorly controlled patients (Wilchesky, Tamblyn, and Huang 2004). On the other hand, population-based administrative health databases provide the opportunity to identify factors associated with sub-optimal care of asthma in ambulatory care.

In Canada, ambulatory data are routinely collected and available through population-based insurance and reimbursement databases. Each province administers a universal health plan for all eligible residents. Payment of physicians on a fee-for-service basis provides a record of each service delivered to a beneficiary. In addition, provincial-specific drug insurance programs provide information on drugs dispensed from community pharmacists for each beneficiary. These data can be used to create a longitudinal record of medical services and prescribed drug use by linking information for an individual using an encrypted health care number.

One of the challenges in using such administrative databases to identify predictors of poor health outcomes is the difficulty in accurately identifying patients with asthma. Several epidemiological studies have used the reason for the medical visit, coded using the International Classification of Diseases 9th version (ICD-9) diagnostic codes in the medical service claims database, as a marker of physician-diagnosed asthma (Smith, Rascati, and McWilliams 2004). However, the validity of the ICD-9 asthma diagnosis codes in the Canadian medical services claims data have been shown to have poor-to-moderate sensitivity in spite of good specificity (Kozyrskyj, Mustard, and Becker 2004; Wilchesky, Tamblyn, and Huang 2004). In contrast, when the utilization of medication from the prescription claims database is used as a marker of asthma, sensitivity is drastically improved (Osborne et al. 1995; Pont et al. 2002; Abrahamowicz et al. 2007).

Address correspondence to Yuko Kawasumi, Ph.D., Centre for Health Services and Policy Research, University of British Columbia, 201-2206 East Mall, Vancouver, BC, Canada V6T 1Z3; e-mail: ykawasumi@chspr.ubc.ca. Yuko Kawasumi, Ph.D., Michal Abrahamowicz, Ph.D., and Robyn Tamblyn, Ph.D., are with the Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, QC, Canada. Michal Abrahamowicz, Ph.D., and Robyn Tamblyn, Ph.D., are with the Department of Medicine, McGill University Health Centre, Montreal, QC, Canada. Pierre Ernst, M.D., M.Sc., F.R.C.P.(c), is with the Center for Clinical Epidemiology, Lady Davis Research Institute, Jewish General Hospital, Montreal, QC, Canada. Pierre Ernst, M.D., M.Sc., F.R.C.P.(c), is with the Department of Medicine, Division of Pulmonary Diseases, Jewish General Hospital, Montreal, QC, Canada.

Several previous studies used the combined information from the medical service and prescription claims databases in identification of asthma (Twiggs et al. 2002; Vollmer et al. 2004); however, to the best of our knowledge, the diagnostic properties and the incremental value of using the medical service as well as prescription claims database have not been evaluated.

STUDY OBJECTIVE

Using physician diagnosis from electronic prescriptions and drug management system for primary care physicians, we developed an algorithm to identify patients with asthma based on potential asthma-specific indicators available from administrative medical services and prescription claims databases.

METHODOLOGY

Study Design and Population

From January 2003 to October 2005, we identified a fixed cohort of patients, 18–65 years old, for whom the diagnosis of asthma was confirmed by their primary care physicians. Physicians and patients were recruited from an existing electronic health record and prescription project (Tamblyn et al. 2006). Administrative database indicators for the study patients were created using data from the Quebec provincial health insurance agency (RAMQ) for the 12-month period from November 2004 to October 2005. We restricted the patient population to those continuously covered by the RAMQ drug plan during the study period in order to have complete information on prescribed medication.

Physician-Confirmed Asthma Status

The Medical Office of the Twenty First Century (MOXXI) system is an electronic prescription and drug management system for primary care physicians, community-based pharmacists, and their 15,398 consenting patients in Montreal, Quebec. The MOXXI system allows physicians to write prescriptions electronically and retrieve information on dispensed prescriptions, diagnostic codes, and dates of all medical visits recorded for a patient from the health insurance program and community pharmacy network.

Two sources of information from the MOXXI system were used to confirm patients' asthma status. The first is the automated problem list. The automated problem list generates potential patient-specific medical conditions based upon two sources of information. Daily downloads of all newly reported ICD-9 diagnosis codes are retrieved from the medical service database and transformed into text-based health problems. The provincial drug and health insurance database was used to identify health problems associated with single-indication drugs that are dispensed to the patient. For each patient, the generated potential list of patient-specific medical conditions is presented to the study physicians at the time they open the patient file in the MOXXI system. Study physicians were asked to verify the status of the generated potential medical conditions as "confirmed," "rejected," or have it remained as a potential problem.

The second source of information was the electronic prescription. To complete each electronic prescription, the physician must select a therapeutic indication. The therapeutic indication of asthma is included in drug-specific dropdown menus for all inhaled corticosteroids, fast-acting β -agonists, long-acting β -agonists, leukotrienes, and oral corticosteroids.

Two groups of patients were included as patients having asthma (Figure 1). First, we included 144 patients with asthma as the generated medical conditions in the problem list, and the status was confirmed as having asthma by a study physician. The second group of potential asthma patients are those with the asthma status in the automated problem list as "yet to be confirmed." In this case, we searched at least one record of written electronic prescription from January 2003 to October 2005 to ensure that study physicians had had an opportunity to acknowledge the condition: "asthma" in the generated medical condition(s) of a given patient. Among 175 patients with the asthma status as "yet to be confirmed," we excluded 117 patients (66.9 percent) if electronic prescriptions were not issued by study physicians. This is because the current unconfirmed status could be due to a lack of opportunity for the confirmation by study physicians.

