

Comparison of the accuracy of classification models to estimate healthcare use and costs associated with COPD exacerbations in Saskatchewan, Canada: A retrospective cohort study

John Paul Kuwornu PhD¹, Gary F. Teare PhD^{1,2}, Jacqueline M. Quail PhD^{1,2}, Evelyn Forget PhD¹, Saman Muthukumarana PhD³, Xiaoyun E. Wang MSc², Meric Osman MA², Lisa M. Lix PhD^{1,2}

JP Kuwornu, GF Teare, JM Quail, E Forget, S Muthukumarana, XE Wang, M Osman, LM Lix. Comparison of the accuracy of classification models to estimate healthcare use and costs associated with COPD exacerbations in Saskatchewan, Canada: A retrospective cohort study. *Can J Respir Ther* 2017;53(3):37–44

Objective: COPD is a high-cost disease and results in frequent contacts with the healthcare system. The study objective was to compare the accuracy of classification models with different covariates for classifying COPD patients into cost groups.

Methods: Linked health administrative databases from Saskatchewan, Canada, were used to identify a cohort of newly diagnosed COPD patients (April 1, 2007 to March 31, 2011) and their episodes of healthcare encounters for disease exacerbations. Total costs of the first and follow-up episodes were computed and patients were categorized as persistently high cost, occasionally high cost, and persistently low cost based on cumulative cost distribution ranking using the 75th percentile cutoff for high-cost status. Classification accuracy was compared for seven multinomial logistic regression models containing socio-demographic characteristics (i.e., base model), and socio-demographic and prior healthcare use characteristics (i.e., comparator models).

Results: Of the 1182 patients identified, 8.5% were classified as persistently high cost, 26.1% as occasionally high cost, and the remainder as persistently low cost. The persistently high-cost and occasionally high-cost patients incurred 10 times (\$12 449 vs \$1263) and seven times (\$9334 vs \$1263) more costs in their first exacerbation episode than persistently low-cost patients, respectively. Classification accuracy was 0.67 for the base model, whereas the comparator model containing socio-demographic and number of prior hospital admissions had the highest accuracy (0.72).

Conclusions: Costs associated with COPD exacerbation episodes are substantial. Adding prior hospitalization to socio-demographic characteristics produced the highest improvements in classification accuracy. Accurate classification models are important for identifying potential healthcare cost management strategies.

Key Words: chronic obstructive; cost analysis; healthcare costs; logistic models; longitudinal studies; pulmonary disease

INTRODUCTION

Patients with chronic obstructive pulmonary disease (COPD) are responsible for up to one-quarter of all hospitalizations and emergency department (ED) visits, and more than one-fifth of ambulatory visits [1]. Previous studies have shown that COPD exacerbations, periods in the disease course that are characterized by worsening patient symptoms, require follow-up care [2], and are therefore major contributors to the total healthcare costs associated with the disease treatment and management. Annual costs are estimated to be 10 times higher among COPD patients who experience exacerbations than among those who do not [3]. The average cost of a severe COPD exacerbation was estimated at \$9557, with the overall economic burden to the Canadian healthcare system projected in the range of \$646–\$736 million per annum [4].

COPD exacerbations often require an ED visit or hospitalization [5, 6]. Patients may also receive follow-up care from their primary care provider or a specialist and might require additional medications [7]. Accordingly, an exacerbation episode may require multiple contacts with different healthcare providers and services. Comprehensive information about costs associated with COPD exacerbations can benefit from an episode-of-care data system, which aggregates healthcare services related to the treatment of the condition [8]. The episode of care provides a clinically meaningful unit for measuring healthcare costs [9] and allows for a detailed analysis of the treatment processes that generate the costs [10].

The phenomenon of a very few individuals, usually the top 5%–15% of healthcare users, accounting for more than 50% of healthcare costs has been consistently reported in the literature [11]. However, recent analyses have also revealed that these high-cost patients are a heterogeneous sub-group, with some patients persistently incurring high costs while others only occasionally incur high-cost services [12]. Understanding this dynamic nature of healthcare expenditures could potentially benefit the development of cost-management strategies. Given that the prevalence of COPD is projected to increase in the future and place even greater economic burden on the healthcare system [4], developing models to predict high-cost groups in early episodes could contribute to the development of timely interventions. The study objectives were to use linked population-based administrative health data to estimate healthcare resource use and costs associated with episodes of COPD exacerbations and to compare the accuracy of classification models with different covariates for classifying patients into cost groups.

METHODS

Data sources

We used administrative health data from the province of Saskatchewan, Canada, which has a population of approximately 1.1 million according to the 2011 Statistics Canada Census. Like all Canadian provinces, Saskatchewan has a universal healthcare program, which means that virtually all residents are eligible for health insurance coverage.

¹Department of Community Health Sciences, University of Manitoba, Winnipeg, MB, Canada; ²Saskatchewan Health Quality Council, 111 Research Drive, Saskatoon, SK, Canada; ³Department of Statistics, Faculty of Science, University of Manitoba, Winnipeg, MB, Canada

Correspondence: John Paul Kuwornu, Department of Community Health Sciences, Faculty of Health Sciences, University of Manitoba, 750 Bannatyne Ave., Winnipeg, MB R3E 0W3, Canada. Tel: 1 519 281 0505. email: kuwornuj@myumanitoba.ca



This open-access article is distributed under the terms of the Creative Commons Attribution Non-Commercial License (CC BY-NC) (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits reuse, distribution and reproduction of the article, provided that the original work is properly cited and the reuse is restricted to noncommercial purposes. For commercial reuse, contact editor@csrt.com

The province maintains multiple administrative health databases in electronic format and they can be anonymously linked via a unique personal health number [13].

