USING THE EPISODE OF CARE APPROACH TO ANALYZE HEALTHCARE USE AND COSTS OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE EXACERBATIONS

By

John Paul Kuwornu

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

Department of Community Health Sciences

Faculty of Health Sciences

University of Manitoba

Winnipeg

ABSTRACT

Healthcare utilizations are typically measured independently of each other; neglecting the interdependencies between services. An episode of care is suitable for measuring healthcare utilizations of patients with complex health conditions because it tracks all contacts throughout the healthcare system. The overall goal of this research was to construct an episode of care data system to study healthcare utilizations and costs of chronic obstructive pulmonary disease (COPD) exacerbations. To achieve this goal, four related studies were undertaken.

The first study (Chapter 2) evaluated the agreement between emergency department (ED) data and hospital records for capturing transitions between the two care settings. Using the κ statistic as a measure of concordance, we found good agreement between the two data sources for intra-facility transfers; but only fair agreement for inter-facility transfers. The results show that linking multiple data sources would be important to identify all related healthcare utilization across care settings.

The second study (Chapter 3) linked hospital data, ED data, physician billing claims, and outpatient drug records to construct an episode of care data system for COPD patients. Latent class analysis was used to identify COPD patient groups with distinct healthcare pathways. Pathways were associated with outcomes such as mortality and costs. A few individuals followed complex pathways and incurred high costs.

Building on the previous study, the next one (Chapter 4) predicted whether high-cost patients in one episode also incurred high costs in subsequent episodes. Using logistic regression models, we found that patient information routinely collected in administrative health data could satisfactorily predict those who become persistent high users.

The final study (Chapter 5) used a cross-validation approach to compare the performance of eight alternative linear regression models for predicting costs of episodes of COPD exacerbations. The results indicate that the robust regression model, a model not often considered for cost prediction, was among the best models for predicting episode-based costs.

Overall, this research demonstrated how population-based administrative health databases could be linked to construct an episode of care data system for a chronic health condition. The resulting data system supported novel investigations of healthcare system-wide utilizations and costs.

ACKNOWLEDGMENTS

It was a great privilege to have Dr. Lisa Lix as my PhD advisor. I am very grateful for her outstanding intellectual guidance, professional direction and encouragement throughout my study. I am also very thankful to my PhD Advisory Committee: Dr. Evelyn Forget, Dr. Saman Muthukumarana, Dr. Jacqueline Quail, and Dr. Gary Teare, who rendered their invaluable support to my research. I am particularly grateful for their quick but thorough review comments on various chapters of my dissertation. Their feedback has improved my research skills tremendously.

I am also indebted to Research Analysts at the Saskatchewan Health Quality Council (HQC) who generously supported my numerous visits to HQC with their time and expertise. Especially, I want to thank Eric Wang, Meric Osman, Nedeene Hudema, and Xinya Lu. In addition, I am grateful for the support I had from students in the secure computer room at HQC; particularly from Shenzhen Yao who helped double-check my SAS codes.

A big thank you to Aynslie, Kristine, Shuman, and lately Lisa and Saeed for being very friendly and tolerant of my uninvited "stop-by-desk" mini talks in the Lab. Thanks to Josie who conscientiously kept all my administrative work in the Lab well-coordinated. I also want to thank Liz Cooper who has been a good study mate for our PhD candidacy examination. I will always remember the several chats I had with Dessalegn over coffee on political dynamics in Africa.

Lastly, I want to thank my family for their love, patience and support throughout my study. And above all, I am grateful to God for the gift of life to complete this work.

My study was supported by Western Regional Training Centre for Health Services Research (WRTC) Fellowship (2012 – 2013) and Research Manitoba Studentship (2014 – 2015).

DEDICATION

I dedicate this thesis to my parents, Winfried Kuwornu and Thecla Alorbi, who made great sacrifices to ensure that I had the opportunity to pursue my academic interests.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AHFS:	American Hospital Formulary System
BIC:	Bayesian-Schwarz Information Criterion
CI:	Confidence Interval
CIHI:	Canadian Institute for Health Information
CMG:	Case Mix Group
COPD:	Chronic Obstructive Pulmonary Disease
CPWC:	Cost per Weighted Case
DA:	Dissemination Area
DIN:	Drug Identification Number
ED:	Emergency Department
EMRs:	Electronic Medical Records
FEV_1:	Forced Expiratory Volume in One Second
FMM:	Finite Mixture Models
FVC:	Forced Vital Capacity
GLM:	Generalized Linear Model
GP:	General Practitioner
HQC:	Saskatchewan Health Quality Council
ICD-9:	International Classification of Diseases, 9^{th} Revision
ICD-10-CA:	International Classification of Diseases, 10^{th} Revision, Canada
LCA:	Latent Class Analysis
MAPE:	Mean Absolute Prediction Error
NACRS:	National Ambulatory Care Reporting System

OLS:	Ordinary Least Squares
OR:	Odds Ratio
QBPs:	Quality-Based Procedures
SHR:	Saskatoon Health Region
RIW:	Resource Intensity Weight
RQHR:	Regina Qu'Appelle Health Region
RMSE:	Root Mean Square Error
SABA:	Short-Acting Beta Agonists
SCU:	Special Care Units
SD:	Standard Deviation

CHAPTER 1: INTRODUCTION

1.1 Background

Healthcare utilizations are typically measured independently of each other, usually in units such as hospital length of stay or number of physician visits. However, different healthcare services are rarely utilized independently of each other. Particularly, patients with complex health conditions usually access care in a series of separate but related delivery encounters to treat or manage their condition. This series of encounters is termed an episode of care,[1] which is defined as a clinically homogenous period of care in which patients are tracked through different care settings to collect all temporally contiguous healthcare services for treating a specific health condition.[2] The episode of care is, therefore, a clinically meaningful unit of analysis to measure health system-wide utilizations and costs of healthcare services.[3]

An episode of care is typically constructed with linkable population-based administrative health databases to ensure that entire population of patients can be tracked; and with the availability of these databases, the episode of care has been adopted in some healthcare jurisdictions. For example, episodes of care are used to measure quality of care,[4] and implement new provider reimbursement schemes such as bundled payments in the US.[5] In Finland, the performance, effectiveness and cost of treatment episodes (PERFECT) project [6] uses episodes of care to measure health system-wide resource utilization for profiling providers. In Canada, Health Quality Ontario is currently developing quality-based procedures based on the episodes of care approach for developing new funding models for selected health conditions, including chronic obstructive pulmonary disease (COPD) (www.hqontario.ca/evidence/evidence-process/episodes-of-care).

COPD has been prioritized in many quality-of-care initiatives [4,7] because it is a leading cause of morbidity and is annually responsible for up to three million deaths worldwide.[8] Also, COPD patients usually experience multiple acute exacerbations of symptoms annually; with treatment often requiring an emergency department (ED) visit or hospitalization.[9,10] They may also receive follow-up care from their primary care provider or a specialist physician, and might require additional medications.[11] With this possibility of contacting different care providers, COPD patients are considered heavy users of healthcare services; and an Ontario study [12] showed that they were responsible for up to one-quarter of all hospitalizations and ED visits, and more than one-fifth of ambulatory visits between April 1, 2008 and March 31, 2011.

Accordingly, an episode of care data system is highly relevant for measuring and describing healthcare utilizations and costs for COPD patients.[13] Understanding the healthcare utilization patterns of these patients across the entire continuum of care is important for identifying opportunities for redesigning healthcare delivery to improve quality and/or lower costs.[14,15]

1.2 Objectives

The overall goal of this research was to use linkable population-based administrative health data from the province of Saskatchewan to construct an episode of care data system to measure and predict healthcare resource utilizations and costs, focusing on the characteristics of episodes of COPD exacerbations. The specific objectives were:

1. To evaluate the agreement between ED records and inpatient hospital data for capturing transitions from the emergency to the acute care settings.
2. To identify and describe care pathways within episodes of care for acute exacerbations of COPD.

3. To estimate healthcare costs associated with episodes of COPD exacerbations, and compare the ability of different models to predict whether a patient will incur high costs persistently.
4. To identify the optimal statistical model(s) for predicting costs of episodes of care for COPD exacerbations.

1.3 Unifying theoretical framework

Since this thesis is focused on patients' healthcare utilization and costs, we adopted the Andersen healthcare utilization model [16] as the unifying theoretical framework. Andersen proposed that healthcare utilization is influenced by an individual's predisposition to use services (i.e., predisposing factors), factors that support or inhibit use (i.e., enabling factors), and an individual's need for services (i.e., need factors). Predisposing factors include socio-demographic characteristics and health beliefs. Factors such as income, social support, availability and accessibility of healthcare services constitute the enabling factors. Finally, the need factors include health status or illness. Also, we used the episodes of care as the unifying unit of analysis to measure healthcare utilizations and costs.

1.4 Episodes of care

The episode of care as a unit of analysis was introduced in health services research in the 1960s,[17] and further developed by Hornbrook et al.[2] in the 1980s. Hornbrook et al.[2] defined an episode of care as a series of health-related events for a particular health problem that exists continuously for a delimited period of time. Thus, in constructing episodes of care, a patient has to be followed over time and across the healthcare system. Electronic medical records and administrative health databases are two potential complementary data resources for constructing episodes. While electronic medical records contain rich clinical and functional information which may be useful for defining severity of conditions, they lack the details of utilization

information such as cost. On the other hand, administrative health data usually contain both diagnosis and cost information but has limited clinical information.[18] In spite of this limitation, studies on episodes of care usually employ administrative health data.[19]

There are several ways to construct an episode of care data system, and no consensus has yet emerged on how to define its components.[2] Generally, an episode of care includes: (1) scope, the range of healthcare services that it includes, (2) trigger(s), the health condition(s) for which it is defined, and (3) demarcation, the separation of one episode from the other.[19]

The scope of an episode of care could either be defined by aggregating related healthcare services by the same provider over time,[1,20] or by including all contiguous services patients receive across multiple settings over a specified period of time; with the latter being the optimal approach to fully capture patients' contacts with the healthcare system.[19,21]

An episode of care could be defined for only one health condition or sets of related health conditions. Episodes of care constructed for multiple conditions are relevant for assessing the effects of multimorbidity, the co-occurrence of two or more conditions; while those constructed for a single condition may be more desirable for payment-related applications, because they focus on a more homogeneous group of patients.[7] However, episodes of care constructed for single health conditions is more common in practice.[22]

The choice of episode start date, episode end date, and demarcation between successive episodes are the other issues to consider in constructing an episode of care. An episode's beginning could be marked by the date on which a treatment is initiated,[2] or by the date of the most recent patient contact with a provider prior to the diagnosis of a health condition.[19] The end date of an episode is defined as the date on which the last care was provided in relation to the condition(s) under consideration. Deciding the demarcation between episodes is, however, more

difficult because it is condition-dependent.[19] For acute conditions, one approach used in the literature is based on clinical consensus; where a group of experts with clinical knowledge of the condition agree on the expected duration of treatment.[23] For chronic conditions, separating one episode from the other is usually based on what is called a "clean period"; a specified period of time in which no services related to the treatment of the health condition(s) were received.[19] Although there are no clear guidelines on selecting the duration of a clean period, a 30-day clean period is commonly used because readmission to inpatient care within 30 days is a key quality of care indicator usually monitored in many healthcare jurisdictions.[19]

An episode of care data system can be used to investigate a number of topics including, patient care pathways and healthcare costs. Because an episode of care data system tracks patients through the entire healthcare system, it allows for detailed study of the care pathways patients follow through care.[4] Care pathway analysis is important because it could identify routes through the healthcare system that would result in best outcomes.[14] A recent study used pathway analysis to identify routes to cancer diagnosis and compared health outcomes associated with each route.[24] Although the study used administrative health data, it did not investigate care pathways during well-defined episodes of care such as periods of COPD exacerbations. Also, an episode of care provides a clinically meaningful unit for measuring healthcare costs;[3] and allows for a detailed analysis of the treatment processes that generate these costs.[25]