Patients were considered "not having asthma" according to the following two criteria during the same study period. First, we identified 24 patients who were identified as potentially having asthma medical service diagnostic codes in the automated problem list, but the diagnosis was rejected (as false) by the study physician. Among these 24 patients, we excluded one patient who received an electronic prescription, where the study physician entered "asthma" as the therapeutic indication because it was not possible to determine the asthma status due to the conflicting information. Second, we

Figure 1: Asthma Diagnostic Classification Using the Two Gold Standards

Asthma as a Potential Health Condition in the Automated Problem List Confirmation Status by Study Physicians

		Confirmed as Yes (n=144)	Not Confirmed (n=175)	Confirmed as No (n=24)	No Asthma Record Generated (n=2,015)
Written Electronic Prescription	Therapeutic Indication for Asthma (N=50)	48	1	1	0
	Other Therapeutic Indication (N=1,369)	81	57	22	1,209
	No Records (N=929)	15	117	1	806
		202 Patients as Asthma Present		1,232 Patients as Asthma Absent	

identified an additional 2,015 patients without any records of asthma as a generated medication condition in the automated problem list. Among those 2,015 patients, we included only 1,209 patients (60 percent) for whom physicians had at least one opportunity to write electronic prescription.

Finally, we developed an indicator to avoid a misclassification between asthma patients and COPD patients. A total of 37 patients with COPD diagnosis were identified using information in the electronic health record problem list and the therapeutic indication of electronic prescriptions during the same study period. Finally, we excluded patients younger than 19 years old.

Cross-Validation Analysis Sample

In total, 1,431 subjects were included in analysis. Cross-validation analysis was conducted using the two groups of subjects based upon the split samples. Among 1,431 subjects, 50 percent of the subjects were selected at random after being stratified by asthma status (asthma/non-asthma). As a result, 714 subjects were used to construct a predictive model (49.9 percent), whereas 717 subjects were used for the model validation (51.1 percent).

DEVELOPMENT OF ADMINISTRATIVE DATABASE INDICATORS

Data Sources

The provincial health insurance agency (RAMQ) provides first dollar coverage for all medical and hospital care for all Quebec residents. The medical services claims database provides information on the beneficiary, date, type, provider, diagnosis (ICD-9 classification), and location of service delivery (e.g., inpatient, emergency, clinic) for all medical services remunerated on a fee-for-service basis (approximately 86 percent of all services) (Regie de l'assurance-maladie du Quebec 2000). The health beneficiary demographic database provides information on age, sex, and postal code for each patient. The prescription claims database provides information on each drug dispensed, including the drug name, quantity, date, and duration for each prescription; the prescribing physician; and the dispensing pharmacy. Beneficiary and physician identification were encrypted.

Indicators to Identify Patients with Asthma

We assessed five groups of asthma indicators from the RAMQ database based on the data from November 2004 to October 2005.

Prescription Claims Database Indicators. *Number of controller medications used* was defined as the number of the prescriptions dispensed for inhaled corticosteroids, long-acting β -agonists, and leukotriene receptor antagonists. *Number of rescue medications use* was defined as the number of prescriptions of short-acting β -agonists. The information of dispensation date and supply days in the prescription claims file was used to determine the start date and the expected end date of each prescription. Prescriptions were considered "active" if the time period between the starting date and the expected end date of a given prescription overlapped with the study period.

Medical Service Claims Database Indicators

Asthma-Related Visits. Three measures from the medical services claims databases were developed: number of asthma-related visits to primary care physicians; number of asthma-related visits to respiratory-related specialists (respirologists, allergists, medical internists, or pediatricians) (Appendix SA2). An assessment of asthma-related visits was measured by physician specialty due to a difference in the process and outcome of care between the

two specialties (Vollmer et al. 1997; Finkelstein et al. 2000; Diette et al. 2001; Vollmer and Swain 2002).

COPD-Related Visits. COPD-related visit was examined based upon the following measure from the medical services claims database: a proportion of ICD-9 asthma diagnosis code/total asthma-related diagnosis codes being ≤ 75 percent or a presence of ICD-9 COPD diagnosis codes (Appendix SA2).

Comorbidity Conditions. Three categories of medical conditions, “upper airways conditions,” “somatic complaints/neurotic disorder,” and “cardiac-related conditions” were created to investigate *whether inclusion or exclusion of these condition would* increase the likelihood of identifying the presence of asthma based upon the following three underlying mechanisms: pathophysiological model (Grupp-Phelan, Lozano, and Fishman 2001), maladaptive illness behavior model (Rief et al. 2005), and confusing origins of chest pain (Edmondstone 2000), respectively.

Health Care System Utilization Indicators. Two measures, (1) total number of health care visits and (2) number of physicians seen, were used to examine if the likelihood of identifying patients with asthma would increase as the number of encounters with health care professionals increased (Starfield et al. 1991). *Total number of health care visits* was defined as the number of days on which a patient received medical service in any health care setting. *Number of physicians seen* was defined as the number of different physicians who provided medical services for a given patient during the same time period. The provider’s identification number in the medical services claims database was used to produce a count for each patient.

Demographics. A number of studies reported that asthma prevalence and severity as well as asthma disease management vary by age and gender (Osborne et al. 1998; Debley, Redding, and Critchlow 2004; Baibergenova et al. 2005). The information on patient’s age and gender was identified from the health beneficiary demographic database in the RAMQ for each patient.

Identification of Patients with Severe Asthma Exacerbations

Asthma-related hospitalization and emergency room visits reflect acute utilization of health service due to the presence of severe, uncontrolled, or progressive disease (Rea et al. 1986; Fernandes et al. 2003; Naureckas et al. 2005). For the assessment of asthma quality of care, it is essential to evaluate the predictive ability of an algorithm to identify patients who experience serious

adverse asthma outcomes. *Patients with severe acute exacerbation* were defined as those who received medical services in an emergency room or were hospitalized for an asthma-related condition for at least 1 day during the follow-up period or filled at least one prescription for oral steroids (Cockcroft and Swystun 1996; Firoozi et al. 2007). The service location code and date in the RAMQ medical service claims database as well as the dispensation records of oral corticosteroid in the RAMQ prescription claims databases were used to identify these events. A dichotomous variable (yes/no) was created for each patient who experienced at least one of the three events asthma during the study period.