Episodes of care for COPD were constructed using databases that capture primary, emergency, and acute care service, for all provincial health insurance beneficiaries, including physician billing claims, ED visit records, hospital discharge abstracts, and prescription drug dispensation records. A hospital discharge abstract is completed when a patient is discharged from an acute care facility. Up to 25 diagnoses are recorded using the International Classification of Diseases, 10th Revision, Canada (ICD-10-CA) codes on each admission record. Information on emergency care is collected in the ED database, which captures up to 16 diagnoses on each record using ICD-10-CA. Physician billing claims contain information submitted by physicians providing care to patients in outpatient settings. A single diagnosis is recorded on each claim using three-digit ICD-9 codes. Prescription drug dispensation records contain information on drugs dispensed in outpatient settings, including the date of dispensation and national drug identification numbers. The population registry and vital statistics registry were also used in the study. They contain demographic information, as well as dates of health insurance coverage and death.

Data were accessed and analyzed at the provincial Health Quality Council in accordance with a standing data-sharing agreement between the organization and the provincial Ministry of Health. Ethics approval for the research was received from the University of Saskatchewan Biomedical Research Ethics Board.

Study design and cohort selection

The study adopted a retrospective cohort design. The cohort was composed of adults (35+ years old) who were newly diagnosed with COPD between April 1, 2007 and March 31, 2011 and were residents of Saskatoon health region (SHR) and Regina Qu'Appelle health region (RQHR), two of 12 health regions in Saskatchewan at the time of the study and the only ones for which ED data were available. Both SHR and RQHR are the only health regions that contain major urban centers (population > 200 000 in each center) and together account for just over half of the provincial population.

We used the following validated case definition to identify individuals with COPD: (i) one or more hospitalizations with a diagnosis of COPD in any diagnosis field or (ii) one or more physician visits with a diagnosis of COPD [14]. This case definition had a sensitivity of 85.0% and a specificity of 78.4% when compared with clinical evaluations by a physician [1]. The index date for COPD diagnosis was the date of the earliest hospitalization admission or physician visit for COPD. Cases were identified from hospital discharge abstracts using the following ICD-10-CA codes: J41, J42, J43, or J44; cases in physician billing claims were identified with ICD-9 codes 491, 492, or 496.

To increase the likelihood that cohort members were newly diagnosed COPD cases, we used a look-back period of 5 years from the index date to determine whether a patient had a prior COPD diagnosis. We selected this duration of time based on previous research [15], which showed that most adults with clinically significant COPD will contact the healthcare system at least once in this period. The cohort was limited to individuals who had continuous provincial health insurance coverage from five years prior to their index date until death or March 31, 2012, whichever came first. This restriction allowed us to identify incidence COPD cases and also capture all insured healthcare contacts during the episode. We restricted the cohort to an incident cohort to study changes in healthcare utilization and costs as the condition progresses. Finally, the study considered only the index (i.e., first) and follow-up episodes among patients who experienced at least two episodes following their COPD diagnosis date.

Defining episodes of care for COPD exacerbations

All episodes of care for COPD exacerbations following the index diagnosis were defined using the healthcare services that initiated, continued, and ended them. We identified episodes of care based on a method developed by the Canadian Institute for Health Information (CIHI), a

national nonprofit organization that provides standardized methods and data sources for health services research, for ascertaining exacerbations [16]. Hospital- or ED-initiated episodes had: (i) a COPD diagnosis in the most responsible diagnosis field or (ii) a diagnosis of an acute lower respiratory tract infection in the most responsible diagnosis field and a diagnosis of other COPD (ICD-10-CA code J44) in the second diagnosis field. Physician visit initiated episodes were identified by an ICD-9 code for COPD or respiratory infection and had to be accompanied by the dispensation of a drug used to treat acute exacerbations of COPD, including antibiotics, systemic corticosteroids, short-acting beta agonists (SABAs), and SABAs combined with anticholinergics within two days of a physician visit.

An episode continued if there were respiratory-related hospitalizations or ED, family practitioners (FP), or specialist visits that followed the initiating service within a 30-day period. All respiratory-related outpatient prescription drugs dispensed during this period were also included in the episode.

An episode ended after either the occurrence of a 30-day clean period, in which there were no respiratory-related healthcare contacts, or death. All patients were followed for at least one year from their index date until March 31, 2012, or death, whichever occurred first. All ongoing episodes at the end of the observation period were excluded to ensure we had complete information to estimate cost of all exacerbation episodes included in the study.

Episode of care costs

The total cost of an episode of care was the sum of all costs associated with healthcare utilization related to respiratory diagnoses incurred between the episode start and end dates. Inpatient hospital costs were estimated based on a standard methodology developed by CIHI [17]. Briefly, inpatient hospital costs were computed by multiplying the resource intensity weight (RIW) of a hospital stay with the cost per weighted case (CPWC). An RIW is a relative value that describes the expected resource consumption of a patient based on: (i) their case mix group; (ii) factors known to affect resource utilization and length of stay including age, comorbidity, hospital-based interventions; and (iii) atypical length of stay such as patients who are transferred between facilities and palliative cases. The CPWC represents the cost of an average patient's hospital stay. We used CPWC figures estimated for Saskatchewan. For the ED cost component, total annual expenditures were obtained from the Ministry of Health and total annual number of visits was extracted from the ED database; these were used to estimate an average cost per visit. The cost of a physician visit was the amount billed by the physician to the provincial Ministry of Health, as recorded in the physician billing claims. Prescription drug costs were based on prices of the active substance plus a dispensing fee, as recorded in the dispensation records.

Episode costs were adjusted for inflation using the health and personal care component of the Saskatchewan consumer price indices [18] and expressed in 2011–2012 constant dollars. All costs were estimated from the perspective of the public payer; individual out-of-pocket expenditures such as copayments were not included in this study.

Study measures

Outcomes: Using the ranked distribution of cumulative total costs in the index and follow-up episodes, we identified high-cost status using the 75th percentile cutoff. Patients were categorized into three cost groups: persistently high cost (i.e., those whose costs were at the 75th percentile and above in the first and subsequent episode), occasionally high cost (i.e., those whose costs were at the 75th percentile and above in either of the episodes), and persistently low cost (i.e., those whose costs were below the 75th percentile in both episodes). The choice of a cutoff point is largely empirically driven [19]; previous studies have used different cutoffs to define high-cost patients including the top 5% [20, 21], the top 10% [22, 23], the top 20% [24], the top 25% [25], or the top tertile [26]. For our data, using more stringent cutoffs such as top 10% would have resulted in sample sizes that were too small to

provide stable estimates in regression models [22]. We also estimated the time between the index and follow-up episode for each cost group.