1.5 Measurement of COPD exacerbations

COPD exacerbations negatively affect quality of life, impose a great burden on the patient, and represent considerable economic costs to society.[26] A number of studies have been undertaken to clearly define and measure COPD exacerbations. A commonly adopted definition of COPD exacerbation is "a sustained worsening of the patient's condition, from the stable state

and beyond normal day-to-day variations that is acute in onset and may warrant additional treatment in a patient with underlying COPD".[27]

Two different methods have been used to measure COPD exacerbations. They include symptoms-defined exacerbations and healthcare-defined exacerbations.[18] Measuring COPD exacerbations by symptoms usually involves interviewing patients on symptoms and treatments received over a given duration,[28] or asking patients to keep diaries of symptoms over a period of time.[29] On the other hand, measuring COPD exacerbations by healthcare utilization records involves identifying specified diagnostic codes, procedures, therapies, and/or drug identification numbers recorded in administrative health databases. Symptoms-defined exacerbations using diaries are considered to be more precise, but they can only be applied in small studies involving patients who have the required cognitive skills and willingness to comply with daily recording of symptoms.[27] For large population-based studies, healthcare-defined exacerbation presents a more practical approach; and methods have been designed to enhance their accuracy.[30]

1.6 Healthcare cost of COPD exacerbations

Although studies have shown that exacerbations contribute significantly to the total costs of managing COPD,[28] estimating these costs is challenging because of the difficulty of determining the exact durations of exacerbations.[18] A few population-based studies that have attempted to estimate the cost of COPD exacerbations were US-based, and they commonly used large administrative data. For instance, Dalal et al.[31] used administrative health data covering 602 hospitals and estimated the average cost of COPD exacerbations to be US$ 647 for ED visits, US$ 7,242 for simple hospitalizations, and US$ 20,757 for complex hospitalizations requiring intensive care or intubation. Another study [18] which also used US-based administrative claims data estimated an average cost of US$ 17,016 for severe exacerbations (i.e., treatment involves

ED visit or hospitalization), and US$ 6,628 for non-severe exacerbations (i.e., treatment does not involve ED visit or hospitalization). Instead of estimating these costs based on episodes of care, the authors rather adopted patient-quarters as unit of analysis to simplify their cost estimation. Unlike the large population-based US studies, the only Canadian study found regarding costs of COPD exacerbation used a prospective observational study; which enrolled approximately 600 adults.[32] This study estimated the average cost of a moderate and severe exacerbations to be CA$ 641 and CA$ 9,557, respectively.

1.7 Statistical models for healthcare costs

Healthcare costs are positive values with unusually high proportions of zeroes,[33] which may denote no utilization of services for the selected health condition. Healthcare cost distribution is also usually positively skewed,[34] because a few patients incur very high costs. There is also a possibility of healthcare cost data having heterogeneous variances;[35] variances increasing over time or with the values of independent variables. These feature complicate the statistical modeling and prediction of healthcare costs; which are necessary for developing fair provider remuneration systems.[36,37]

As a result, previous studies [35,38,39] have compared a variety of statistical models for predicting healthcare costs. The ordinary least squares (OLS) regression model performs well in predicting mean costs, although the distribution of error terms tends to be non-normal.[38,40] Skewness in the cost distribution has been addressed using Box-Cox transformations [41] to normalize the distribution and stabilize its variance.[34] Although these transformations are useful for modeling purposes, they can make interpretation of the results difficult because the data are no longer in the original scale. The generalized linear model (GLM) has been an alternative approach to deal with skewness in cost data. GLM allows for flexibility in choosing more

appropriate distributions for skewed data.[42] Aside from the OLS and GLM models, some studies also investigated more flexible distributions such as the generalized gamma [43] and the generalized beta of the second kind [44,45] for predicting healthcare costs.

1.8 Organization of thesis

This thesis is composed of four related manuscripts; each addressing one of the specified research objectives. Given that a number of patients who visit EDs for COPD exacerbations may later be hospitalized during the same encounter, the first manuscript (Chapter 2) investigated the agreement between the ED database and hospital discharge abstracts for capturing transitions from the emergency to the acute care settings. This study is important because, in spite of the important roles EDs may play in managing acute exacerbations of COPD,[46] previous studies [47] which assessed the accuracy of Saskatchewan's administrative health databases did not include the ED data. As at the time this research was conducted, there was no provincial ED database. Only Saskatoon health region (SHR) and Regina Qu'Appelle health region (RQHR) had ED data available. SHR and RQHR are the two largest of twelve health regions in Saskatchewan, and together account for just over half of the provincial population. eHealth Saskatchewan created a database using the ED data contributed by both SHR and RQHR. This is the ED database referenced throughout this thesis.

After constructing the episode of care data system, the second manuscript (Chapter 3) focuses on defining and describing patient care pathways during episodes of COPD exacerbations. Care pathway analysis is useful for guiding quality and efficiency improvement initiatives by identifying routes through the healthcare system that may result in increased probability of survival and/or lowest cost. Measuring care pathways is, however, challenging because patients with complex health conditions often exhibit heterogeneity in their use of

healthcare services. The manuscript provides a model-based approach for defining and describing patient care pathways.

The third manuscript (Chapter 4) estimates healthcare costs associated with episodes of COPD exacerbations, and compares the ability of different models to predict whether a patient will incur high costs persistently. The phenomenon of few individuals accounting for more than 50% of healthcare costs was reported over three decades ago, and recent analyses also revealed that these high-cost patients are a heterogeneous sub-group. Some of these patients persistently incur high cost, while others occasionally incur high cost. Identifying these persistent high users and their utilization patterns would be important for timely interventions.

The final manuscript (Chapter 5) compares the prediction accuracy of eight linear regression models to identify the best model(s) for predicting costs of an episode of care for COPD exacerbations. Being able to accurately predict episodes of care costs is essential for designing innovative provider reimbursement schemes such as bundled payments. The thesis ends with a summary of findings and directions for further studies (Chapter 6).

Ethics approvals to conduct the research were obtained from the Health Research Ethics Board, University of Manitoba (Appendix A) and Biomedical Research Ethics Board, University of Saskatchewan (Appendix B).

Each manuscript describes the research methodologies employed and the characteristics of the study samples. Because the studies are related, some repetitions occurred across the various chapters. No permissions are required to reprint the article (Chapter 3) published in *Medicine* and the article (Chapter 5) published in *Health Services and Outcomes Research Methodology*.

1.9 References

1. Casale AS, Paulus RA, Selna MJ, et al. "ProvenCareSM": a provider-driven pay-for-performance program for acute episodic cardiac surgical care. *Ann Surg*. 2007;246(4):613–21.

2. Hornbrook MC, Hurtado AV, Johnson RE. Health care episodes: definition, measurement and use. *Med Care Res Rev*. 1985;42(2):163–218.

3. 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):319–325.

4. National Quality Forum. Measurement framework: evaluating efficiency across patient-focused episodes of care. Washington, DC. 2009. Available at: http://www.qualityforum.org/Publications/2010/01/Measurement_Framework__Evaluating_Efficiency_Across_Patient-Focused_Episodes_of_Care.aspx. Accessed January 12, 2014.

5. Cutler DM, Ghosh K. The potential for cost savings through bundled episode payments. *N Engl J Med*. 2012;366(12):1075–1077.

6. Peltola M, Juntunen M, Häkkinen U, et al. A methodological approach for register-based evaluation of cost and outcomes in health care. *Ann Med*. 2011;43 Suppl 1(January):S4–13.

7. Health Quality Ontario & Ministry of Health and Long-Term Care. Quality-based procedures: clinical handbook for chronic obstructive pulmonary disease. Toronto, ON. 2013. Available at: http://health.gov.on.ca/en/pro/programs/ecfa/docs/qbp_copd.pdf. Accessed January 15, 2015.

8. World Health Organization. Chronic obstructive pulmonary disease. Geneva. 2015. Available at: http://www.who.int/mediacentre/factsheets/fs315/en/. Accessed June 20, 2015

9. Seemungal TA, 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):1608–1613.

10. Strassels SA, Smith DH, Sullivan SD, Mahajan PS. The costs of treating COPD in the United States. *Chest*. 2001;119(2):344–352.

11. Sharifabad MA, Parsons JP. COPD. *BMJ Best Practice*. Oct 2014. Available at: http://bestpractice.bmj.com/best-practice/monograph/7/follow-up/recommendations.html. Accessed January 20, 2015.

12. 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.

13. Mechanic D, Schlesinger M, McAlpine DD. Management of mental health and substance abuse services: state of the art and early results. *Milbank Q*. 1995;73(1):19–55.

14. Adeyemi S, Demir E, Chaussalet T. Towards an evidence-based decision making healthcare system management: modelling patient pathways to improve clinical outcomes. *Decis Support Syst*. 2013;55(1):117–125.

15. Dobson A, DaVanzo JE, El-Gamil A, et al. Clinically appropriate and cost-effective placement study. *Final Report by Dobson DaVanzo & Associates, LLC*. 2012. Available at: http://ahhqi.org/images/pdf/cacep-report.pdf. Accessed December 22, 2014.

16. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? *J Health Soc Behav*. 1995;36(1):1–10.

17. Solon J, Feeney J. Delineating episodes of medical care. *Am J Public Health*. 1967;57(3):401–408.

18. Yu AP, Yang H, Wu EQ, et al. Incremental third-party costs associated with COPD exacerbations: a retrospective claims analysis. *J Med Econ*. 2011;14(3):315–323.

19. 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.

20. Liu CF, Subramanian S, Cromwell J. Impact of global bundled payments on hospital costs of coronary artery bypass grafting. *J Health Care Finance*. 2001;27(4):39–54.

21. Haas-Wilson D. Demand for mental health services: an episode of treatment approach. *South Econ J*. 1989;56(1):219–232.

22. Wells KB, Keeler E, Manning WG. Patterns of outpatient mental health care over time: some implications for estimates of demand and for benefit design. *Health Serv Res*. 1990;24(6):773–89.

23. Moscovice I. A method for analyzing resource use in ambulatory care settings. *Med Care*. 1977;15(12):1024–1044.

24. Elliss-Brookes L, McPhail S, Ives A, et al. Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. *Br J Cancer*. 2012;107(8):1220–6.

25. Miller HD. From volume to value: better ways to pay for health care. *Health Aff*. 2009;28(5):1418–1428.

26. Effing TW, Kerstjens HAM, et al. Definitions of exacerbations: does it really matter in clinical trials on COPD? *Chest*. 2009;136(3):918–923.

27. Burge S, Wedzicha JA. COPD exacerbations: definitions and classifications. *Eur Respir J Suppl*. 2003;41:46s–53s.

28. Andersson F, Borg S, Jansson SA, et al. The costs of exacerbations in chronic obstructive pulmonary disease (COPD). *Respir Med*. 2002;96(9):700–708.

29. O'Reilly JF, Williams AE, Holt K, Rice L. Defining COPD exacerbations: impact on estimation of incidence and burden in primary care. *Prim Care Respir J*. 2006;15(6):346–353.

30. 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 October 24, 2014.

31. Dalal AA, Shah M, D'Souza AO, Rane P. Costs of COPD exacerbations in the emergency department and inpatient setting. *Respir Med*. 2011;105(3):454–460.

32. 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.

33. Mullahy J. Econometric modeling of health care costs and expenditures. *Med Care*. 2009;47(7):104–108.

34. Basu A, Manning WG. Issues for the next generation of health care cost analyses. *Med Care*. 2009;47(7 Suppl 1):S109–14.