Statistical Analysis

Descriptive statistics were used to characterize the study population and to evaluate differences in the distribution of each asthma-specific markers between asthma patients and patients without asthma. Multivariable logistic regression was used to estimate the probability of the presence of asthma as a function of relevant indicators developed from administrative databases. We examined the following four algorithms: (1) Asthma-COPD classification algorithm, (2) Medical Service Claims Algorithm, (3) Prescription Claims Algorithms, and (4) Medical Service/Prescription combined algorithm.

Based on the results of the regression model, we constructed a receiver operating characteristic (ROC) curve by plotting sensitivity against the false-positive rate ($1 - \text{specificity}$) over the range of cut-off values for the estimated probability of the presence of asthma. The optimal cut-off point for the probability was selected by evaluating the upper left-hand corner of the graph, to correspond to a combination of maximum gain of sensitivity with a minimum reduction in specificity. Sensitivity, specificity, and positive predictive values were evaluated on the basis of the identified optimal cut-off for the probability of “having asthma.”

Several criteria were used to assess the optimal combination of indicators and the incremental value of each regression model. The *C*-statistic, representing the area under the ROC curve, was used to evaluate the ability of the identified algorithm to correctly classify subjects according to the asthma status (Hanley and Mcneil 1982). An area of one represents an algorithm with perfect sensitivity and specificity, while an area of 0.5 represents a “random” classification without any explanatory capacities.

In order to examine an incremental value of the different combination of the hypothesized indicators, an ROC curve was compared with the one from

the preceding algorithm using a nonparametric test for pair-wise comparisons while accounting for correlation between the curves developed based upon the same subjects (DeLong, DeLong, and Clarke-Pearson 1988). The p -value $> .05$ indicates that the ROC curve is statistically different from the ROC curve of the preceding algorithm (e.g., Model 5 versus Model 4). Finally, Akaike Information Criteria were used to assess the goodness-of-fit to the data while taking into account the number of independent variables in the models to avoid potential overfitting (Akaike 1974).

The probability of the presence of asthma was calculated based on the estimated multiple logistic regression models. The optimal cut-off point for the probability of the presence of asthma was used to evaluate the number of patients with severe asthma, who would have been identified using a given algorithm. Based on this number, we evaluated the performance of diagnostic algorithm in identifying patients with acute severe exacerbation of asthma or patients with COPD. Finally, in order to validate the predictive model, the cut-off probability of having asthma based upon the parameter estimates from the predictive model was applied to evaluate the diagnostic performance in the validation model.

RESULTS

Study Population

In total, 1,431 patients were included in our study. According to the two gold standards, automated problem list and electronic prescription, we identified 193 patients (13.5 percent) as “having asthma” and 1,238 patients (86.5 percent) as “not having asthma.”

Table 1 shows the patient characteristics and hypothesized algorithm indicators of the two groups: study patients used to construct predictive model ($n = 714$) and those in the cross-validation model ($n = 717$). Overall, patients with asthma were more likely to be younger and female. As expected, patients with asthma showed a greater tendency of having asthma-related medical visits and asthma drug use during the study follow-up period. Asthma patients in the predictive model sample were more likely to have these characteristics, compared with the asthma patients in the validation sample.

Among the three groups of comorbidity conditions considered, the prevalence of neurotic disorder/somatic complaints and upper airway conditions were slightly higher in patients with asthma. On the other hand, patients

Table 1: Characteristics of Study Patients for the Predictive Model and the Validation Model

	Predictive Model Sample (n = 714)		Validation Model Sample (n = 717)	
	Patients with Asthma (n = 97)	Patients without Asthma (n = 617)	Patients with Asthma (n = 96)	Patients without Asthma (n = 621)
Patient demographics				
Age; mean [SD; range]	49.1 [12.4; 21–65]	51.6 [11.6; 19–65]	49.5 [11.9; 25–65]	51.1 [11.9; 20–65]
Gender, female	73 (75.3%)	393 (63.7%)	64 (66.7%)	400 (64.4%)
Medical service file indicators				
Number of visits to general practitioners, mean n (%)				
0	59 (60.8%)	537 (87.0%)	62 (64.6%)	554 (89.2%)
≥ 1	38 (39.2%)	80 (13.0%)	34 (35.4%)	67 (10.8%)
Number of respiratory-related specialist visits, mean n (%)				
0	93 (95.9%)	605 (98.1%)	90 (93.8%)	607 (97.8%)
≥ 1	4 (4.1%)	12 (1.9%)	6 (6.3%)	14 (2.2%)
Prescription claims file indicators				
Number of controller medication dispensed, mean n (%)				
0	45 (46.4%)	600 (97.2%)	51 (53.1%)	600 (96.6%)
≥ 1	52 (53.6%)	17 (2.8%)	45 (46.9%)	21 (3.4%)
Number of rescue medication dispensed, mean n (%)				
0	50 (51.6%)	596 (96.6%)	66 (68.9%)	601 (96.8%)
1	19 (19.6%)	12 (1.9%)	23 (7.3%)	7 (1.1%)
≥ 2	28 (28.9%)	9 (1.5%)	23 (24.0%)	13 (2.1%)

Comorbidity indicators					
Cardiac-related conditions					
Neurotic disorder/somatic complaints	8 (8.3%)	57 (9.2%)	7 (7.3%)	45 (7.3%)	
Upper airway conditions	13 (13.4%)	58 (9.4%)	14 (14.6%)	70 (11.3%)	
Health service utilization indicators					
Number of unique MD seen, mean	15 (15.5%)	45 (7.3%)	12 (12.5%)	36 (5.8%)	
[SD; range]	6.3 [5.2; 1-37]	4.8 [5.2; 1-69]	6.5 [5.4; 1-35]	4.7 [1-61]	
Number of health care visits, mean	11.3 [10.7; 1-81]	8.8 [9.8; 1-90]	11.7 [11.0; 1-70]	8.8 [11.0; 1-155]	
[SD; range]					
Number of patients with acute asthma	8 (8.3%)	30 (4.9%)	10 (10.4%)	22 (3.5%)	
Number of patients with severe asthma	11 (11.3%)	22 (3.6%)	10 (10.4%)	28 (4.5%)	
Number of patients with COPD	17 (2.3%)			20 (2.7%)	

with asthma were more likely to see a greater number of unique physicians and had more health care visits during the follow-up period (Table 1).