Health services utilization measures: For each patient, we tracked the number and duration of use of various healthcare services in each episode (see Table A1 in the appendix for the definitions of these utilization variables). These included visits to EDs, FPs, and specialists as well as hospital admissions to general wards and specialized care units (SCUs). The number of dispensed drugs was calculated using the American hospital formulary service pharmacologic-therapeutic classification system by summing the number of different four-digit drug classifications for each cohort member.

Patient and disease characteristics: The patient and disease characteristics included in the analysis were guided by the Andersen healthcare utilization model [27]. Andersen proposed that an individual's healthcare use is influenced by three broad groups of factors, namely predisposing, enabling, and need. The predisposing factors were sex (i.e., male or female) and age group (i.e., 35–54, 55–74, or 75+). The enabling factor was residence location (i.e., urban or rural); urban residents were those whose postal codes were in a census metropolitan or agglomeration area (i.e., 10 000+ population). Finally, the need factor examined in this study was the level of comorbidity, which was defined using the Charlson comorbidity index [28]. This index was based on diagnoses in the hospital discharge abstract and the physician billing claims data. The index score for each individual in the study cohort was categorized as 0, 1, 2, or ≥ 3. The Charlson comorbidity index has previously been used to predict healthcare utilization in Saskatchewan [29]. We also included the fiscal year of COPD diagnosis (i.e., 2007–2008, 2008–2009, 2009–2010, or 2010–2011) in the model, as this may influence follow-up care patterns. All variables were defined as of the index date of COPD diagnosis except for the Charlson comorbidity index score, which was calculated using data for the 365-day period prior to the index date.

Statistical analysis

We described overall and individual cost components of episodes of COPD exacerbations with means and standard deviations (SDs). The χ^2 statistic was used to test for differences in patients' healthcare encounters in the three cost groups. All hypotheses tests were conducted using two-tailed test at the significance level of 0.05. We plotted the duration (in days) of healthcare utilization measures during the episodes of care.

A multinomial logistic regression model was fit to the data to predict cost group membership using information on patients' age, sex, residence location, comorbidities, and fiscal year of COPD diagnosis (i.e., base model). A previous study [21] has shown that including the number of previous healthcare services would enhance a model's ability to predict future high-cost patients. To evaluate the improvement in classification accuracy, we included the number of times different healthcare services were utilized in the index episode. To the base model, these subsequent models added: number of hospital admissions (model 1), number of ED visits (model 2), number of FP visits (model 3), number of specialist visits (model 4), number of types of drugs dispensed (model 5), and all five healthcare utilization measures (model 6). We added each of the five healthcare services to the base model one at a time to construct models 1 to 5, whilst model 6 comprised of the base model and all five healthcare services.

To evaluate model performance, we used measures of goodness-of-fit (i.e., the log-likelihood and Bayesian information criterion, BIC) and classification accuracy. Classification accuracy was evaluated by comparing the proportional-by-chance accuracy rate of the data with each model's classification accuracy rate [30]. The proportional-by-chance accuracy rate is calculated by summing the square of the proportions of the categories of the dependent variable (i.e., proportion of cohort in each cost group). Models with at least 25% improvement over the proportional-by-chance accuracy rate were accepted as having adequate classification accuracy [30]. To compare our results with previous studies [31, 32], we conducted two pairwise logistic regression models using the same predictors as discussed above, comparing the c statistic from these models. The first model compared the persistently high-cost group with the

TABLE 1
Episode of care cost components by cost group and episode of care following chronic obstructive pulmonary disease diagnosis

	Index episode					Follow-up episode					
	Hospital	Emergency department	Physician	Drugs	Total episode cost	Hospital	Emergency department	Physician	Drugs	Total episode cost	
Persistently high cost (n = 100)											
Mean costs (SD)	11665.57 (12329.75)	191.06 (298.43)	448.58 (675.60)	144.77 (405.02)	12449.99 (13183.42)	15255.92 (33406.66)	264.59 (308.48)	500.47 (769.45)	74.70 (138.88)	16095.70 (33957.90)	
Mean costs, given any use (SD) ^a	11665.57 (12329.75)	415.35 (317.34)	540.46 (707.68)	268.10 (522.23)	12449.99 (13183.42)	15255.92 (33406.66)	426.77 (289.99)	562.33 (794.28)	162.40 (166.95)	16095.70 (33957.90)	
Occasionally high cost (n = 309)											
Mean costs (SD)	8785.21 (25306.51)	153.59 (238.84)	304.99 (484.90)	90.82 (168.78)	9334.61 (25644.98)	7870.76 (21323.29)	175.32 (341.13)	287.84 (480.82)	106.25 (257.39)	8440.17 (21471.14)	
Mean costs, given any use (SD) ^a	11751.65 (28680.88)	402.20 (222.10)	354.29 (505.72)	147.70 (194.87)	9334.61 (25644.98)	13740.48 (26731.71)	451.44 (418.93)	316.52 (495.18)	153.42 (297.54)	8440.17 (21471.14)	
Persistently low cost (n = 773)											
Mean costs (SD)	1015.67 (1933.40)	56.73 (133.31)	116.52 (166.91)	74.52 (140.11)	1263.45 (2028.50)	462.83 (1364.14)	70.34 (185.77)	102.69 (183.56)	78.54 (124.14)	714.40 (1450.62)	
Mean costs, given any use (SD) ^a	4538.24 (835.73)	339.94 (100.79)	125.27 (169.87)	91.73 (150.30)	1263.45 (2028.50)	4259.17 (971.23)	385.63 (260.49)	110.40 (188.08)	90.88 (129.28)	714.40 (1450.62)	

^aMean costs were estimated only in those who utilized the listed healthcare services, not in the total number of individuals in the cost group. Costs are reported in CAD constant dollars for 2011–2012.

persistently low-cost group whilst the second compared the occasionally high-cost group with the persistently low-cost group. SAS® version 9.3 (SAS Institute Inc., Cary, NC, USA) was used for all analyses.