35. Dodd S, Bassi A, Bodger K, Williamson P. A comparison of multivariable regression models to analyse cost data. *J Eval Clin Pract*. 2006;12(1):76–86.

36. Davis K. Paying for care episodes and care coordination. *N Engl J Med*. 2007;356(11):1166–8.

37. Averill RF, Goldfield NI, Hughes JS, Eisenhandler J, Vertrees JC. Developing a prospective payment system based on episodes of care. *J Ambul Care Manage*. 2009;32(3):241–51.

38. Dunn G. Describing, explaining or predicting mental health care costs: a guide to regression models: methodological review. *Br J Psychiatry*. 2003;183(5):398–404.

39. Larimer ME, Malone DK, Garner MD, et al. Health care and public service use and costs before and after provision of housing for chronically homeless persons with severe alcohol problems. *JAMA*. 2009;301(13):1349–57.

40. Kilian R, Matschinger H, Löeffler W, Roick C, Angermeyer MC. A comparison of methods to handle skew distributed cost variables in the analysis of the resource consumption in schizophrenia treatment. *J Ment Health Policy Econ*. 2002;5(1):21–31.

41. Cox DR, Box GEP. An analysis of transformation. *J R Stat Soc Ser B*. 1964;26:211–243.

42. Diehr P, Yanez D, Ash A, Hornbrook M, Lin DY. Methods for analyzing health care utilization and costs. *Annu Rev Public Health*. 1999;20:125–44.

43. Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. *J Health Econ*. 2005;24(3):465–88.

44. McDonald J, Xu Y. A generalization of the beta distribution with applications. *J Econom*. 1995;66:133–152.

45. Sun J, Frees EW, Rosenberg MA. Heavy-tailed longitudinal data modeling using copulas. *Insur Math Econ*. 2008;42(2):817–830.

46. Hasegawa K, Tsugawa Y, Tsai C-L, Brown DF, Camargo CA. Frequent utilization of the emergency department for acute exacerbation of chronic obstructive pulmonary disease. *Respir Res*. 2014;15(1):40.

47. Lix LM, Yan L, Blackburn D, et al. Validity of the RAI-MDS for ascertaining diabetes and comorbid conditions in long-term care facility residents. *BMC Health Serv Res*. 2014;14:17.

CHAPTER 2: MEASURING CARE TRAJECTORIES USING HEALTH ADMINISTRATIVE DATABASES: AN INVESTIGATION OF TRANSITIONS FROM EMERGENCY TO ACUTE CARE

2.1 Chapter overview

This manuscript investigates the agreement between emergency department (ED) data and inpatient hospital records by assessing the accuracy of patient transition information between the ED and inpatient hospitalization. Using the hospital discharge abstracts as the reference, patient transition information recorded in the ED data were compared with those recorded in the hospital discharge abstracts. This study is important for two reasons. First, the ED may serve as the first point of contact with the healthcare system for patients with acute chronic obstructive pulmonary disease (COPD) exacerbations; making the ED an important service in describing patients' journey through care. Second, it provides justification for constructing episodes of care data system by linking multiple databases to describe patients' utilization across the healthcare system.

Publication details:

Kuwornu JP, Lix LM, Quail JM, Wang XE, Osman M, Teare GF. Measuring care trajectories using health administrative databases: an investigation of transitions from emergency to acute care. *BMC Health Services Research* [Revised and resubmitted]

2.2 Abstract

Objective: This study describes patient transitions from the emergency department (ED) to other healthcare settings and estimates agreement between ED and inpatient hospital records for capturing transitions from emergency to acute care services.

Methods: Administrative health databases from the province of Saskatchewan, Canada (population 1.1 million) were used to identify patients with at least one ED visit to provincial teaching hospitals (n = 5) between April 1, 2006 and March 31, 2012. Discharge disposition from ED was described using frequencies and percentages; and it includes categories such as home, transfer to other facilities, and died. The kappa statistic with 95% confidence intervals (95% CIs) was used to measure agreement between the discharge disposition on the ED record and mode of entry on the hospitalization record.

Results: We identified N=1,062,861 visits for 371,480 patients to EDs over the six-year study period. Three-quarters of the discharges were to home, 16.1% were to acute care in the same facility in which the ED was located, and 1.6% resulted in a patient transfer to a different acute care facility. Agreement between ED and inpatient hospital records for measuring patient transitions was good when the emergency and acute care departments were in the same facility (κ = 0.77, 95% CI 0.77, 0.77). For transfers to a different acute care facility, agreement was fair (κ = 0.36, 95% CI 0.35, 0.36).

Conclusions: The majority of patients who attend EDs do not transition to another healthcare setting. For those who transition to acute care, agreement between ED and hospital records depends on whether the two services are in the same facility. The results suggest that ED data does not contain accurate information about transitions to acute care services in different

facilities. Therefore, data linkage is required to accurately identify patient trajectories across different healthcare settings.

2.3 Background

A patient's care trajectory,[1] the patient's sequence of contacts with care providers, is important to study because it is expected to affect resource use [2] and healthcare outcomes.[3,4] Accurate construction of care trajectories is particularly useful for anticipating the needs of patients and allocating healthcare resources.[5] However, there are no standard methods of constructing care trajectories. Population-based administrative health databases are potentially valuable resources for constructing care trajectories for entire populations, provided they can accurately capture patient transitions between healthcare services.

Previous studies have focused on the quality of individual administrative health databases,[6] and examined their utility for capturing patient data in single healthcare settings.[7] Only a few studies have compared multiple databases across settings, either by evaluating the accuracy of the linkage process for such data sources as ambulance, emergency department (ED), and hospital,[8] or by examining the completeness of integrated data sets to describe the patient's journey through the healthcare system.[9] No study, to the best of our knowledge, has described the accuracy of administrative health data for capturing patient transition information between healthcare services.

This study investigated whether the discharge disposition recorded in the ED data was a reliable source of patient transition information for constructing care trajectories. Specifically, the objectives were to: (a) describe patient transitions from the ED to other healthcare settings, and (b) estimate agreement between ED and inpatient hospital records for capturing transitions from the emergency to the acute care settings. The ED often serves as the starting point for the

receipt of services in the care trajectory, particularly among patients without a regular source of primary care.[10] ED encounters may require admission to hospital; one of the most expensive care settings. Thus, the ED and hospital are two of the most important healthcare services in describing patient care trajectories.

2.4 Methods

2.4.1 Data sources

ED data were obtained from all teaching hospitals (n = 5) in the province of Saskatchewan, Canada, which has a population of approximately 1.1 million.[11] This province, like other Canadian provinces, has a system of universal healthcare, which means that virtually all residents of the province are eligible for health insurance coverage. Only non-residents and individuals such as inmates in federal penitentiaries and members of the armed forces are not covered under the provincial insurance program. The teaching hospitals are located in Saskatoon and Regina Qu'Appelle health regions, two of 12 health regions in the province and the only regions that contain major urban centres (population > 200,000 in each centre). Three EDs started capturing patient data in electronic records in April 1, 2002, and by April 1, 2006 all five EDs were doing so. This study focused on the following components of the electronic ED data: location, visit and discharge dates, and discharge disposition. Discharge disposition provides information on where the patient goes after treatment in the ED, and includes home, transfer to other facilities, left without being seen, and died. The relevant categories for describing transitions between ED and acute care were *admitted to the acute care hospital in which the ED was located* and *transferred for admission to an acute care hospital in a different facility*.

Electronic hospital discharge abstracts and population registry files were also used to conduct the research. A hospital discharge abstract is completed when a patient is discharged from an

acute care facility. Hospital discharge data are available for all inpatient hospitalizations in the province. For this study, the relevant components of the hospital discharge abstracts were the mode of entry to the hospital and the admission dates.

The population registry file contains demographic information, such as date of birth and residence location. It also captures dates of health insurance coverage.

All healthcare databases can be linked via a unique, anonymized personal health identification number. 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.

2.4.2 Study cohort

The study adopted a population-based retrospective cohort design, which comprised all provincially insured individuals who had at least one visit to any of the five EDs between 2006/07 and 2011/12 fiscal years (a fiscal year extends from April 1 to March 31).

2.4.3 Study measures

Patient characteristics selected for investigation were based on the Andersen healthcare utilization model,[12] and included predisposing factors of age group (0 – 19, 20 – 39, 40 – 59, 60 – 79, 80+) and sex (male, female), and enabling factors of residence location (urban, rural) and health region affiliation (Saskatoon, Regina Qu'Appelle). Urban residents were those whose postal codes were in a census metropolitan or agglomeration area (i.e., 10,000+ population). All variables were measured as of the date of ED visit.

2.4.4 Statistical analysis

To achieve the first study objective, the ED visit discharge dispositions were described using frequencies and percentages. To achieve the second objective, agreement between the ED data and hospital discharge abstracts database in capturing patient transitions between the two services was estimated using the kappa statistic (κ);[13] 95% confidence intervals (CIs) were also computed. The κ statistic has been used in previous studies to measure agreement between administrative health databases.[14, 15] We selected κ statistic as the measure of agreement because our variable of interest is binary (i.e., whether or not patient transitions from emergency care to acute care were recorded in both the ED and hospital databases).[16] The value of κ is 1 when perfect agreement exists between two sources, 0 when agreement is equal to that expected assuming independence, and negative when agreement is less than expected by chance.[17] The interpretation of agreement adopted here is: poor ($\kappa < 0.20$), fair ($\kappa = 0.20$ to 0.39), moderate ($\kappa = 0.40$ to 0.59), good ($\kappa = 0.60$ to 0.79), and very good ($\kappa = 0.80$ to 1.00).[17] When measuring agreement, we linked patient transitions from ED to acute care by matching hospital admissions which occurred on the same day as the ED discharge; as well as allowing up to 3 days between ED discharge date and hospital admission date in a sensitivity analysis.[9] Patient transitions between ED and acute care were assessed where the ED and the acute care hospital were in the same facility, as well as where the ED and the admitting acute care hospital were not in the same facility. Analyses were stratified by fiscal year and ED location.

Percentages (95% CIs) were used to describe the differences between cohort members whose transition information was and was not missing. A patient's transition information was considered missing if the ED discharge disposition indicated that the patient was admitted to an

acute care hospital either in the same facility as the ED or in a different facility but the patient's admission information was not recorded in the hospital discharge abstracts.

2.5 Results

A total of 383,860 patients had at least one visit to an ED in the province's teaching hospitals between 2006/07 and 2011/12 fiscal years. Of this number, 12,380 (3.2%) patients did not have provincial insurance coverage (e.g., were not residents of the province) and were therefore excluded. Thus, the study cohort was comprised of 371,480 patients (96.8%) with a total of 1,062,861 ED visits over the study period.

Table 2-1 summarizes the discharge dispositions for all ED visits by the study cohort members. Three-quarters of ED visits resulted in a discharge to the patient's home, while 16.1% resulted in admission to the acute care hospital in which the ED was located, and 1.6% resulted in a transfer for admission to an acute care in a different facility.

The contingency table statistics used to calculate the overall agreement between the ED data and hospital records are summarized in Table 2-2. Of the 170,584 visits recorded in the ED data with a discharge disposition of *admitted to the acute care hospital in which the ED was located* (Table 2-1), 143,633 of these visits were found in the hospital discharge abstracts with the same admission date as the ED discharge date (cell A in the upper half of Table 2-2). Similarly, of the 16,951 visits recorded in the ED data with a discharge disposition of *transferred for admission to an acute care hospital in a different facility* (Table 2-1), only 6,633 of these visits were found in the hospital discharge abstracts with the same admission date as the ED discharge date (cell A in the lower half of Table 2-2).