Multivariable Logistic Regression Analysis

Table 2 presents the incremental value of the regression coefficients of the five groups of hypothesized indicators in the logistic models in relationship with the presence of asthma. The use of controller medication was the strongest predictor of identifying patients with asthma (OR = 22.0 for ≥ 1 dispensed controller medication), followed by the use of rescue medication (OR = 11.6 for one dispensed rescue medication and OR = 3.68 for ≥ 2 dispensed rescued medication).

Diagnostic Performance

Table 3 presents the optimal probability cut-offs and the resulting sensitivity, specificity, and positive predictive values across four logistic regression models based upon predictive model sample.

The prescription claims-based indicators (Model 4) and the full model (Model 5) demonstrated nearly equal level of sensitivity, but Model 5 provided a slightly higher level of specificity (0.94 for Model 5; 0.874 for Model 4) and a higher but moderate PPV value (0.648), compared with the Model 4 (0.486). A comparison of the ROC curves obtained from the selected logistic models (Figure 2) showed that the incremental value of adding medical services claims-based indicators to prescription claims algorithm (Model 4) was very small, but the difference was statistically significant ($p = .03$).

Identification of COPD Patients and Patients with Severe Acute Exacerbations

The performances of hypothesized indicators to avoid a misclassification between asthma patients and COPD patients are presented in Table 3. Model 1 demonstrated that the particular indicator alone only identified one COPD patient as an asthma patient. However, the addition of other indicators in the model increased the number of COPD patients who were falsely identified as having asthma up to 12 out of 17 patients (Model 5).

Both Models 4 and 5 demonstrated a similar performance in correctly identifying patients with asthma. Using the probability cut-off of 0.117 (Model 4) as well as 0.153 (Model 5), both models correctly identified all the asthma patients with severe asthma exacerbations; however, Model 5 demonstrated a slight better performance, compared with Model 4, in reducing a

Table 2: Multivariable Logistic Regression Models in Identifying Patients with Asthma Using Algorithm Indicators from Health Care Administrative Databases Based upon Predictive Model Sample ($n = 714$)

Variables	OR (95% CI)		
	Simple Linear Regression	Medical Services Claims-Specific Model	Prescription Claims-Specific Model
Asthma-COPD classification indicator			
Percentage of ICD-9 asthma/total respiratory diagnosis (dx)	0.05 (0.02–0.11)	0.08 (0.03–0.23)	0.11 (0.026–0.44)
Codes < 75% or presence of ICD-9 COPD dx Code			
Medical service visits indicators			
Number of GP visits			
0	Reference	Reference	Reference
≥ 1	4.32 (2.70–6.92)	2.35 (1.18–4.67)	0.90 (0.35–2.30)
Number of respiratory-specialists visits			
0	Reference	Reference	Reference
≥ 1	2.17 (0.69–6.87)	0.72 (0.18–2.92)	0.52 (0.08–3.45)
Comorbidity indicators			
Absence of cardiac conditions	1.13 (0.52–2.45)	1.79 (0.69–4.67)	1.16 (0.49–5.44)
Presence of neurotic disorder/somatic complaints	1.49 (0.78–2.84)	1.00 (0.47–2.14)	1.15 (0.44–3.02)
Presence of upper airway conditions	2.33 (1.24–4.36)	2.29 (1.16–4.47)	1.17 (0.46–3.01)
Health services usage indicators			
Total number of health care visits	1.02 (1.00–1.04)	0.99 (0.96–1.05)	0.99 (0.94–1.05)
Number of unique physician seen	1.04 (1.01–1.08)	1.04 (0.96–1.13)	1.05 (0.94–1.16)
Prescription indicators			
Number of control medication used			
0	Reference	Reference	Reference
≥ 1	40.8 (21.8–76.2)	22.5 (10.3–49.4)	21.1 (8.73–51.0)

continued

Table 2. Continued

<i>Variables</i>	<i>OR (95% CI)</i>			
	<i>Simple Linear Regression</i>	<i>Medical Services Claims-Specific Model</i>	<i>Prescription Claims-Specific Model</i>	<i>Full Model</i>
Number of rescue medication used				
0	Reference		Reference	Reference
1	18.9 (8.67–41.1)		11.1 (4.3–28.3)	11.6 (4.21–32.1)
≥ 2	37.9 (16.6–82.9)		4.99 (1.72–14.5)	3.68 (1.18–11.5)
Demographics				
Sex: female	1.73 (1.06–2.83)	1.89 (1.09–3.26)	0.80 (0.43–1.48)	0.87 (0.46–1.65)
Age: >45 years	0.55 (0.35–0.86)	0.72 (0.44–1.18)	0.38 (0.21–0.68)	0.44 (0.23–0.81)
AIC	—	519.89	377.0	378.2

AIC, Akaike Information Criteria; GP, general practitioner; ICD-9, International Classification of Diseases 9th version.