RESULTS

Cohort selection and characteristics

A total of 12 543 COPD cases were identified between April 1, 2007 and March 31, 2011. After exclusion criteria were applied (i.e., previous healthcare utilization with a COPD diagnosis within a 5-year look-back period (38.0%), and not having continuous provincial health insurance coverage (5.4%)), a total of 7099 individuals were eligible for study inclusion. During an average follow-up time of 3.7 years, 2659 individuals had a total of 5348 episodes. The final cohort (n = 1182) was comprised of all individuals with at least two COPD episodes of care during the follow-up period.

Based on the 75th percentile cutoff of the cumulative total episode cost distribution, 100 (8.5%) patients were classified as persistently high cost, 309 (26.1%) as occasionally high cost, and 773 (65.4%) as persistently low cost. The average time between the last date of the index episode and the first date of the follow-up episode was longer for the persistently high-cost patients (374.2 days; SD = 361.8 days) than for the occasionally high-cost (351.2; SD = 325.0) and persistently low-cost (341.9; SD = 313.9) patients. In general, patients in the persistently high-cost group were older (74.6 years; SD = 11.7 years) than those in the occasionally high-cost (71.8 years; SD = 12.0 years) and persistently low-cost (65.5 years; SD = 12.5 years) groups. The persistently high-cost group was composed of 52.0% males, and this percentage was similar for the other two cost groups.

Episode costs

Average episode costs are summarized in Table 1. The persistently high-cost patients incurred about 10 times more costs than the persistently low-cost patients in the index episode (\$12 449.99 vs \$1263.45). Similarly, the occasionally high-cost patients incurred a little over seven times more costs than the persistently low-cost patients in the index episode (\$9334.61 vs \$1263.45). Hospital cost was the major component of total episode costs. Specifically, it constituted over 90% of total costs for the

persistently high-cost and occasionally high-cost patients. However, when hospital costs were excluded from total episode costs, the persistently high-cost group still had higher costs than the other two cost groups. Similarly, patients in the persistently high-cost group incurred higher average total costs in the follow-up episode than patients in the other groups.

Health services utilization during episodes

The number of hospital admissions, SCU admissions, ED visits, FP visits, and specialist visits were significantly different among the three cost groups in both episodes ($p < 0.001$ for all services) (Table 2). All patients in the persistently high-cost group were admitted to hospitals during both episodes, whilst lower percentages of patients in the occasionally high-cost and persistently low-cost groups were hospitalized during these episodes. Similarly, a higher percentage of patients in the persistently high-cost group was admitted to SCUs during their hospitalizations and had ED and specialist visits more than patients in the other two groups. However, a higher percentage of patients in the persistently low-cost group utilized more FP services and out-patient drug dispensations than patients in the two high-cost groups.

Patients in the persistently high-cost group had the longest hospital and SCU stays, followed by the occasionally high-cost group and then the persistently low-cost group (Figure 1). The average number of days in EDs was similar among the cost groups. Overall, the average number of days in episodes was higher in the persistently high-cost group than in the other two cost groups.

Multinomial logistic regression results

In the multinomial logistic regression models (Table 3), compared with patients who were 75+ years of age, those in age group 35–54 years (odds ratio (OR) = 0.19, 95% CI 0.09–0.41) or age group 55–74 years (OR = 0.53, 95% CI 0.33–0.85) were much less likely to be in the persistently high-cost group than the persistently low-cost group. Also, compared with those with no comorbid conditions, patients with a Charlson comorbidity score of 1 (OR = 2.68, 95% CI 1.51–4.77), 2 (OR = 2.28, 95% CI 1.17–4.42), or ≥ 3 (OR = 4.29, 95% CI 2.30–8.00) were more likely to be in the persistently high-cost group than the persistently low-cost group.