Table 2-3 provides the results for the assessment of patient transitions from ED to acute care settings, stratified by fiscal year. For the case where patients were admitted to acute care in the

same facility in which the ED was located, the overall agreement between the ED and hospital records was good ($\kappa = 0.77$, 95% CI = 0.77, 0.77). The agreement between the two data sources was lowest ($\kappa = 0.55$, 95% CI = 0.54, 0.55) in 2006/07 fiscal year and increased steadily to the highest value ($\kappa = 0.94$, 95% CI = 0.94, 0.94) in 2011/12 fiscal year. For ED discharges to acute care in a different facility, the overall agreement was only fair ($\kappa = 0.36$, 95% CI = 0.35, 0.36); agreement varied by fiscal year, reaching its peak ($\kappa = 0.40$, 95% CI = 0.39, 0.42) in 2009/10.

The level of agreement between ED and hospital records also varied across facilities (Figure 2-1); with κ estimates ranging from good (i.e., $\kappa = 0.68$ for ED #5) to very good (i.e., $\kappa = 0.87$ for ED #4).

In the sensitivity analysis, when a lag of up to three days between ED discharge and hospital admission dates was allowed, agreement increased slightly both for admissions to acute care in the same facility in which the ED was located (from $\kappa = 0.77$ to $\kappa = 0.80$), and for transfers to an acute care hospital located in a different facility (from $\kappa = 0.36$ to $\kappa = 0.43$).

Table 2-4 compares the characteristics of cohort members whose transition information was and was not missing. Those whose information was missing from the ED to the acute care in which the ED was located had a total of 3,601 ED visits over the study period; and they were younger (62.8% vs 38.6% in the 20 to 59 years age group) and were more likely to be male (59.4% vs 50.6%) than those whose information was not missing. No major differences were observed between the two groups in terms of residence location and health region affiliation. Similarly, patients whose transition information was missing when they were transferred from the ED to an acute care in a different facility had a total of 1,218 ED visits over the study period; and were younger (64.1% vs 43.3% in the 20 to 59 years age group), and with more representation of males (61.9% vs 49.7%) than those whose information was not missing.

2.6 Discussion

The vast majority of patients who visited EDs in Saskatchewan's teaching hospitals during the study period did not transition between healthcare services; they were discharged home. Although the accuracy of the ED discharge to home information could not be ascertained within our study, our estimate of 74.9% of ED visits being discharged home is consistent with another Canadian study which found 73.8% of ED visits being discharged home during a similar time period.[18] Similarly, our results for discharges to acute care were quite similar to those recorded in the National Ambulatory Care Reporting System (NACRS) for Canada,[19] especially for visits transferred for admission to an acute care in a different facility (average of 1.6% over our study period versus an average of 1.1% in NACRS over the same period). We found higher rates for admissions to the acute care hospital in which the ED was located than the NACRS data (16.1% vs 9.4%), and this might have occurred because ED visits rates reported in NACRS exclude scheduled visits.

The overall agreement between the two data sources in capturing transitions from the ED to the acute care hospital in which the ED was located was good; a high proportion of all ED visits recorded as being discharged to an acute care hospital located within the same facility were identified in the hospital records as being admitted on the same date. We also noted a steady improvement in agreement between the two data sources over time. This may be an indication of improvements in data quality over time.

Among ED visits transferred to an acute care hospital located in a different facility, agreement between the two data sources was only fair. Thus, a high proportion of all ED visits recorded as being discharged to an acute care hospital located in a different facility were not

identified in the hospital records as being admitted to the facility on the same date. Also, the measure of agreement did not improve consistently over time for this category of ED discharges. The highest agreement was observed in 2009/10 fiscal year, and may be associated with a major outbreak of H1N1 flu virus in Canada.[20] A high level of inter-facility collaboration between the EDs and acute care hospitals likely occurred during this period.

Patient transition information between the ED and acute care were more likely to be missing at some ED locations than others. ED numbers 1 and 2 are located in the Regina Qu'Appelle health region whilst the other three EDs are located in Saskatoon health region. Although the data used in this study has been anonymized and harmonized by Saskatchewan eHealth, two separate information systems were used by the health regions for collecting the ED data. However, there was no clear influence of the difference in the information system on the results; since the EDs with the highest and lowest agreement were both located in the Saskatoon health region. Although all the five EDs are located in teaching hospitals, ED number 3 is located in the biggest teaching hospital in the province and serves as the main trauma center for the province.

Allowing for up to three days between the ED discharge date and hospital admission date in a sensitivity analysis resulted in some improvement in agreement between the two data sources. The possibility of lags between ED discharge and hospital inpatient admission for the same episode of care was reported by a previous study.[9] Nonetheless, the overall agreement between the two data sources for ED visits discharged to an acute care hospital in a different facility remained low.

One potential reason for the lower agreement between the ED data and hospital records for ED discharges to an acute care in a different facility may be due to a lack of clarity of discharge dispositions labelled as *transferred for admission to acute care hospital in a different facility*. It

is likely that some of these institutions labelled as acute care hospitals may not be contributing data to the hospital discharge abstracts database to which the ED data was linked. Another reason may be that patients did not reach the hospital to which they were being transferred either due to death in the case of severely ill patients or a decision to seek care elsewhere. A third potential reason is data entry errors. Peabody et al.[21] found inaccuracies in the coding of primary and secondary diagnosis in administrative data; similar coding errors might exist in the ED data. Further investigations, including chart reviews, could explore the cause(s) of the lack of agreement.

We were unable to evaluate the accuracy of the ED discharges to other types of care setting (i.e., non-acute care) because we did not have access to any linkable data sources containing these healthcare services. This study could be expanded to include healthcare utilizations prior to the arrival in ED such as ambulance services. By linking ambulance services, ED, and hospital data, future studies would be able to assess the accuracy of transition information captured in these databases for care trajectories related to the entire emergency care to acute care journey.

2.7 Conclusion

In summary, we used a population-based administrative health data to describe discharge dispositions from EDs and then evaluate patient transition between the ED and acute care settings. The majority of patients who attended EDs during the study period did not transition to another healthcare setting. For those who transitioned to acute care, agreement between ED and hospital records depended on whether the two services were provided in the same facility. The agreement between the two data sources for capturing transitions from ED to the acute care hospital in which the ED was located was good, whilst the agreement for transitions which occurred between the ED and an acute care hospital located in a different facility was only fair.

Compared to hospital records, electronic ED records appear not to contain accurate or complete information about transitions to acute care services in different facilities. Therefore, studies of patient care trajectories that describe transitions from the ED to acute care should not rely on the discharge disposition field recorded in the ED data, but rather be conducted by linking patient-specific records across the two care settings.

2.8 References

1. Le Meur N, Gao F, Bayat S. Mining care trajectories using health administrative information systems: the use of state sequence analysis to assess disparities in prenatal care consumption. *BMC Health Serv Res*. 2015;15(1).

2. Peltola M, Juntunen M, Häkkinen U, et al. A methodological approach for register-based evaluation of cost and outcomes in health care. *Ann Med*. 2011;43 Suppl 1(January):S4–S13.

3. Adeyemi S, Demir E, Chaussalet T. Towards an evidence-based decision making healthcare system management: Modelling patient pathways to improve clinical outcomes. *Decis Support Syst*. 2013;55(1):117–125.

4. Elliss-Brookes L, McPhail S, Ives A, et al. Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. *Br J Cancer*. 2012;107(8):1220–6.

5. Jay N, Nuemi G, Gadreau M, Quantin C. A data mining approach for grouping and analyzing trajectories of care using claim data: the example of breast cancer. *BMC Med Inform Decis Mak*. 2013;13:130.

6. Roos LL, Gupta S, Soodeen R-A, Jebamani L. Data quality in an information-rich environment: Canada as an example. *Can J Aging / La Rev Can du Vieil*. 2005;24(S1):153–170.

7. Jacke CO, Kalder M, Wagner U, Albert U-S. Valid comparisons and decisions based on clinical registers and population based cohort studies: assessing the accuracy, completeness and epidemiological relevance of a breast cancer query database. *BMC Res Notes*. 2012;5(1):700.

8. Crilly J, O'Dwyer J, O'Dwyer M. Linking ambulance, emergency department and hospital admissions data: understanding the emergency journey. *Med J Aust*. 2011;194(4).

9. Wong A, Kozan E, Sinnott M, Spencer L, Eley R. Tracking the patient journey by combining multiple hospital database systems. *Aust Health Rev*. 2014;38(3):332–6.

10. McHale P, Wood S, Hughes K, et al. Who uses emergency departments inappropriately and when - a national cross-sectional study using a monitoring data system. *BMC Med*. 2013;11(258).

11. Statistics Canada, Population and dwelling counts for Canada, provinces and territories, 2011 and 2006 censuses, Ottawa, ON. Available at: http://www12.statcan.gc.ca/census-recensement/2011/dp-pd/hlt-fst/pd-pl/Table-Tableau.cfm?LANG=Eng&T=101&S=50&O=A (2012). Accessed September 14, 2014

12. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter?

J Health Soc Behav. 1995;36(1):1–10.

13. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37–46.

14. Chen G, Faris P, Hemmelgarn B, Walker RL, Quan H. Measuring agreement of administrative data with chart data using prevalence unadjusted and adjusted kappa. *BMC Med Res Methodol*. 2009;9(1).

15. Turner D, Hildebrand KJ, Fradette K, S L. Same question, different data source, different answers? Data source agreement for surgical procedures on women with breast cancer. *Heal Policy*. 2007;3(1):46–54.

16. Hallgren KA. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutor Quant Methods Psychol*. 2012;8(1):23–34.

17. Warrens MJ. Cohen's kappa is a weighted average. *Stat Methodol*. 2011;8(6):473–484.

18. Doupe M, Kozyrskyj A, Soodeen R-A, et al. An initial analysis of emergency departments and urgent care in Winnipeg, Winnipeg, MB. Available at: http://mchp-appserv.cpe.umanitoba.ca/reference/EDFull.pdf (2008). Accessed October 13, 2014.

19. Canadian Institute for Health Information, The impact of the H1N1 pandemic on Canadian hospitals, Ottawa, ON. Available at: https://secure.cihi.ca/free_products/H1N1_AIB_final_EN.pdf (2010). Accessed August 20, 2014.

20. Rolland E. Comparing Telehealth and NACRS Influenza Activity to Fluwatch Surveillance Using EARS-X. *Adv Dis Surveill*. 2007;4:189.

21. Peabody J, Luck J, Jain S. Assessing the accuracy of administrative data in health information systems. *Med Care*. 2004;42(11):1066–1072.