Table 3: Diagnostic Performance of Five Predictive Models in Identifying Patients with Asthma Based upon Predictive Model ($n = 714$)

Diagnostic Performance of Predictive Model						
Diagnostic Performance Indicators	Model 1: Asthma-COPD Classification	Model 2: Medical Service Claims Indicators		Model 3: Model 1+2	Model 4: Prescription Claims Model	Model 5: Full Model
C-statistics	0.614 (0.571–0.657)	0.732 (0.677–0.786)	0.743 (0.687–0.799)	0.841 (0.788–0.895)	0.872 (0.825–0.918)	
p-value	.157	<.001	.306	.006	.030	
Optimal probability cut-off	0.7097	0.1490	0.1325	0.1167	0.1530	
Sensitivity	0.227 (0.148–0.323)	0.536 (0.423–0.625)	0.577 (0.442–0.644)	0.733 (0.621–0.808)	0.701 (0.699–0.866)	
Specificity	0.985 (0.976–0.995)	0.825 (0.791–0.852)	0.818 (0.805–0.864)	0.874 (0.854–0.906)	0.940 (0.918–0.957)	
Number of correctly identified asthma patients (<i>n</i> = 101)	22	52	56	71	68	
Number of identified patients as false positive (<i>n</i> = 634)	9	108	112	75	37	
Positive predictive value	0.709	0.325	0.333	0.486	0.648	
Identification of Patients with Severe Acute Exacerbations						
Emergency room visits/hospitalization (<i>n</i> = 38)	Asthma (<i>n</i> = 8)	No asthma (<i>n</i> = 30)	Asthma (<i>n</i> = 8)	No asthma (<i>n</i> = 30)	Asthma (<i>n</i> = 8)	No asthma (<i>n</i> = 30)
	1	0	7	17	8	6
Oral corticosteroid use (<i>n</i> = 34)	Asthma (<i>n</i> = 11)	No asthma (<i>n</i> = 23)	Asthma (<i>n</i> = 11)	No asthma (<i>n</i> = 23)	Asthma (<i>n</i> = 11)	No asthma (<i>n</i> = 23)
	4	1	7	11	9	7
COPD patients falsely identified as having asthma (<i>n</i> = 17)	1	9	8	11	11	12

Figure 2: Receiver Operator Characteristics (ROC) Curves for Various Combination of Hypothesized Algorithm Indicators by Fitting Multivariable Logistic Regression Models Based upon the Predictive Model Sample

number of falsely identified patients with severe asthma exacerbations (oral corticosteroid use) (7 out of 23 patients; PPV = 0.611) (Table 3).

Cross-Validation Analysis

Using the cut-off probability from the predictive model, the performance of sensitivity, specificity, and positive predictive value, based upon the validation sample, are presented in Table 4. Across any models, the specificity in the validation model was similar to those derived from the predictive model; whereas, the sensitivity of validation model in Model 4 and Model 5 was lower than those from the predictive model. Figure 3 shows the selected results of ROC curves, compared between the predictive model and cross-validation model. A greatest difference in the validated ROC curve and the predicted ROC curve was seen in the prescription claims-specific model (Figure 3).

DISCUSSION

The availability of administrative health care databases creates an extremely valuable opportunity for researchers to capture and assess a wide range of

Table 4: Diagnostic Performance of Cross-Validation Model Using the Cut-Off Probability from the Predictive Models ($n = 717$)

Diagnostic Performance Indicators	Diagnostic Performance of Cross-Validation Model							
	Model 1: Asthma-COPD Classification		Model 2: Medical Services Claims Indicators		Model 3: Model 1+2		Model 4: Prescription Claims Indicators	
	Asthma	No asthma	Asthma	No asthma	Asthma	No asthma	Asthma	No asthma
C-statistics	0.606 (0.564–0.648)		0.704 (0.642–0.765)		0.724 (0.778–0.786)		0.782 (0.719–0.833)	0.828 (0.778–0.879)
p-value	< .001		.002		.003		.159	.006
Sensitivity	0.240 (0.158–0.338)		0.510 (0.406–0.613)		0.500 (0.386–0.594)		0.583 (0.479–0.683)	0.563 (0.458–0.664)
Specificity	0.989 (0.977–0.996)		0.831 (0.799–0.860)		0.836 (0.805–0.864)		0.855 (0.825–0.882)	0.952 (0.932–0.967)
Number of correctly identified asthma patients ($n = 96$)	23		49		47		56	54
Number of identified patients as false positive ($n = 621$)	7		105		91		90	30
Positive predictive value	0.767		0.318		0.348		0.383	0.643
Identification of Patients with Severe Acute Exacerbations								
Emergency room visits/hospitalization ($n = 32$)	Asthma ($n = 10$) 2	No asthma ($n = 22$) 0	Asthma ($n = 10$) 8	No asthma ($n = 22$) 15	Asthma ($n = 10$) 7	No asthma ($n = 22$) 15	Asthma ($n = 10$) 7	No asthma ($n = 22$) 3
Oral corticosteroids use ($n = 38$)	Asthma ($n = 10$) 4	No asthma ($n = 28$) 1	Asthma ($n = 10$) 8	No asthma ($n = 28$) 15	Asthma ($n = 10$) 7	No asthma ($n = 28$) 9	Asthma ($n = 10$) 8	No asthma ($n = 28$) 11
COPD patients falsely identified as having asthma ($n = 20$)		4		15		13		18
								17

Figure 3: Receiver Operator Characteristics (ROC) Curves for a Predictive and a Cross-Validation Model from Various Combinations of Indicators

clinical information at the population level in a cost-effective manner. As asthma continues to be a serious public health problem, administrative health care databases have a great advantage; they provide a method to monitor quality of care and monitor asthma patients' control status throughout the health care continuum. In achieving this objective, this study attempted to develop a predictive algorithm to accurately identify patients' asthma status to enable optimal use of administrative databases for this purpose.