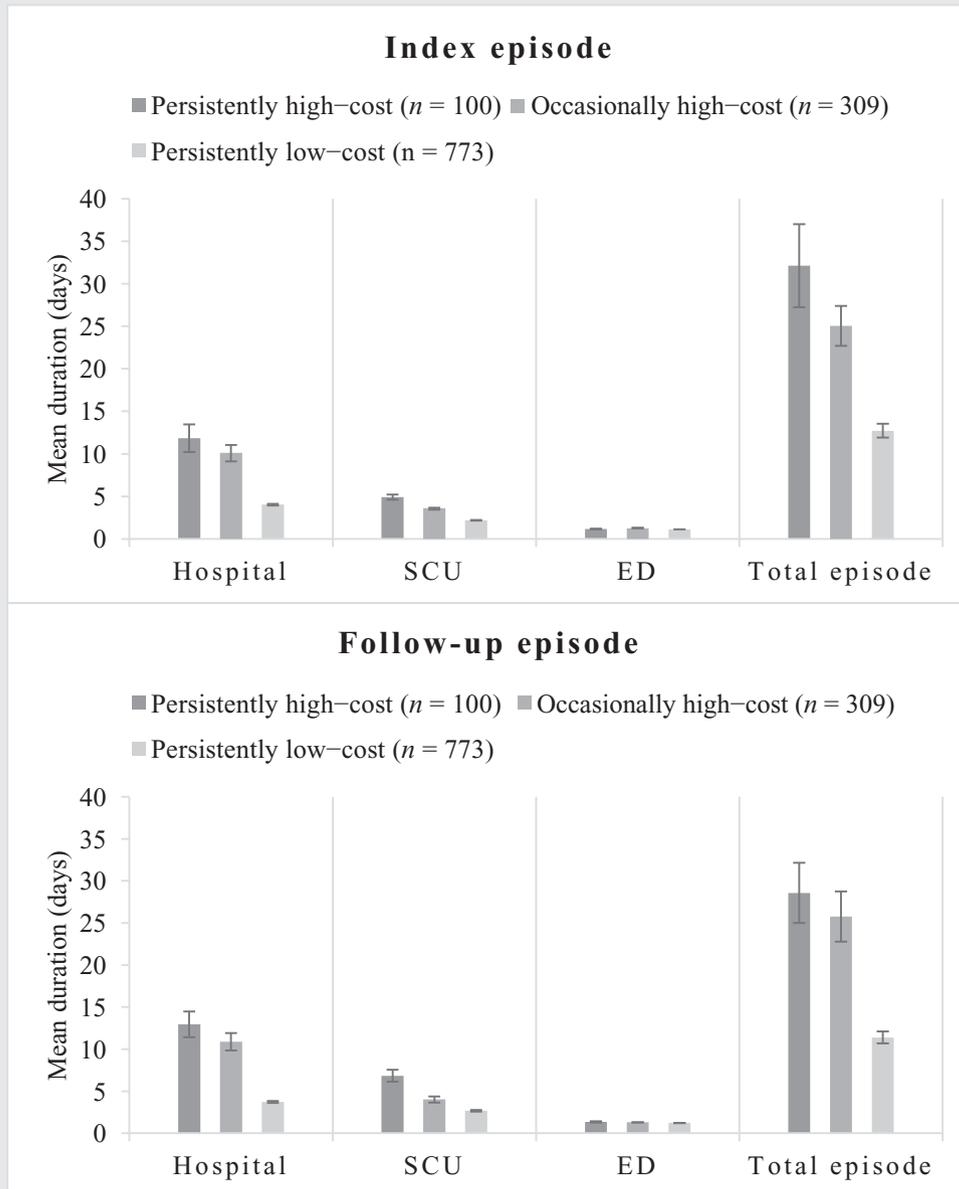
TABLE 2
Frequency of healthcare services utilization by episode cost group

	Index episode (%)			Follow-up episode (%)		
	Persistently high cost (n = 100)	Occasionally high cost (n = 309)	Persistently low cost (n = 773)	Persistently high cost (n = 100)	Occasionally high cost (n = 309)	Persistently low-cost (n = 773)
No. of hospital admissions*						
0	0.0	25.2	77.2	0.0	42.4	88.6
1	85.0	68.3	22.8	92.0	50.8	11.4
2+	15.0	6.5	0.0	8.0	6.8	0.0
No. of SCU admissions*						
0	89.0	93.5	98.7	89.0	93.8	99.2
1+	11.0	6.5	1.3	11.0	6.2	0.8
No. of ED visits*						
0	54.0	61.8	83.3	38.0	61.2	81.7
1	38.0	31.4	15.7	51.0	32.4	15.8
2+	8.0	6.8	1.0	11.0	6.4	2.5
No. of FP visits*						
0	39.0	31.0	15.7	38.0	24.9	12.7
1	19.0	24.0	43.6	19.0	33.3	48.9
2+	42.0	45.0	40.7	43.0	41.8	38.4
No. of specialist visits*						
0	40.0	51.8	77.1	35.0	54.7	81.4
1	8.0	11.7	11.0	13.0	16.2	10.2
2+	52.0	36.5	11.9	52.0	29.1	8.4
No. of different drugs						
Mean (SD)	10.0 (11.5)	7.1 (7.4)	4.7 (6.3)	9.5 (11.4)	8.6 (18.5)	4.3 (5.0)
Median	6	4	3	7	5	3

Note: *Utilization distributions in the three cost groups are significantly different using a χ^2 test at $p < 0.001$. SCU, special care unit; ED, emergency department; FP, family practitioner; SD, standard deviation.

FIGURE 1

Average health services utilization and episode durations by cost group. ED, emergency department; SCU, special care unit.



Similarly, patients in age groups 35–54 or 55–74 years were less likely to be in the occasionally high-cost group than the persistently low-cost group compared with patients aged 75 years and above. Again, compared with those with no comorbid conditions, patients with Charlson comorbidity scores of 1, 2, or ≥ 3 were more likely to be in the occasionally high-cost group than the persistently low-cost group. The associations of sex, residence location, and fiscal year of the COPD diagnosis with cost group membership were not statistically significant.

Models' prediction performance

Model 1 (i.e., the model containing patients' demographic and disease characteristics as well as the number of hospital admissions in the first episode) had the best fit to the data based on the BIC (Table 4). Although the classification accuracy differed substantially across the multinomial logistic regression models, each of the models provided more than 25% improvement over the proportional-by-chance accuracy rate of 0.50 for

our data. Thus, all the models had adequate classification; but model 1 had the highest classification accuracy rate. The *c* statistic from the logistic regression models ranged from 0.74 to 0.88 for the models comparing persistently high cost with persistently low cost, and from 0.68 to 0.83 for the models comparing occasionally high cost with persistently low cost.

DISCUSSION

In this study, we estimated the healthcare costs associated with episodes of COPD exacerbation and examined high-cost persistence using population-based administrative health data from Saskatchewan, Canada. By using the episode of care as the unit of analysis, our study uniquely characterizes the critical link between utilization patterns and healthcare costs. The episode-of-care approach reveals how the use of different services are related during COPD exacerbations. This provides a comprehensive understanding of the key drivers of overall episode-of-care costs associated with COPD exacerbations.