Table 2-1: Emergency department discharge dispositions for the study cohort, April 1, 2006 to March 31, 2012

Discharge disposition	***n* (%)**
Home	795,823 (74.9)
Admitted to the acute care hospital in which the ED was located	170,584 (16.1)
Unplanned discharge (i.e., left before being seen, left against medical advice after being seen by a doctor, signed out)	61,274 (5.7)
Transferred for admission to an acute care hospital in a different facility	16,951 (1.6)
Institutional place of residence (e.g., long-term care, jail)	10,129 (1.0)
Transfer within same facility (i.e., day surgery, out-patient care)	3,482 (0.3)
Transfer to external non-acute care facility	3,407 (0.3)
Died	1,211 (0.1)
Total	1,062,861 (100.0)

Notes: ED = emergency department

Table 2-2: Contingency table statistics used to calculate overall agreement between emergency department and hospital records

ED visits admitted to the acute care hospital in which the ED was located			
		Hospital records	
		Admitted to attached hospital on same day as discharge from ED	Not admitted to attached hospital on same day as discharge from ED
ED Records	Admitted to acute care in the attached hospital	A: True positives *n* = 143,633	B: False positives *n* = 26,951
	Not admitted to acute care in the attached hospital	C: False negatives *n* = 41,243	D: True negatives *n* = 851,034
ED visits transferred to an acute care hospital in a different facility			
		Hospital records	
		Admitted to a different hospital on same day as discharge from ED	Not admitted to a different hospital on same day as discharge from ED
ED Records	Transferred to an acute care in a different facility	A: True positives *n* = 6,633	B: False positives *n* = 10,318
	Not transferred to an acute care in a different facility	C: False negatives *n* = 12,495	D: True negatives *n* = 1,033,415

Notes: ED = emergency department

Table 2-3: Agreement between emergency department (ED) and hospital records for capturing patient transition from ED to acute care, stratified by fiscal year

	Transition from ED to the acute care hospital in which the ED was located	Transition from ED to acute care in a different facility
Fiscal Year *	κ (95% CI)	κ (95% CI)
Overall	0.77 (0.77, 0.77)	0.36 (0.35, 0.36)
2006/07	0.55 (0.54, 0.55)	0.33 (0.31, 0.34)
2007/08	0.67 (0.67, 0.68)	0.34 (0.33, 0.36)
2008/09	0.73 (0.73, 0.74)	0.36 (0.34, 0.38)
2009/10	0.83 (0.82, 0.83)	0.40 (0.39, 0.42)
2010/11	0.86 (0.86, 0.86)	0.34 (0.33, 0.35)
2011/12	0.94 (0.94, 0.94)	0.36 (0.34, 0.38)

Notes: CI = confidence interval

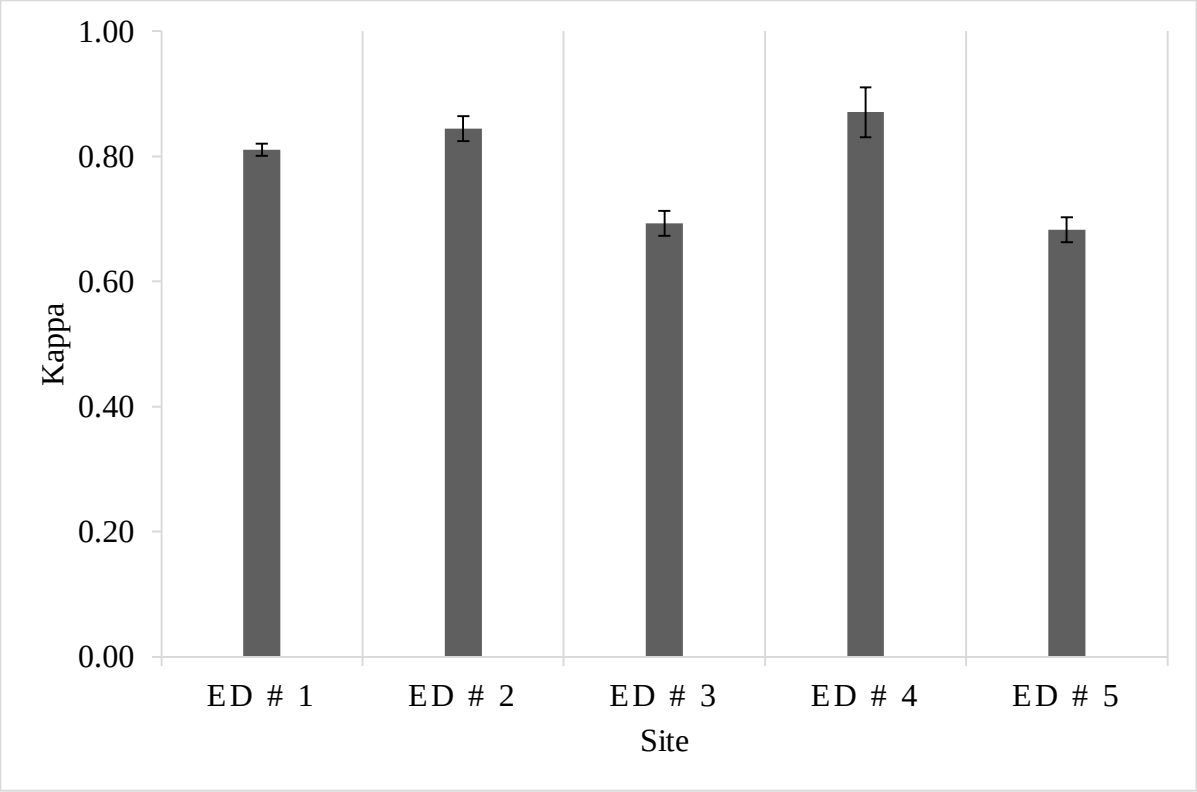
* A fiscal year extends from April 1 to March 31

Table 2-4: Comparison of study cohort by status of transition information from emergency department (ED) to acute care

	Transition from ED to the acute care hospital in which the ED was located		Transition from ED to acute care in a different facility	
	Cohort without missing transition information from ED to the hospital in which the ED was located (*N*=36,254)	Cohort with missing transition information from ED to the hospital in which the ED was located (*N*=1,559)	Cohort without missing transition information from ED to a hospital located in a different facility (*N*=3,742)	Cohort with missing transition information from ED to a hospital located in a different facility (*N*=239)
	% (95% CI)			
Age group				
0 - 19	17.0 (16.7 – 17.3)	16.2 (14.5 – 18.1)	19.5 (18.3 – 20.8)	23.4 (18.5 – 29.2)
20 - 39	16.7 (16.5 – 16.9)	36.0 (33.6 – 38.4)	21.5 (20.3 – 22.9)	42.3 (34.6 – 46.9)
40 - 59	21.9 (21.6 – 22.1)	26.8 (24.6 – 29.0)	21.8 (20.5 – 23.2)	21.8 (17.0 – 27.4)
60 - 79	26.9 (26.6 – 27.1)	14.2 (12.5 – 16.0)	22.4 (21.0 – 23.7)	4.5 (2.6 – 8.1)
80+	17.5 (17.3 – 17.8)	6.8 (5.1 – 7.9)	14.8 (13.7 – 16.0)	8.0 (6.5 – 14.0)
Sex				
Female	49.4 (49.0 – 49.8)	40.6 (37.8 – 42.7)	50.3 (48.7 – 51.9)	38.1 (32.2 – 44.4)
Male	50.6 (50.3 – 50.9)	59.4 (56.9 – 61.8)	49.7 (48.1 – 51.3)	61.9 (54.4 – 66.7)
Residence location				
Urban	99.3 (99.3 – 99.4)	99.5 (98.9 – 99.7)	99.8 (99.6 – 99.9)	93.7 (89.9 – 96.2)
Rural	0.7 (0.7 – 0.8)	0.5 (0.3 – 1.0)	0.2 (0.1 – 0.4)	6.3 (3.8 – 10.1)
Health region				
Regina Qu'Appelle	50.5 (50.2 – 50.8)	50.1 (48.2 – 53.2)	48.0 (46.4 – 49.6)	54.4 (48.1 – 60.6)
Saskatoon	49.5 (49.1 – 49.9)	48.9 (46.5 – 51.4)	52.0 (50.4 – 53.6)	45.6 (39.4 – 51.9)

Notes: ED = emergency department

Figure 2-1: Agreement between emergency department (ED) and hospital records for capturing patient transition from ED to acute care, stratified by ED site; error bars represent 95% confidence intervals



CHAPTER 3: IDENTIFYING DISTINCT HEALTHCARE PATHWAYS DURING EPISODES OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE EXACERBATIONS

3.1 Chapter overview

This manuscript defines patient care pathways during episodes of care for patients with chronic obstructive pulmonary disease (COPD) exacerbations. Understanding healthcare pathways is important because they affect outcomes such as cost and mortality. Following our findings in the previous study, episodes of care were constructed using Saskatchewan's population-based administrative health data, including the emergency department (ED) data, hospital discharge abstracts, physician billing claims, and outpatient dispensation drugs data. The paper focuses on using a model-based approach to define care pathways and describe patient characteristics associated with these pathways.

Publication details:

Kuwornu JP, Lix LM, Quail JM, Forget E, Muthukumarana S, Wang XE, Osman M, Teare GF. Identifying distinct healthcare pathways during episodes of chronic obstructive pulmonary disease exacerbations. *Medicine.* 2016; 95(9):e2888

3.2 Abstract

Objective: The objective of this study was to identify and describe healthcare pathways during episodes of COPD exacerbations.

Methods: Linked administrative databases from Saskatchewan, Canada were used to identify a cohort of newly-diagnosed COPD patients and their episodes of healthcare use for disease exacerbations. Latent class analysis (LCA) was used to classify the cohort into homogeneous pathways using indicators of respiratory-related hospitalizations, emergency department (ED) visits, general and specialist physician visits, and outpatient prescription drug dispensations. Multinomial logistic regression models tested patient demographic and disease characteristics associated with pathway group membership. The most frequent healthcare contact sequences in each pathway were described. Tests of mean costs across groups were conducted using a model-based approach with χ^2 statistics.

Results: LCA identified three distinct pathways for patients with hospital- (*n*=963) and ED-initiated (*n*=364) episodes. For the former, pathway group 1 members followed complex pathways in which multiple healthcare services were repeatedly used and incurred substantially higher costs than patients in the other pathway groups. For ED patients, pathway group 1 members also had higher costs than other groups. Pathway groups differed with respect to patient demographic and disease characteristics. A minority of patients were discharged from ED or hospital but did not have any follow-up care during the remainder of their episode.

Conclusion: Patients who followed complex pathways could benefit from case management interventions in order to improve quality of care. The minority of patients whose pathways were not consistent with recommended follow-up care should be further investigated to fully align COPD treatment in the province with recommended care practices.

3.3 Introduction

Chronic obstructive pulmonary disease (COPD) is a leading cause of morbidity and is annually responsible for up to three million deaths worldwide.[1] COPD patients are heavy users of healthcare services and are responsible for up to one-quarter of all hospitalizations and ED visits, and more than one-fifth of ambulatory visits.[2] COPD patients may experience multiple acute exacerbations of symptoms annually; treatment often requires an ED visit or hospitalization.[3,4] Patients may also receive follow-up care from their primary care provider or a specialist physician, and might require additional medications.[5] Accordingly, an exacerbation episode may require contact with a number of different healthcare providers and services and the care pathway, the patient's journey through the healthcare system during COPD exacerbations, will not be the same for every patient.[6]

Studies have investigated the economic burden of COPD exacerbations,[2,4] and the risk factors associated with them,[7,8] but little is known about the care pathways that patients follow during episodes of exacerbation, and how these influence health outcomes such as cost. Understanding care pathways during exacerbations could help care providers anticipate the needs of patients.[2] Also, care pathway analysis can guide quality of care initiatives by identifying patients who may benefit the most from case management interventions.[9]

Care pathways can be investigated using an episode-of-care data system; a data system which collects all healthcare services related to the treatment of a specific health condition.[10] Health services utilization within a care pathway can be characterized by duration,[11] intensity (i.e., frequency of contacts),[12] and sequence of contacts (i.e., trajectories).[6] Latent class analysis (LCA) has previously been used to characterize healthcare utilization profiles based on intensity

of use.[13,14] However, the method could also be used to identify pathways during episodes of care for patients with complex health conditions.

The objective of this study was to identify and describe healthcare pathways during episodes of COPD exacerbations, focusing on hospital- and ED-initiated episodes. These exacerbations, which are moderate to severe in nature,[15] routinely incur the highest healthcare costs.[16]

3.4 Methods

3.4.1 Data sources

We used population-based 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. The province maintains multiple administrative health databases in electronic format, which can be anonymously linked via a unique personal health number.[17]

Episodes of care for COPD exacerbations were constructed using databases that capture primary, emergency, and acute care services for provincial health insurance beneficiaries; these include 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 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, with a single diagnosis 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, costs, 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.