We found that the combined use of asthma-specific indicators from medical services claims and indicators from prescription claims databases showed a best fitting algorithm with a sensitivity of 70.1 percent and specificity of 94.0 percent and positive predictive value of 0.65. Several other U.S. studies demonstrated similar findings as our study. Vollmer et al. (2004) demonstrated a benefit of using multiple claims databases for the case identification of asthma. Using the electronic database of a large Health Maintenance Organization, the combined use of both medical service claims and pharmacy databases identified 12,492 patients with asthma; however, the pharmacy database identified 47 percent of the asthmatics, whereas the medical service claims databases identified 81 percent (Vollmer et al. 2004). However, they failed to examine the level of specificity using patients without asthma.

The combined use of medical and pharmacy claims files have several advantages in identifying a wide range of patients with asthma who represent very different aspects of the spectrum of asthma management. Previous studies reported that there is a difference between patients with an asthma diagnosis alone and patients who receive prescription medication for asthma. For instance, suboptimal use of asthma medication is one of the major barriers in achieving optimal asthma management, and both physician and patients may contribute to under-use of control medication (Haltermann et al. 2002; Fitz-Gerald et al. 2006). Even when asthma medication is prescribed, nearly one-third of the prescriptions are not filled by patients within a 1-year period (Watts et al. 1997). Therefore, the magnitude of identifying uncontrolled asthma will likely be underestimated, if the pharmacy claims data were used alone.

Some patients receive a prescription for asthma medications but do not have a record of either asthma or asthma-like diagnosis in the ambulatory care settings (Pont et al. 2002). This particular mismatch is complex, and it could be related to a distinctive pattern of health service use, such as frequent visitors to the emergency room for asthma-related conditions (Schatz et al. 2004). Alternatively, this discrepancy could be related to the difference in the requirement of reporting between medical services and prescription databases.

Our finding of the sensitivity and specificity from the medical service database is lower than another Quebec study, in which the information from medical charts was used as a gold standard. The Quebec study found sensitivity of up to 94 percent and specificity of up to 95 percent. However, approximately half of the study population was from respiratory specialty clinics where asthma is more likely to be recorded as the reason for the visit (Blais et al. 2006).

One of strengths in our study is our demonstration of a potential practical utility of an indicator in distinguishing between asthma patients and COPD patients using a medical service claims-specific indicator. As COPD and asthma are treated with same medication, use of pharmaceutical claims database could easily misclassify patients with asthma and COPD (Canadian Thoracic Society 2003). We identified a relevant and sensitive indicator based upon ICD-9 diagnosis codes for COPD as well as respiratory-related conditions (McKnight et al. 2005) resulting in a minimum number of misclassified COPD patients as asthma patients (1/17 COPD patient). However, as expected, the inclusion of other hypothesized indicators in the algorithm increased misclassification (up to 12 COPD patients).

Our study also demonstrated the interesting performance of the prescription claims algorithm. This particular algorithm demonstrated a nearly

equal level of sensitivity as the medical service/prescription claims combined model. However, the specificity and the positive predictive value of prescription claims model were lower, compared with the algorithm based upon the medical service/prescription claims combined model (sensitivity = 0.73, specificity = 0.87, and PPV = 0.49). The performance and predictive properties of algorithms tested in our study are consistent with prior research. Pont and colleagues in the Netherlands found that prescription of one or more inhaled bronchodilator and one or more inhaled corticosteroid in 12 months had 41 percent sensitivity and 99 percent specificity with PPV = 0.84; whereas any prescriptions of antiasthma medication (FABA, LABA, ICS, theophylline, and cromoglycates) in 12 months resulted in 95 percent sensitivity and 99 percent specificity with PPV = 0.70, when using at least one asthma diagnosis in computerized patient records of general practitioners as the gold standard (Pont et al. 2002).

Furthermore, our study attempted to identify those asthmatic patients who are most likely to experience morbidity and resource utilization. Our prescription claims-based algorithm showed predictive properties for emergency room/hospitalization for asthma with 100 percent sensitivity, 80 percent specificity, and PPV = 0.57. Schatz et al. (2006) developed a similar risk stratification scale from computerized pharmacy data, including any asthma medication use or having previous emergency room visits, and achieved a maximum sensitivity of 83 percent and specificity of 98.3 percent as well as the maximum positive predictive value of 46.9 percent (Schatz et al. 2006).

There are several limitations in interpreting our results. The optimal cut-off point of our best-fitting algorithm failed to identify approximately 30 percent of patients who were identified as "having asthma." We defined "having asthma" for the gold standard based upon an approximately 3 years' worth of information from January 2003 to October 2005. On the other hand, algorithm indicators were developed from administrative database based on the 12-month period from November 2004 to October 2005. According to Ernst et al. (2002), 23 percent of patients with asthma who are initially treated with therapy appropriate for mild asthma are rarely treated with therapy when followed up to 5 years. Therefore, asthma severity as well as an intensity of treatment may have changed over time among patients whose asthma was confirmed earlier in the study period. An extension of the time period to over 1-year period in constructing hypothesized indicators could have resulted in a better performance (Vollmer et al. 2004).

Similarly, the specificity demonstrated in our study may generate a considerable number of falsely identified patients as having asthma. An

application of our current algorithm may falsely identify approximately 20–30 percent of patients as having severe acute exacerbations. Although the number of patients with severe acute exacerbations is very small, an application of our algorithm into the general population should be carefully considered.

Medical services for asthma patients are delivered by both general practitioners and specialists, but only a small proportion of patients with asthma receive care from both (Diette et al. 2001). Patients who are seen by specialists are distinctively different from those who are cared for by general practitioners (Vollmer et al. 1997). In the present study, a gold standard was established based on the confirmation of asthma status by primary care physicians in the Quebec primary care network. Thus, our finding may not be generalizable to patients whose asthma is cared solely by respiratory-related specialists.