TABLE 3

Baseline characteristics of the study cohort and ORs from the multinomial logistic regression models

Characteristic	Persistently high-cost patients	Occasionally high-cost patients	Persistently low-cost patients	All	Persistently high-cost patients	Occasionally high-cost patients
	(n = 100) No. (%)	(n = 309) No. (%)	(n = 773) No. (%)	(n = 1182) No. (%)	(n = 100) ^a ORs (95% CIs)	(n = 309) ^a ORs (95% CIs)
Age group, y						
35–54	9 (9.0)	45 (14.6)	250 (32.3)	304 (25.7)	0.19 (0.09–0.41)*	0.35 (0.23–0.52)*
55–74	37 (37.0)	129 (41.8)	317 (41.0)	483 (40.9)	0.53 (0.33–0.85)*	0.71 (0.52–0.96)*
75+	54 (54.0)	135 (43.7)	206 (26.7)	395 (33.4)	ref	ref
Sex						
Female	52 (52.0)	148 (47.9)	375 (48.5)	575 (48.7)	1.25 (0.81–1.93)	1.03 (0.78–1.36)
Male	48 (48.0)	161 (52.1)	398 (51.5)	607 (51.4)	ref	161 (52.1)
Residence location						
Urban	21 (21.0)	77 (24.9)	191 (24.7)	893 (75.5)	1.33 (0.79–2.25)	1.05 (0.76–1.45)
Rural	79 (79.0)	232 (75.1)	582 (75.3)	289 (24.5)	ref	ref
Charlson comorbidity index						
0	42 (42.0)	155 (50.2)	567 (73.4)	764 (64.6)	ref	ref
1	22 (22.0)	70 (22.7)	91 (11.8)	183 (15.5)	2.68 (1.51–4.77)*	2.43 (1.68–3.52)*
2	15 (15.0)	40 (12.9)	66 (8.4)	121 (10.2)	2.28 (1.17–4.42)*	1.83 (1.17–2.86)*
≥3	21 (21.0)	44 (14.2)	49 (6.3)	114 (9.6)	4.29 (2.30–8.00)*	2.67 (1.69–4.22)*
Fiscal year of COPD diagnosis						
2007–2008	27 (27.0)	90 (29.1)	266 (34.4)	383 (32.4)	ref	ref
2008–2009	29 (29.0)	87 (28.2)	194 (25.1)	310 (26.2)	1.22 (0.68–2.16)	1.18 (0.82–1.69)
2009–2010	24 (24.0)	84 (27.2)	158 (20.4)	266 (22.5)	1.45 (0.79–2.64)	1.56 (1.07–2.25)
2010–2011	20 (20.0)	48 (15.5)	155 (20.1)	223 (18.9)	1.12 (0.60–2.11)	0.85 (0.56–1.29)

Note: *Statistically significant at $\alpha = 0.05$. OR, odds ratio; COPD, chronic obstructive pulmonary disease.

^aReference group was persistently low-cost patients.

TABLE 4

Comparison of goodness-of-fit and classification accuracy between models

Performance metric	Base model	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Goodness-of-fit, multinomial models							
–2 Log-likelihood	1842.98	1455.33	1762.55	1835.90	1711.69	1833.55	1429.88
BIC	1998.64	1625.13	1932.33	2005.70	1881.49	2003.35	1656.28
Classification accuracy, multinomial models							
PBCAR ^a	0.50	0.50	0.50	0.50	0.50	0.50	0.50
Model classification accuracy	0.67	0.72	0.66	0.67	0.69	0.67	0.71
% improvement over PBCAR	34.0	44.0	32.0	34.0	38.0	34.0	42.0
Logistic regression comparing persistently high-cost with persistently low-cost (n = 873)							
c statistic	0.74	0.87	0.80	0.75	0.85	0.76	0.88
Logistic regression comparing occasionally high-cost with persistently low-cost (n = 1082)							
c statistic	0.68	0.82	0.73	0.69	0.74	0.68	0.83

Notes: Base model = age, sex, residence, Charlson comorbidity index; Model 1 = base model + no. of hospital admission in index episode; Model 2 = base model + no. of emergency department visits in index episode; Model 3 = base model + no. of family practitioner visits in index episode; Model 4 = base model + no. of specialist visits in index episode; Model 5 = base model + no. of drugs dispensed in index episode; Model 6 = base model + no. of all the above healthcare services in index episode. BIC, Bayesian information criterion; PBCAR, proportional-by-chance accuracy rate.

^aPBCAR = $(100/1182)^2 + (309/1182)^2 + (773/1182)^2 = 0.50$.

The average episode of care costs for the persistently high-cost patients were between 10 and 22 times higher than that of the persistently low-cost patients in the baseline and follow-up episodes, respectively. Similarly, the average episode costs for the occasionally high-cost patients were between 7 and 12 times higher than that of the persistently low-cost patients in the baseline and follow-up episodes, respectively. Although overall average cost was lower in the follow-up episode for the entire cohort, this cost increased for the persistently high-cost patients by 29.3%. The increase in costs among persistently high-cost patients is likely due to the increase in hospital length of stay (i.e., number of days in hospital) as well as the number of days spent in specialized units during hospitalizations in the follow-up episode.

Previous studies [19] showed that older patients are more likely than younger ones to be in the persistently high-cost group. We found that older age (75+ years) was associated with both persistently high-cost and occasionally high-cost groups. Although long-term care is expensive and usually places its users in the high-cost group, this care setting is deemed the most appropriate for the frail elderly, who are typically not

the focus of intensive case management interventions [20]. Instead of including home care or long-term care costs in the total episode costs, we rather calculated the proportion of patients who were users of these services before or during their episodes of care and found that only 11.0% of the persistently high-cost patients used these services. Thus, it is likely that the great majority of the persistently high-cost patients might be suitable candidates for case management interventions.

Being able to predict whether individual patients will continue to incur high healthcare costs over time is useful for understanding patterns of healthcare utilization and identifying individuals for case management interventions [33]. We found that each of the multinomial logistic regression models compared in our study had more than 25% improvement over the proportional-by-chance accuracy rate, demonstrating that each of these models had adequate classification accuracy. However, model 1 (i.e., the model containing patient demographic and disease characteristics as well as the number of hospital admissions in the previous episode) had the highest classification accuracy rate and should be preferred over the other models. Unlike our study, previous studies

have developed logistics regression models to predict patients who might become high-cost users in the future, with *c* statistics ranging from 0.81 to 0.85 [31, 32]. For comparison purposes, we also developed logistic regression models and found that the models that predicted persistently high-cost patients had *c* statistics ranging between 0.74 and 0.88, whilst those that predicted occasionally high-cost patients had *c* statistics ranging from 0.68 to 0.83. One of the key differences between the *c* statistic reported in our study and those in the cited studies is that the cited studies did not distinguish between persistently high-cost and occasionally high-cost patients. Our results indicate that predictions of the persistently high-cost group, the group more likely to benefit from case management interventions, are more accurate compared with the occasionally high-cost group.