3.4.2 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 twelve health regions in Saskatchewan and the only ones for which ED data were available. Both SHR and RQHR are the only health regions which 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: (1) one or more hospitalizations with a diagnosis of COPD in any diagnosis field, or (2) one or more physician visits with a diagnosis of COPD.[18] This case definition has a sensitivity of 85.0% and a specificity of 78.4% when compared with clinical evaluation by a physician.[2] The index date of COPD diagnosis was the earliest hospitalization or physician visit date for COPD. Cases were identified from hospital discharge abstracts using the following ICD-10-CA codes: J41, J42, J43

or J44; whilst cases in the physician billing claims were identified with ICD-9 codes 491, 492 or 496.

We used a look-back period of five years from the index date to determine whether or not a patient had a prior COPD diagnosis. We selected this duration of time based on previous research,[19] 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.

3.4.3 Defining episodes of care for COPD exacerbations

All episodes of COPD exacerbations following the index diagnosis were defined using the healthcare services which initiated, continued, and ended them. We identified hospital-initiated, ED-initiated, general practitioner (GP) visit-initiated, and specialist visit-initiated episodes of care based on a method developed by the Canadian Institute for Health Information (CIHI), a national non-profit organization that provides standardized methods and data sources for health services research, for ascertaining COPD exacerbations.[20] Hospital-initiated or ED-initiated episodes had: (1) a COPD diagnosis in the most responsible diagnosis field, or (2) 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 algorithm was used to define the

episode initiating service in cases where multiple services were used on the date the episode started. When physician services and hospital (or ED) services were received on the same date, and the billing claims indicated that the physician provided the service in a hospital (or ED), the episode initiating service was taken as hospitalization (or ED visit). Otherwise, the episode initiating service was taken as physician visit. When ED visit and hospitalization occurred on the same date, the ED visit was the initiating service. Since the study focused on moderate to severe exacerbations,[15] only the ED- and hospital-initiated episodes were included.

An episode continued if there were respiratory-related hospitalizations, ED, GP, 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 captured. These drugs included those that were listed to define physician visit-initiated episodes described earlier, as well as those that are used for chronic management of COPD and other respiratory-related conditions (e.g., long-acting beta-agonists and long-acting bronchodilator combined with anti-inflammatory medications). Inclusion of all respiratory-related drugs increases the possibility of capturing all relevant drug utilizations during 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 1 year from their index date until March 31, 2012 or death, whichever occurred first. All episodes that were ongoing at the end of the observation period were excluded. This study was limited to index episodes, that is, all first episodes of care following the COPD diagnosis date.

3.4.4 Study measures

3.4.4.1 Indicators of care pathways

The indicators of care pathways were based on all respiratory-related services received following the episode-initiating service. Specifically, binary measures of hospitalization, ED visit, GP visit, specialist visit, and outpatient prescription drug dispensation were used to define care pathways.

3.4.4.2 Episode of care cost and duration

The episode cost was the sum of all respiratory-related costs incurred between the episode start and end dates. Hospital costs were estimated based on a standard methodology developed by CIHI for all Canadian provinces.[21] For the ED cost component, total annual expenditures were obtained from the Ministry of Health and total annual numbers of visits were 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. Prescription drug costs were based on the price of the active substance plus a dispensing fee. Episode costs were adjusted for inflation using the health and personal care components of the Saskatchewan consumer price indices,[22] and expressed in 2011/12 constant dollars.

Duration was measured in days; starting from the first day of the episode initiating service to the date the last service was received. Episode intensity was measured by the number of times the patient contacted care providers during the episode.

3.4.4.3 Patient and disease characteristics

The patient and disease characteristics included in the analysis were based on the Andersen healthcare utilization model.[23] Andersen proposed that an individual's healthcare use is influenced by 3 broad groups of factors, namely predisposing, enabling, and need. The

predisposing factors included in this study were sex and age. 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 factors were comorbid conditions, which were measured using the Charlson index,[24] as well as by the three most prevalent comorbid conditions in the study cohort; including congestive heart failure (i.e., yes or no), diabetes (i.e., yes or no) and hypertension (i.e., yes or no). These comorbid conditions were defined using ICD-9 and ICD-10-CA codes, [25] and were based on diagnoses in the hospital discharge abstract and the physician billing claims data. A higher score of the Charlson index indicates greater comorbidity, and we categorized the index as 0, 1, or ≥ 2. We also included the fiscal year of COPD diagnosis (i.e., 2007/08, 2008/09, 2009/10, or 2010/11). All variables were defined as of the index date of COPD diagnosis except for the comorbid conditions, which were defined using data from the year prior to the index date.

3.4.5 Statistical Analysis

LCA was used to classify patients into distinct care pathway groups based on healthcare services in the index episode. LCA assumes that each individual in the study belongs to one of a set of mutually exclusive and exhaustive classes.[25] We fit models to the data having between two and seven classes. The Bayesian-Schwarz Information Criterion (BIC),[26] which is a penalized measure of the likelihood function, as well as entropy, a measure of the quality of class separation, were used to guide the selection of the optimal number of classes.[27] Entropy values range from 0 to 1, with values greater or equal to 0.8 often recommended as indicative of good class separation; while a smaller BIC for a particular model suggests that it is preferable on the basis of the trade-off between fit and parsimony. Final model selection was based on these indices, as well as the usefulness and interpretability of results.[13] The usefulness of results was

assessed based on the number of individuals in each latent class, with classes of less than 10 individuals not being considered.

Subsequent to choosing the final model, the hospital- and ED-initiated pathways were described. Multinomial logistic regression models were used to test patient demographic and disease characteristics associated with pathway groups. In order to avoid multicollinearity between the Charlson comorbidity index and the three most prevalent comorbid conditions in the study cohort, two sets of multinomial logistic regression models were conducted. The first model included patient demographics (i.e., age, gender, and residence location), Charlson comorbidity index and COPD diagnosis fiscal year as covariates. The second model included the three most prevalent comorbid conditions, and adjusted for patient demographics and COPD diagnosis fiscal year. Only the adjusted odds ratios (OR) (with the 95% confidence intervals) of the three most prevalent comorbid conditions were reported from the second multinomial logistic regression model. Further, we plotted the means (with the 95% confidence intervals) of episode duration and healthcare contacts by pathway groups. The plots helped to examine the variations in these episode characteristics by pathway groups. We also described the sequence in which patients contacted care providers, focusing on the most frequent patterns in the study cohort. We calculated the percentage of deaths within the most frequent sequence in each care pathway. Finally, we used χ^2 statistics to test for differences in outcomes (i.e., costs and mortality) between care pathways using a model-based method.[28] The method empirically derives and summarizes the class-dependent density functions of outcomes with categorical, continuous, or count distributions.

SAS® version 9.3 (SAS Institute Inc., Cary, NC, USA) was used for the descriptive analyses and data manipulation. Mplus® version 7.2 (Muthén & Muthén, Los Angeles, CA, USA) was used to perform the LCA.

3.5 Results

3.5.1 Cohort selection

A total of 12,543 COPD patients 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%]) the remaining 7,099 individuals were eligible for study inclusion. During follow-up, 2,659 individuals experienced at least one COPD episode of care, which resulted in a total of 5,348 episodes. We excluded a total of 156 (2.9%) episodes because they were still ongoing at the end of study period. Further, since the study focused on moderate to severe exacerbations, we excluded all those whose episodes were initiated by a physician visit (1,332; 50.1%). The final cohort (n = 1,327) was comprised of all individuals whose index episode was initiated by a hospitalization (n=963; 72.6%) or an ED visit (n=364; 27.4%). Separate analyses were conducted for the two groups.

3.5.2 Latent class model selection

The BIC value was smallest for the 3-class model in both the hospital- and ED-initiated groups. Entropy was highest (0.91) for the 3-class model among individuals with hospital-initiated episodes, but was highest (0.78) for the 5-class model among the individuals with ED-initiated episodes. However, the 5-class model had at least 1 class comprised of less than 10 individuals, rendering it not useful for further analysis. Given that the improvement in entropy of

the 5-class model over the 3-class model among the ED-initiated episodes was only marginal (0.04), we selected the 3-class model for both groups.

3.5.3 Characteristics of care pathways

Table 3-1 summarizes the results of the care pathways of individuals derived from the LCA models. Among individuals with a hospital-initiated episode, Pathway 1 (7.2%) was comprised of those who had a high probability of ED visit(s), GP visit(s), and specialist visit(s) after their initial hospital discharge. Also, they had a high probability of having another hospitalization during the same episode. Pathway 2 (35.6%) was comprised of individuals who had a high probability of follow-up visit(s) to GPs and outpatient drug dispensations after their episode-initiating hospitalization. The majority of individuals were in Pathway 3 (57.2%); they had a lower probability of having any follow-up contacts after their episode-initiating hospital discharge.

Similarly, among individuals with an ED-initiated episode (Table 3-1), those in Pathway 1 (27.2%) had a higher probability of hospitalization, GP and specialist visits, and outpatient drug dispensations following their initial ED visit. The majority of individuals, who belonged to Pathway 2 (48.6%), had a lower probability of having any follow-up utilization after their initial ED discharge. Pathway 3 (24.2%) was comprised of individuals who had a higher probability of follow-up visit(s) to GPs and outpatient drug dispensations after their episode-initiating ED visit.

3.5.4 Patient characteristics and care pathways

Table 3-2 shows the results of the multinomial logistic regression models, which examined the association between pathway membership, and patient demographic and comorbidity characteristics. Among individuals with a hospital-initiated episode, those in the 55 – 74 years age group were more likely to be in Pathway 1 compared to Pathway 3 (OR = 2.08, 95% CI 1.11

to 3.90). Individuals in Pathway 2 were more likely to be female compared to individuals in Pathway 3 (OR = 1.58, 95% CI 1.20 – 2.09). Also, individuals with a Charlson comorbidity index score of 1 (OR = 0.66, 95% CI 0.47 – 0.95) or ≥ 2 (OR = 0.46, 95% CI 0.33 – 1.59) were less likely to be in Pathway 2 compared to Pathway 3. Among individuals with an ED-initiated episode, those in Pathway 3 had a significantly lower odds of being female (OR = 0.55, 95% CI 0.30 – 0.99), and having Charlson comorbidity score ≥ 2 (OR = 0.32, 95% CI 0.11 – 0.93) compared to individuals in Pathway 2.

Among individuals with a hospital-initiated episode, those in Pathway 1 had a significantly higher odds of congestive heart failure (OR = 1.31, 95% CI 1.04 – 1.65) compared to those in Pathway 3 (Table 3-2). Also, individuals in Pathway 2 had a lower odds of hypertension (OR = 0.59, 95% CI 0.37 – 0.92) compared to those in Pathway 3. Among individuals with an ED-initiated episode, those in care Pathway 1 had significantly higher odds of diabetes (OR = 1.33, 95% CI 1.02 – 1.74) and hypertension (OR = 1.44, 95% CI 1.04 – 2.00) compared to those in Pathway 2.

3.5.5 Episode duration, care contacts, and care sequence by care pathways

Figure 3-1 describes the mean (95% confidence intervals) duration of episodes (a) and mean healthcare contacts (b) by care pathways. Among individuals with a hospital-initiated episode, those assigned to Pathway 1 had the longest episode duration as well as the highest number of healthcare contacts. Similarly, among individuals with an ED-initiated episode, those in Pathway 1 had the longest episode duration and highest number of healthcare contacts. However, although individuals in Pathway 3 had longer duration than those in Pathway 2, the number of healthcare contacts in the two pathways was similar.

Among individuals with a hospital-initiated episode, almost one quarter of those in Pathway 1 had a specialist visit following hospital discharge, followed by ED and specialist visits (Figure 3-2a). Among individuals in Pathway 3, 41.2% did not have follow-up care after hospital discharge. Further analysis indicates that only 20.3% of the individuals who followed this sequence died during the episode (Figure 3-3a). For individuals with an ED-initiated episode, 18.2 % of those in Pathway 1 had a hospitalization following the ED discharge followed by specialist visit; and close to one fifth of those in Pathway 3 did not have any follow-up care although they were discharged from the ED and remained alive throughout the episode (Figure 3-2b). Of the 57 people in Pathway 2 who did not have any follow-up care (Figure 3-2b), only 14.0% of them died during the episode (Figure 3-3b).