In addition, the gold standard is established using the two sources of information: electronic prescription and automated problem list. In particular, the automated problem list generates potential patient-specific medical conditions based upon daily downloads of all newly reported ICD-9 diagnosis codes from the medical service claims database. Although study physicians were asked to verify the status of the generated potential medical conditions, the algorithm using the indicators from medical service claims database may potentially be overestimated.

In conclusion, the current study demonstrated a practical approach of using administrative claims databases to identify patients with asthma. Our findings can be useful to identify patients with asthma from administrative claims databases for the future assessment of asthma management and its predictors.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This study was supported by the the Régie de l'assurance maladie du Québec and the Canadian Institutes of Health Research.

REFERENCES

- Abrahamowicz, M., Y. Xiao, R. Ionescu-Ittu, and D. Lacaille. 2007. "Simulations Showed That Validation of Database-Derived Diagnostic Criteria Based on a Small Subsample Reduced Bias." *Journal of Clinical Epidemiology* 60 (6): 600–9.
- Akaike, H. 1974. "New Look at Statistical-Model Identification." *IEEE Transactions on Automatic Control* AC19 (6): 716–23.

- Baibergenova, A., L. Thabane, N. Akhtar-Danesh, M. Levine, A. Gafni, R. Moineddin, and I. Pulcins. 2005. "Effect of Gender, Age, and Severity of Asthma Attack on Patterns of Emergency Department Visits due to Asthma by Month and Day of the Week." *European Journal of Epidemiology* 20 (11): 947–56.
- Blais, L., C. Lemiere, D. Menzies, and D. Berbiche. 2006. "Validity of Asthma Diagnoses Recorded in the Medical Services Database of Quebec." *Pharmacoepidemiology and Drug Safety* 15 (4): 245–52.
- Boulet, L. P., T. R. Bai, A. Becker, D. Berube, R. Beveridge, D. M. Bowie, K. R. Chapman, J. Cote, D. Cockcroft, F. M. Ducharme, P. Ernst, J. M. Fitzgerald, T. Kovesi, R. V. Hodder, P. O'Byrne, B. Rowe, M. R. Sears, F. E. Simons, and S. Spier. 2001. "What Is New Since the Last (1999) Canadian Asthma Consensus Guidelines? (Comment)." *Canadian Respiratory Journal* 8 (Suppl 27A): 5A–27A.
- Canadian Thoracic Society. 2003. "Executive Summary: Canadian Thoracic Society Recommendations for Management of Chronic Obstructive Pulmonary Disease." *Canadian Respiratory Journal* 10 (Suppl A May/June): 5A–33A.
- Cockcroft, D. W., and V. A. Swystun. 1996. "Asthma Control versus Asthma Severity." *Journal of Allergy and Clinical Immunology* 98 (6, Part 1): 1016–8.
- Debley, J. S., G. J. Redding, and C. W. Critchlow. 2004. "Impact of Adolescence and Gender on Asthma Hospitalization: A Population-Based Birth Cohort Study." *Pediatric Pulmonology* 38 (6): 443–50.
- DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson. 1988. "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach." *Biometrics* 44 (3): 837–45.
- Diette, G. B., E. A. Skinner, T. T. Nguyen, L. Markson, B. D. Clark, and A. W. Wu. 2001. "Comparison of Quality of Care by Specialist and Generalist Physicians as Usual Source of Asthma Care for Children." *Pediatrics* 108 (2): 432–7.
- Edmondstone, W. M. 2000. "Chest Pain and Non-Respiratory Symptoms in Acute Asthma." *Postgraduate Medical Journal* 76 (897): 413–4.
- Ernst, P., B. Cai, L. Blais, and S. Suissa. 2002. "The Early Course of Newly Diagnosed Asthma." *American Journal of Medicine* 112 (1): 44–8.
- Fernandes, A. K., F. Mallmann, A. M. Steinhurst, F. L. Nogueira, E. M. Avila, D. Z. Saucedo, F. J. Machado, M. G. Raymundi, S. S. Barreto, and P. T. Dalcin. 2003. "Characteristics of Acute Asthma Patients Attended Frequently Compared with Those Attended Only Occasionally in an Emergency Department." *Journal of Asthma* 40 (6): 683–90.
- Finkelstein, J. A., P. Lozano, R. Shulruff, T. S. Inui, S. B. Soumerai, M. Ng, and K. B. Weiss. 2000. "Self-Reported Physician Practices for Children with Asthma: Are National Guidelines Followed?" *Pediatrics* 106 (4, Suppl): 886–96.
- Firoozi, F., C. Lemiere, M. F. Beauchesne, A. Forget, and L. Blais. 2007. "Development and Validation of Database Indexes of Asthma Severity and Control." *Thorax* 62 (7): 581–7.
- FitzGerald, J. M., L. P. Boulet, R. A. McIvor, S. Zimmerman, and K. R. Chapman. 2006. "Asthma Control in Canada Remains Suboptimal: The Reality of Asthma Control (TRAC) Study." *Canadian Respiratory Journal* 13 (5): 253–9.