The study has some limitations. First, a common limitation of studies that use administrative health data to construct episodes of care is the inability to make distinctions between scheduled and unscheduled visits to healthcare providers; this information is not routinely collected in some databases such as ED databases [34]. A second potential limitation of the study is that we only considered a clean period of 30 days to distinguish one episode from another, although this is a common approach to defining episodes of care [35]. Scheduled visits beyond 30 days may be counted as part of a new episode. However, recommended practice [7] suggests that follow-up visits be scheduled within two to four weeks of discharge from acute care; hence, the possibility of scheduled visits distorting our episode construction may be minimal. Third, we used simple average costs for some cost components such as ED costs. This did not take acuity or complexity of patients' conditions into account. However, hospitalization, which was the major component of episode costs, was based on a standard methodology developed by CIHI to reflect variations in resource utilization. Fourth, the prediction accuracy of the models compared in this study was based on the model building dataset only. There is the need to validate these models in independent datasets. Fifth, our study did not include all potential confounders such as smoking status, physical activity, and body mass index. Our inability to account for these variables, because they were not routinely collected in the data sources used in our study, may possibly have led to spurious findings. Future research should consider including these potential confounders. Sixth, there was the possibility of underestimating healthcare utilizations and costs if patients sought treatment outside the two health regions included in the study. However, given that these health regions contained the major urban centers with the main healthcare facilities, the likelihood of patients receiving treatment outside these regions may be minimal. Seven, there was a possibility of survival bias in our study, which could bias the results toward the null, particularly among the elderly age group. Lastly, the generalizability of the findings is limited to the health regions included in our study.

Despite these limitations, this study demonstrates a practical approach to link various administrative health databases to characterize healthcare costs of patients with a complex health condition. Healthcare costs have been increasing at an unsustainable rate in many jurisdictions, and some governments are currently instituting cost-controlling provider reimbursement reforms such as bundled payment, which pays providers for an entire episode of care [36, 37]. Understanding healthcare costs based on the episodes of care, as demonstrated in our study, is important for adopting new provider payments schemes.

CONCLUSION

The costs associated with episodes of COPD exacerbations are substantial; some patients incur high healthcare expenditures persistently. Adding prior hospitalizations to socio-demographic characteristics produced the highest improvements in classification accuracy of patients into their respective high-cost groups. Being able to identify persistently high-cost patients is important for implementing strategies to manage costs and improve quality of life.

REFERENCES

- Gershon AS, Guan J, Victor JC, Goldstein R, To T. Quantifying health services use for chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2013;187(6):596–601. doi: 10.1164/rccm.201211-2044OC.
- Burge S, Wedzicha J. COPD exacerbations: Definitions and classifications. *Eur Respir J Suppl* 2003;41:46s–53s. doi: 10.1183/09031936.03.00078002.
- Blanchette CM, Gross NJ, Altman P. Rising costs of COPD and the potential for maintenance therapy to slow the trend. *Am Health Drug Benefit* 2014;7:98–106.
- Strassels SA, Smith DH, Sullivan SD, Mahajan PS. The costs of treating COPD in the United States. *Chest* 2001;119(2):344–52. doi: 10.1378/chest.119.2.344.
- Seemungal T, Donaldson GC, Bhowmik A, Jeffries DJ, Wedzicha JA. Time course and recovery of exacerbations in patients with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2000;161(5):1608–13. doi: 10.1164/ajrccm.161.5.9908022.
- Sharifabad MA, Parsons J. COPD. *BMJ best pract*. Available at: <http://bestpractice.bmj.com/best-practice/monograph/7/follow-up/recommendations.html> (Accessed January 20, 2015).
- Hornbrook MC, Hurtado AV, Johnson RE. Health care episodes: Definition, measurement and use. *Med Care Res Rev* 1985;42(2):163–218. doi: 10.1177/107755878504200202.
- Bassin E. Episodes of care a tool for measuring the impact of healthcare services on cost and quality. *Dis Manag Heal Outcomes* 1999;6(6):319–25. doi: 10.2165/00115677-199906060-00002.
- Miller HD. From volume to value: Better ways to pay for health care. *Health Aff* 2009;28(5):1418–28. doi: 10.1377/hlthaff.28.5.1418.
- Roos NP, Shapiro E, Tate R. Does a small minority of elderly account for a majority of health care expenditures? A sixteen-year perspective. *Milbank Q* 1989;67(3–4):347–69. doi: 10.2307/3350220.
- Riley GF. Long-term trends in the concentration of medicare spending. *Health Aff* 2007;26(3):808–16. doi: 10.1377/hlthaff.26.3.808.
- Mittmann N, Kuramoto L, Seung SJ, et al. The cost of moderate and severe COPD exacerbations to the Canadian healthcare system. *Respir Med* 2008;102(3):413–21. doi: 10.1016/j.rmed.2007.10.010.
- Roos LL, Gupta S, Soodeen R-A, Jebamani L. Data quality in an information-rich environment: Canada as an example. *Can J Aging* 2010;24(Suppl 1):153–170.
- Gershon A, Wang C. Identifying individuals with physician diagnosed COPD in health administrative databases. *COPD J Chronic Obstr Pulm Dis* 2009;6:388–94. doi: 10.1080/15412550903140865.
- Gershon AS, Wang C, Wilton AS, Raut R, To T. Trends in chronic obstructive pulmonary disease prevalence, incidence, and mortality in Ontario, Canada, 1996 to 2007: A population-based study. *Arch Intern Med* 2010;170(6):560–5. doi: 10.1001/archinternmed.2010.17.
- Canadian Institute for Health Information. Health indicators: Definitions, data sources and rationale. Ottawa, ON; 2011. Available at: http://www.cihi.ca/CIHI-ext-portal/pdf/internet/DEFINITIONS_062011_EN (Accessed March 24, 2015).
- Canadian Institute for Health Information. Cost of a standard hospital stay. Ottawa, ON; 2014. Available at: <http://indicatorlibrary.cihi.ca/display/HSPIL/Cost+of+a+Standard+Hospital+Stay> (Accessed January 15, 2015).
- Statistics Canada. Consumer price index. Ottawa, ON; 2014. Available at: <http://www.statcan.gc.ca/tables-tableaux/sum-som/101/cst01/econ161a-eng.htm> (Accessed April 20, 2015).
- Coughlin TA, Long SK. Health care spending and service use among high-cost medicare beneficiaries, 2002–2004. *Inquiry* 2009;46:405–17. doi: 10.5034/inquiryjrn1.46.4.405.
- Ronksley PE, McKay JA, Kobewka DM, Mulpuru S, Forster AJ. Patterns of health care use in a high-cost inpatient population in Ottawa, Ontario: A retrospective observational study. *CMAJ Open* 2015;3(1):E111–18. doi: 10.9778/cmajo.20140049.
- Chechulin Y, Nazerian A, Rais S, Malikov K. Predicting patients with high risk of becoming high-cost healthcare users in Ontario (Canada). *Healthc Policy* 2014;9(3):68–79. doi: 10.12927/hcpol.2014.23710.
- Fleishman JA, Cohen JW. Using information on clinical conditions to predict high-cost patients. *Health Serv Res* 2010;45(2):532–52. doi: 10.1111/j.1475-6773.2009.01080.x.
- Weir S, Awah G, Clark RE. Case selection for a medicare chronic care management program. *Health Care Financ Rev* 2008;30(1):61–74.
- Baser O, Burkan A, Baser E, et al. High cost patients for cardiac surgery and hospital quality in Turkey. *Health Policy* 2012;109(2):143–9. doi: 10.1016/j.healthpol.2012.09.015.
- Reschovsky JD, Hadley J, Saiontz-Martinez CB, Boukus ER. Following the money: Factors associated with the cost of treating high-cost medicare beneficiaries. *Health Serv Res* 2011;46(4):997–1021. doi: 10.1111/j.1475-6773.2011.01242.x.