3.5.6 Comparison of healthcare outcomes between care pathways

Table 3-3 shows the results for the tests of differences in healthcare outcomes between care pathways for both hospital- and ED-initiated episodes. Among individuals with a hospital-initiated episode, those in Pathway 1 had a higher mean cost ($29,480; SD = 54,907) than those in Pathway 2 ($9,817; SD = 21,865) and Pathway 3 ($10,619; SD = 21,042) (*P* value = 0.049). A higher percentage (18.0%) of individuals in Pathway 3 died during their index episode compared to individuals in Pathway 1 (15.9%) and Pathway 2 (6.1%) (*P* value < 0.001).

Among individuals with an ED-initiated episode, those in Pathway 1 had a higher mean cost ($9,204; SD = $10,019) than those in Pathway 2 ($3,493; SD = $5,279) and Pathway 3 ($943; SD = $2,992) (*P* value < 0.001). A higher percentage (10.7%) of individuals in Pathway 2 died during their index episode compared to individuals in Pathway 1 (10.1%) and Pathway 3 (0.0%) (*P* value = 0.006).

3.6 Discussion

In this study, we applied a model-based approach to characterize healthcare pathways for COPD patients focusing on hospital- and ED-initiated episodes of care. Understanding the pathways that these patients follow after they are discharged from the acute care setting is important for identifying opportunities for redesigning healthcare delivery to improve quality of care and lower cost.[6] This population-based study uniquely presents the relationship between pathways and healthcare outcomes.

Three distinct care pathways were identified among individuals with a hospital-initiated episode. Individuals in these pathways differed in terms of the health services they used, as well as the sequence of contacts with care providers. A small group of individuals (7.2%) made repeated use of a wide variety of healthcare services during their episode, which resulted in their cost being about three times higher than the pathway group with the lowest cost. Similarly, three distinct pathways were identified among the individuals with an ED-initiated episode. More than one-quarter of individuals whose episode were initiated by an ED visit used multiple services, and their healthcare costs were approximately 10 times higher than the pathway group with the lowest cost.

Analysis of the duration, intensity, and sequence of healthcare contacts provides a clear description of the care processes experienced by individual patients during COPD exacerbation. Information on the sequence in which patients contacted care providers may point at deviations from recommended care practices. For example, it is recommended that patients treated for COPD exacerbations should have a follow-up visit to a primary care provider between 2 to 4 weeks after an inpatient acute care discharge.[5] Our results indicate that this recommendation is not closely followed. We found that close to 19% of the entire study cohort were discharged

from ED or hospital and remained alive during the episode but did not have any follow-up contacts with the healthcare systems. Although these patients represented a small percentage of the cohort, quality of care improvement programs could be targeted towards them in order to fully align COPD treatment with recommended care practice.

A previous result [30] showed that the way in which substance abuse episodes were initiated (e.g., inpatient hospitalization) was associated with outcomes such as cost. Our study of a cohort of newly-diagnosed COPD patients indicate that considerable heterogeneity exists in healthcare costs even among individuals whose episodes of care were initiated by the same healthcare service, and these variations may be partly explained by patient demographic and disease characteristics. Previous studies identified age [13] and comorbidities [30] as predictors of healthcare utilization. We found that patient age, gender, and comorbid conditions were associated with care pathways. Although these characteristics are not modifiable, they are useful for identifying patients who might follow either complex pathways or pathways which are not aligned with recommended care practices, especially in subsequent episodes.

The study has some limitations. First, a common limitation of studies that have used administrative health data to construct episodes of care, is the difficulty to consistently distinguish between scheduled and unscheduled healthcare visits. Although information about healthcare visit schedules might be used to assess the quality of care, this information is not routinely collected in some databases.[32] 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.[29] Lastly, we used per diem rates and charges to estimate some cost components. However, hospitalization, which was the major component of

episode costs, was based on a standard methodology developed by CIHI to reflect actual resource utilization.

Despite these limitations, this study demonstrates a practical approach for describing patient care pathways. By linking various administrative health databases, our study describes patient journeys covering the primary, emergency, and acute care settings. This is the first study to provide such a comprehensive description of care pathways; important information for anticipating and allocating resources to meet the needs of this patient sub-group. This is particularly relevant because this patient sub-group is expected to increase in the future, placing even higher economic burden on the healthcare system.[7]

In conclusion, the LCA approach identified the major pathways patients with COPD followed through a provincially-insured healthcare system during episodes of the disease exacerbation. Patients who followed complex pathways could benefit from case management interventions in order to improve quality of care. The minority of patients whose pathways were not consistent with the recommended follow-up care should be further investigated to fully align COPD treatment in the province with recommended care practices. Our modeling approach could be applied to other resource-intensive health conditions, such as hypertension and diabetes.

3.7 References

1. World Health Organization. Chronic obstructive pulmonary disease. Geneva. 2015. Available at: http://www.who.int/mediacentre/factsheets/fs315/en/. Accessed June 20, 2015.

2. 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.

3. Seemungal TA, 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):1608–1613.

4. Strassels SA, Smith DH, Sullivan SD, Mahajan PS. The costs of treating COPD in the United States. *Chest*. 2001;119(2):344–352.

5. Sharifabad MA, Parsons JP. COPD. *BMJ Best Practice*. Oct 2014. Available at: http://bestpractice.bmj.com/best-practice/monograph/7/follow-up/recommendations.html. Accessed January 20, 2015.

6. Dobson A, DaVanzo JE, El-Gamil A, et al. Clinically appropriate and cost-effective placement study. *Final Report by Dobson DaVanzo & Associates, LLC*. 2012. Available at: http://ahhqi.org/images/pdf/cacep-report.pdf. Accessed December 22, 2014.

7. Mittmann N, Kuramoto L, Seung SJ, Haddon JM, Bradley-Kennedy C, Fitzgerald JM. The cost of moderate and severe COPD exacerbations to the Canadian healthcare system. *Respir Med*. 2008;102(3):413–21.

8. Oostenbrink JB, Rutten-van Mölken MPMH. Resource use and risk factors in high-cost exacerbations of COPD. *Respir Med*. 2004;98(9):883–891.

9. Adeyemi S, Demir E, Chaussalet T. Towards an evidence-based decision making healthcare system management: Modelling patient pathways to improve clinical outcomes. *Decis Support Syst*. 2013;55(1):117–125.

10. Hornbrook MC, Hurtado AV, Johnson RE. Health care episodes: definition, measurement and use. *Med Care Res Rev*. 1985;42(2):163–218.

11. Carey K. Measuring the hospital length of stay/readmission cost trade-off under a bundled payment mechanism. *Health Econ*. 2015;24:790–802.

12. Moos RH, Moos BS. Long-term influence of duration and frequency of participation in alcoholics anonymous on individuals with alcohol use disorders. *J Consult Clin Psychol*. 2004;72(1):81–90.

13. Van Dijk CE, Hoekstra T, Verheij RA, et al. Type II diabetes patients in primary care: profiles of healthcare utilization obtained from observational data. *BMC Health Serv Res*. 2013;13:7.

14. Hastings SN, Whitson HE, Sloane R, Landerman LR, Horney C, Johnson KS. Using the past to predict the future: latent class analysis of patterns of health service use of older adults in the emergency department. *J Am Geriatr Soc*. 2014;62(4):711–5.

15. Yu AP, Yang H, Wu EQ, Setyawan J, Mocarski M, Blum S. Incremental third-party costs associated with COPD exacerbations: a retrospective claims analysis. *J Med Econ*. 2011;14(3):315–323.

16. Dalal AA, Shah M, D'Souza AO, Rane P. Costs of COPD exacerbations in the emergency department and inpatient setting. *Respir Med*. 2011;105(3):454–460.

17. Roos LL, Gupta S, Soodeen R-A, Jebamani L. Data quality in an information-rich environment: Canada as an example. *Can J Aging / La Rev Can du Vieil*. 2005;24(S1):153–170.

18. Gershon A, Wang C. Identifying individuals with physcian diagnosed COPD in health administrative databases. *COPD J Chronic Obstr Pulm Dis*. 2009;6(October):388–394.

19. 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.

20. Canadian Institute for Health Information. Health indicators: definitions, data sources and rationale. 2011. Available at: http://www.cihi.ca/CIHI-ext-portal/pdf/internet/DEFINITIONS_062011_EN. Accessed March 24, 2015.

21. Canadian Institute for Health Information: Cost of a standard hospital stay. 2014. Available at: http://indicatorlibrary.cihi.ca/display/HSPIL/Cost+of+a+Standard+Hospital+Stay. Accessed January 15, 2015.

22. Statistics Canada: Comsumer price index. 2014. Available at: http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/econ161a-eng.htm. Accessed April 20, 2015.

23. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? *J Health Soc Behav*. 1995;36(1):1–10.

24. 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):373–383.

25. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care*. 2005;43(11):1130–1139.

26. Lanza ST, Collins LM. A new SAS procedure for latent transition analysis: transitions in dating and sexual risk behavior. *Dev Psychol*. 2008;44(2):446–56.

27. Schwartz G. Estimating the dimension of a model. *Ann Statist.* 1978;6:461–464.

28. Lanza ST, Collins LM, Lemmon DR, Schafer JL. PROC LCA: A SAS Procedure for Latent Class Analysis. *Struct Equ Model A Multidiscip J*. 2007;14(4):671–694.

29. Lanza ST, Tan X, Bray BC. Latent class analysis with distal outcomes: A flexible model-based approach. *Struct Equ Modeling*. 2013;20(1):1–26.

30. 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.

31. El Fakiri F, Foets M, Rijken M. Health care use by diabetic patients in the Netherlands: patterns and predicting factors. *Diabetes Res Clin Pract*. 2003;61(3):199–209.

32. Doupe M, Doupe M, Kozyrskyj A, Soodeen R-A, et al. An initial analysis of emergency departments and urgent care in Winnipeg. 2008. Available at: http://mchp-appserv.cpe.umanitoba.ca/reference/EDFull.pdf. Accessed November 15, 2014.