- Grupp-Phelan, J., P. Lozano, and P. Fishman. 2001. "Health Care Utilization and Cost in Children with Asthma and Selected Comorbidities." *Journal of Asthma* 38 (4): 363–73.
- Halterman, J. S., H. L. Yoos, J. M. Kaczorowski, K. McConnochie, R. J. Holzhauer, K. M. Conn, S. Lauver, and P. G. Szilagyi. 2002. "Providers Underestimate Symptom Severity among Urban Children with Asthma." *Archives of Pediatrics and Adolescent Medicine* 156 (2): 141–6.
- Hanley, J. A., and B. J. Mcneil. 1982. "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve." *Radiology* 143 (1): 29–36.
- Kozyrskyj, A. L., C. A. Mustard, and A. B. Becker. 2004. "Identifying Children with Persistent Asthma from Health Care Administrative Records." *Canadian Respiratory Journal* 11 (2): 141–5.
- McKnight, J., A. Scott, D. Menzies, J. Bourbeau, L. Blais, and C. Lemiere. 2005. "A Cohort Study Showed That Health Insurance Databases Were Accurate to Distinguish Chronic Obstructive Pulmonary Disease from Asthma and Classify Disease Severity." *Journal of Clinical Epidemiology* 58 (2): 206–8.
- Naureckas, E. T., V. Dukic, X. Bao, and P. Rathouz. 2005. "Short-Acting Beta-Agonist Prescription Fills as a Marker for Asthma Morbidity." *Chest* 128 (2): 602–8.
- Osborne, M. L., W. M. Vollmer, R. E. Johnson, and A. S. Buist. 1995. "Use of an Automated Prescription Database to Identify Individuals with Asthma." *Journal of Clinical Epidemiology* 48 (11): 1393–7.
- Osborne, M. L., W. M. Vollmer, K. L. Linton, and A. S. Buist. 1998. "Characteristics of Patients with Asthma within a Large HMO: A Comparison by Age and Gender." *American Journal of Respiratory and Critical Care Medicine* 157 (1): 123–8.
- Pont, L. G., G. T. van der Werf, P. Denig, and F. M. Haaijer-Ruskamp. 2002. "Identifying General Practice Patients Diagnosed with Asthma and Their Exacerbation Episodes from Prescribing Data." *European Journal of Clinical Pharmacology* 57 (11): 819–25.
- Rea, H. H., R. Scragg, R. Jackson, R. Beaglehole, J. Fenwick, and D. C. Sutherland. 1986. "A Case–Control Study of Deaths from Asthma." *Thorax* 41 (11): 833–9.
- Regie de l'assurance-maladie du Quebec. 2000. *Statistiques Annuelles*. Quebec: Regie de l'assurance-maladie du Quebec.
- Rief, W., A. Martin, A. Klaiberg, and E. Brahler. 2005. "Specific Effects of Depression, Panic, and Somatic Symptoms on Illness Behavior." *Psychosomatic Medicine* 67 (4): 596–601.
- Schatz, M., R. Nakahiro, C. H. Jones, R. M. Roth, A. Joshua, and D. Petitti. 2004. "Asthma Population Management: Development and Validation of a Practical 3-Level Risk Stratification Scheme." *American Journal of Managed Care* 10 (1): 25–32.
- Schatz, M., R. S. Zeiger, W. M. Vollmer, D. Mosen, A. J. Apter, T. B. Stibolt, A. Leong, M. S. Johnson, G. Mendoza, and E. F. Cook. 2006. "Development and Validation of a Medication Intensity Scale Derived from Computerized Pharmacy Data that Predicts Emergency Hospital Utilization for Persistent Asthma." *American Journal of Managed Care* 12 (8): 478–84.

- Seung, S. J., and N. Mittmann. 2005. "Urgent Care Costs of Uncontrolled Asthma in Canada, 2004." *Canadian Respiratory Journal* 12 (8): 435–6.
- Smith, M. J., K. L. Rascati, and B. C. McWilliams. 2004. "Inhaled Anti-Inflammatory Pharmacotherapy and Subsequent Hospitalizations and Emergency Department Visits among Patients with Asthma in the Texas Medicaid Program." *Annals of Allergy, Asthma and Immunology* 92 (1): 40–6.
- Starfield, B., J. Weiner, L. Mumford, and D. Steinwachs. 1991. "Ambulatory Care Groups: A Categorization of Diagnoses for Research and Management." *Health Services Research* 26 (1): 53–74.
- Tamblyn, R., A. Huang, Y. Kawasumi, G. Bartlett, R. Grad, A. Jacques, M. Dawes, M. Abrahamowicz, R. Perreault, L. Taylor, N. Winslade, L. Poissant, and A. Pinsonneault. 2006. "The Development and Evaluation of an Integrated Electronic Prescribing and Drug Management System for Primary Care." *Journal of the American Medical Informatics Association* 13 (2): 148–59.
- Twigg, J. E., J. Fifield, A. J. Apter, E. A. Jackson, and R. A. Cushman. 2002. "Stratifying Medical and Pharmaceutical Administrative Claims as a Method to Identify Pediatric Asthma Patients in a Medicaid Managed Care Organization." *Journal of Clinical Epidemiology* 55 (9): 938–44.
- Vollmer, W. M., E. A. O'Connor, M. Heumann, E. A. Frazier, V. Breen, J. Villnave, and A. S. Buist. 2004. "Searching Multiple Clinical Information Systems for Longer Time Periods Found More Prevalent Cases of Asthma." *Journal of Clinical Epidemiology* 57 (4): 392–7.
- Vollmer, W. M., M. O'Hollaren, K. M. Ettinger, T. Stibolt, J. Wilkins, A. S. Buist, K. L. Linton, and M. L. Osborne. 1997. "Specialty Differences in the Management of Asthma. A Cross-Sectional Assessment of Allergists' Patients and Generalists' Patients in a Large HMO." *Archives of Internal Medicine* 157 (11): 1201–8.
- Vollmer, W. M., and M. C. Swain. 2002. "Role of the Specialist in the Treatment of Asthma." *Current Opinion in Allergy and Clinical Immunology* 2 (3): 189–94.
- Watts, R. W., G. McLennan, I. Bassham, and O. el Saadi. 1997. "Do Patients with Asthma Fill Their Prescriptions? A Primary Compliance Study." *Australian Family Physician* 26 (Suppl 1): S4–6.
- Wilchesky, M., R. M. Tamblyn, and A. Huang. 2004a. "Validation of Diagnostic Codes within Medical Services Claims." *Journal of Clinical Epidemiology* 57 (2): 131–41.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix SA2: ICD9 Diagnostic Codes and Asthma Medication.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.