26. Freeborn DK, Pope CR, Mullooly JP, McFarland BH. Consistently high users of medical care among the elderly. *Med Care* 1990;28(6):527-40. doi: 10.1097/00005650-199006000-00005.

27. Andersen RM. Revisiting the behavioral model and access to medical care: Does it matter? *J Health Soc Behav* 1995;36(1):1-10. doi: 10.2307/2137284.

28. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis* 1987;40(5):373-83. doi: 10.1016/0021-9681(87)90171-8.

29. Quail JM, Lix LM, Osman BA, Teare GF. Comparing comorbidity measures for predicting mortality and hospitalization in three population-based cohorts. *BMC Health Serv Res* 2011;11(1):146. doi: 10.1186/1472-6963-11-146.

30. Petrucci CJ. A primer for social worker researchers on how to conduct a multinomial logistic regression. *J Soc Serv Res* 2009;35(2):193-205. doi: 10.1080/01488370802678983.

31. Meenan RT, Goodman MJ, Fishman PA, et al. Using risk-adjustment models to identify high-cost risks. *Med Care* 2003;41(11):1301-12. doi: 10.1097/01.MLR.0000094480.13057.75.

32. DeSalvo KB, Jones TM, Peabody J, et al. Health care expenditure prediction with a single item, self-rated health measure. *Med Care* 2009; 47(4):440-7. doi: 10.1097/MLR.0b013e318190b716.

33. Monheit AC. Persistence in health expenditures in the short run: Prevalence and consequences. *Med Care* 2003;41(7 Suppl):III53-64. doi: 10.1097/00005650-200307001-00007.

34. Doupe M, Kozyrskyj A, Soodeen R-A, et al. An initial analysis of emergency departments and urgent care in Winnipeg. Available at: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:An+initial+analysis+of+emergency+department+and+urgent+care+in+Winnipeg#0> (Accessed January 7 2015).

35. Goodman AC, Hankin JR, Kalist DE, Peng Y, Spurr SJ. Estimating determinants of multiple treatment episodes for substance abusers. *J Ment Health Policy Econ* 2001;4(2):65-77.

36. Davis K, Guterman S. Rewarding excellence and efficiency in Medicare payments. *Milbank Q* 2007;85(3):449-68. doi: 10.1111/j.1468-0009.2007.00495.x.

37. DeJong G. Bundling acute and postacute payment: From a culture of compliance to a culture of innovation and best practice. *Phys Ther* 2010;90(5):658-62. doi: 10.2522/ptj.2010.90.5.658.

APPENDIX

TABLE A1

Definitions of healthcare utilization variables

Variable	Definition	Analysis in which variable was included	
		Descriptive	Prediction
No. of hospital admissions	The number of times a patient was admitted to hospital during episode	√	√
No. of days in hospital	Total number of days a patient spent in hospitals during episode	√	
No. of SCU admissions ^a	The number of times a patient was admitted to SCUs during hospital stays in the episode	√	
No. of days in SCUs ^a	Total number of days a patient spent in SCUs during hospital admissions in the episode	√	
No. of ED visits	The number of times a patient visited EDs during episode	√	√
No. of days in ED	Total number of days a patient spent in EDs during episode	√	
No. of FP visits	The number of times a patient visited FPs during episode	√	√
No. of specialist visits	The number of times a patient visited specialist physicians during episode	√	√
No. of different drugs	The number of different types of out-patient drugs dispensed during the episode	√	√
No. of days in episode	The total number of days the episode covered, starting from the first date of the episode to the last date	√	

Notes: ^aThese variables were defined for only those who had hospital admission. √ = variable was included in the specified analysis; SCU, special care units; ED, emergency department; FP, family practitioner.