Table 3-1: Item-response probabilities estimated from latent class analysis to identify care pathways by episode initiation

Hospital-initiated episodes	Pathway 1 (*n=69*)	Pathway 2 (*n=343*)	Pathway 3 (*n=551*)
		λ (standard error)	
ED visit	1.00 (0.00)	0.01 (0.01)	0.01 (0.01)
GP visit	0.71 (0.07)	0.82 (0.02)	0.32 (0.03)
Specialist visit	0.91 (0.05)	0.46 (0.03)	0.35 (0.02)
Drug dispensation	0.44 (0.08)	0.94 (0.07)	0.00 (0.00)
Another hospitalization	0.57 (0.08)	0.08 (0.02)	0.00 (0.01)
Emergency department-initiated episodes	Pathway 1 (*n=99*)	Pathway 2 (*n=177*)	Pathway 3 (*n=88*)
		λ (standard error)	
Hospitalization	0.91 (0.19)	0.39 (0.05)	0.00 (0.00)
GP visit	0.73 (0.05)	0.19 (0.06)	0.65 (0.07)
Specialist visit	0.70 (0.05)	0.24 (0.04)	0.25 (0.14)
Drug dispensation	0.72 (0.07)	0.00 (0.00)	0.83 (0.13)
Another ED visit	0.23 (0.04)	0.02 (0.02)	0.12 (0.06)

Note: ED = emergency department; GP = general practitioner; λ = item-response probability that the healthcare service was used by individuals in the pathway

Table 3-2: Results of multinomial logistic regression predicting patient care pathway membership

	Hospital-initiated episodes (*n*=963)		Emergency department-initiated episodes (*n*=364)	
	Pathway 1 [a]	Pathway 2 [a]	Pathway 1 [b]	Pathway 3 [b]
	Odds ratio (95% confidence interval)			
Age				
35 – 54	2.03 (0.82 – 5.03)	0.84 (0.52 – 1.37)	0.88 (0.36 – 2.12)	1.01 (0.39 – 2.63)
55 – 74	2.08 (1.11 – 3.90) *	1.17 (0.87 – 1.54)	0.93 (0.53 – 1.63)	1.23 (0.65 – 2.33)
75+ (Reference)				
Gender				
Female	0.85 (0.47 – 1.54)	1.58 (1.20 – 2.09) *	1.43 (0.85 – 2.40)	0.55 (0.30 – 0.99) *
Male (Reference)				
Residence location				
Urban	1.85 (0.90 – 3.81)	1.19 (0.87 – 1.63)	0.75 (0.36 – 1.56)	1.91 (0.67 – 5.46)
Rural (Reference)				
Charlson comorbidity index				
0 (Reference)				
1	0.44 (0.18 – 1.07)	0.66 (0.47 – 0.95) *	1.53 (0.68 – 3.44)	0.36 (0.11 – 1.16)
≥2	0.83 (0.43 – 1.59)	0.46 (0.33 – 0.65) *	0.51 (0.23 – 1.16)	0.32 (0.11 – 0.93) *
COPD diagnosis fiscal year				
2007/08 (Reference)				
2008/09	0.96 (0.41 – 2.25)	0.99 (0.64 – 1.53)	0.82 (0.37 – 1.79)	1.10 (0.51 – 2.40)
2009/10	0.94 (0.40 – 2.22)	0.89 (0.57 – 1.39)	1.29 (0.61 – 2.73)	0.50 (0.20 – 1.26)
2010/11	0.69 (0.32 – 1.48)	0.96 (0.66 – 1.39)	1.83 (0.95 – 3.54)	0.97 (0.46 – 2.06)
Congestive heart failure [c]				

No (Reference)				
Yes	1.31 (1.04 – 1.65) *	0.99 (0.67 – 1.47)	0.86 (0.39 – 1.86)	1.32 (0.45 – 3.88)
Diabetes [c]				
No (Reference)				
Yes	0.63 (0.28 – 1.42)	0.76 (0.51 – 1.13)	1.33 (1.02 – 2.74) *	0.73 (0.22 – 2.45)
Hypertension [c]				
No (Reference)				
Yes	0.84 (0.45 – 1.56)	0.59 (0.37 – 0.92) *	1.44 (1.04 – 2.00) *	1.42 (0.77 – 2.64)

Note: COPD = chronic obstructive pulmonary disease; [a] Reference group was Pathway 3; [b] Reference group was Pathway 2; [c] The odds ratios (95% confidence intervals) reported were adjusted for age, gender, residence location, and COPD diagnosis fiscal year in separate multinomial logistic regression model.

* Significant at $\alpha = 0.05$

Table 3-3: Characteristics of health outcomes overall and by care pathways

Hospital-initiated episodes	Overall (*n=963*)	Pathway 1 (*n=69*)	Pathway 2 (*n=343*)	Pathway 3 (*n=551*)	P-value [a]
Costs, mean (SD) [b]	11373.3 (24531.3)	29479.7 (54907.9)	9817.1 (21865.9)	10619.1 (21042.9)	0.044
Death during episode (%)	13.6	15.9	6.1	18.0	<0.001
Emergency department-initiated episodes	Overall (*n=364*)	Pathway 1 (*n=99*)	Pathway 2 (*n=177*)	Pathway 3 (*n=88*)	
Costs, mean (SD) [b]	4557.5 (7170.4)	9204.1 (10019.2)	3493.3 (5279.7)	943.2 (2992.2)	<0.001
Death during episode (%)	8.0	10.1	10.7	0.0	0.006

Note: COPD = chronic obstructive pulmonary disease; SD = standard deviation

[a] Differences between pathway groups were tested with the χ^2 statistics, using the Auxiliary option of the Variable command in Mplus. We specified (DCON) for cost and (DCAT) for death. These options produce the results for the method proposed by Lanza et al. [29]

[b] Costs are reported in CAD constant dollars for 2011/12

Figure 3-1: Mean (95% confidence intervals) episode duration (a) and care contacts (b) by care pathway

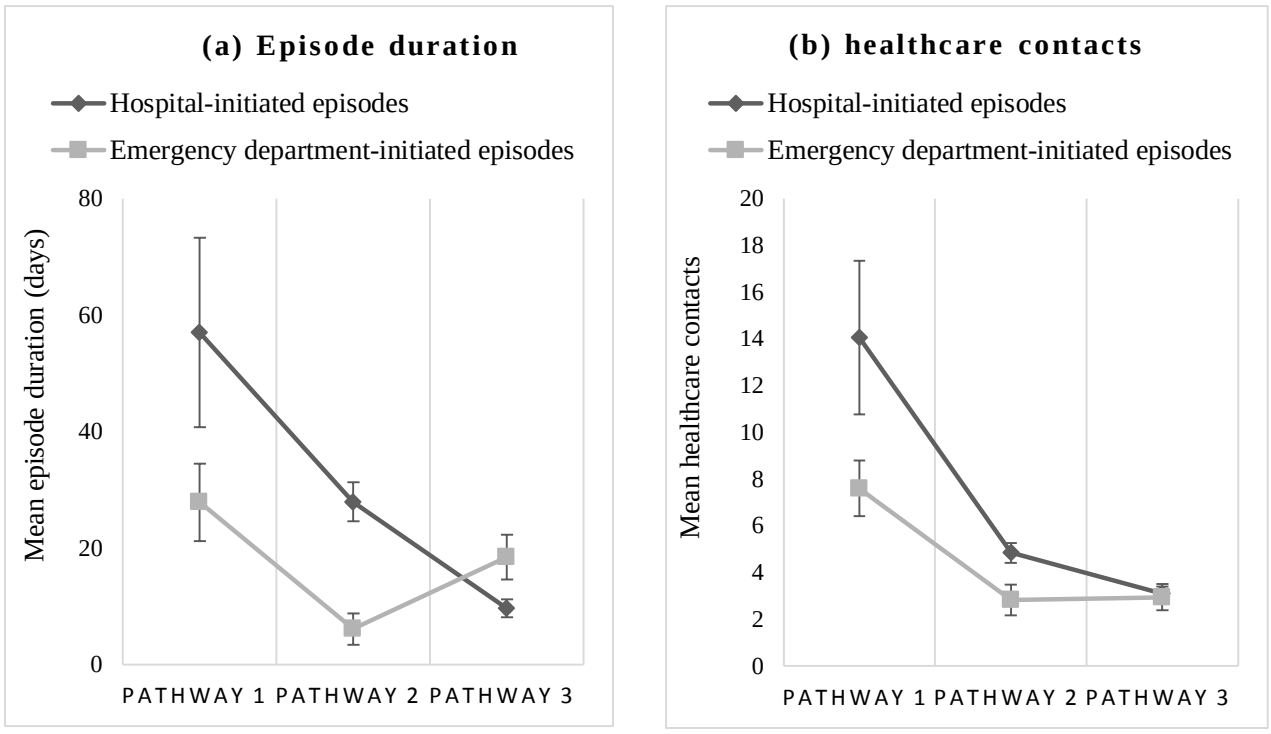
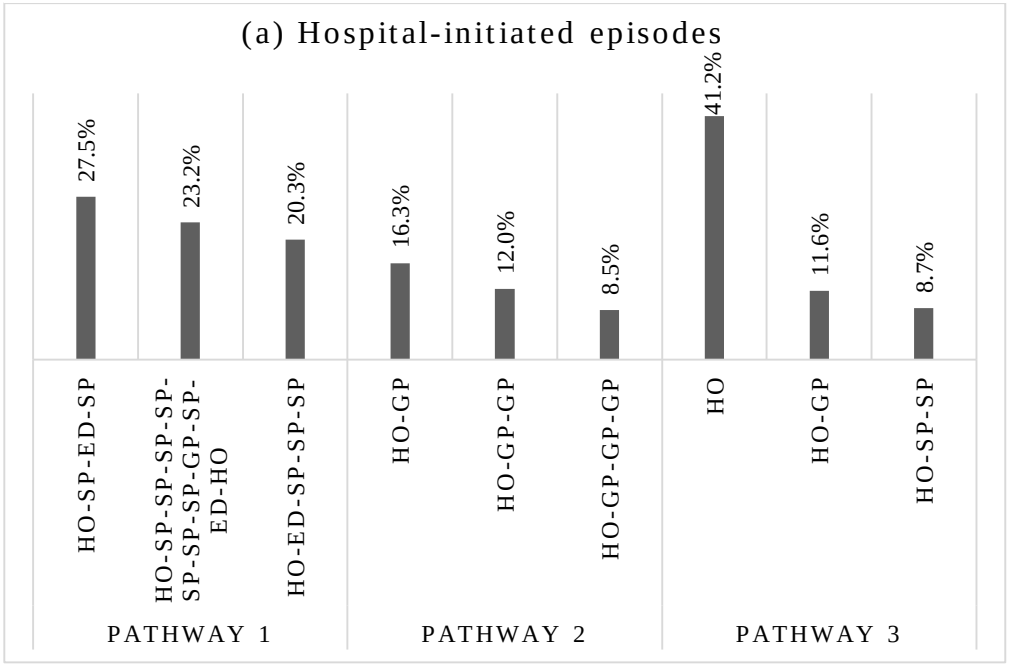
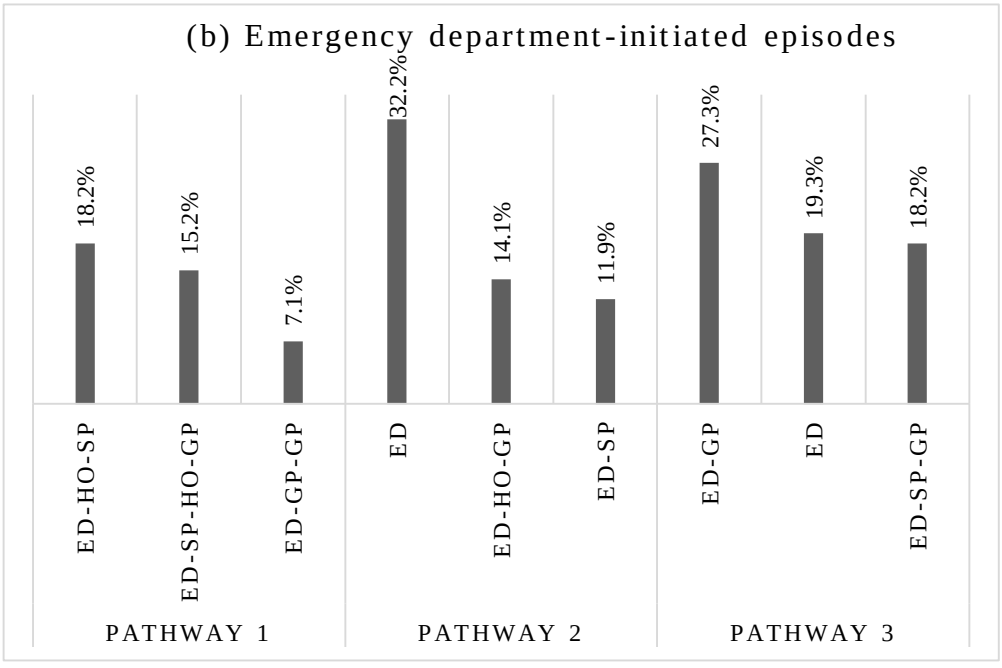


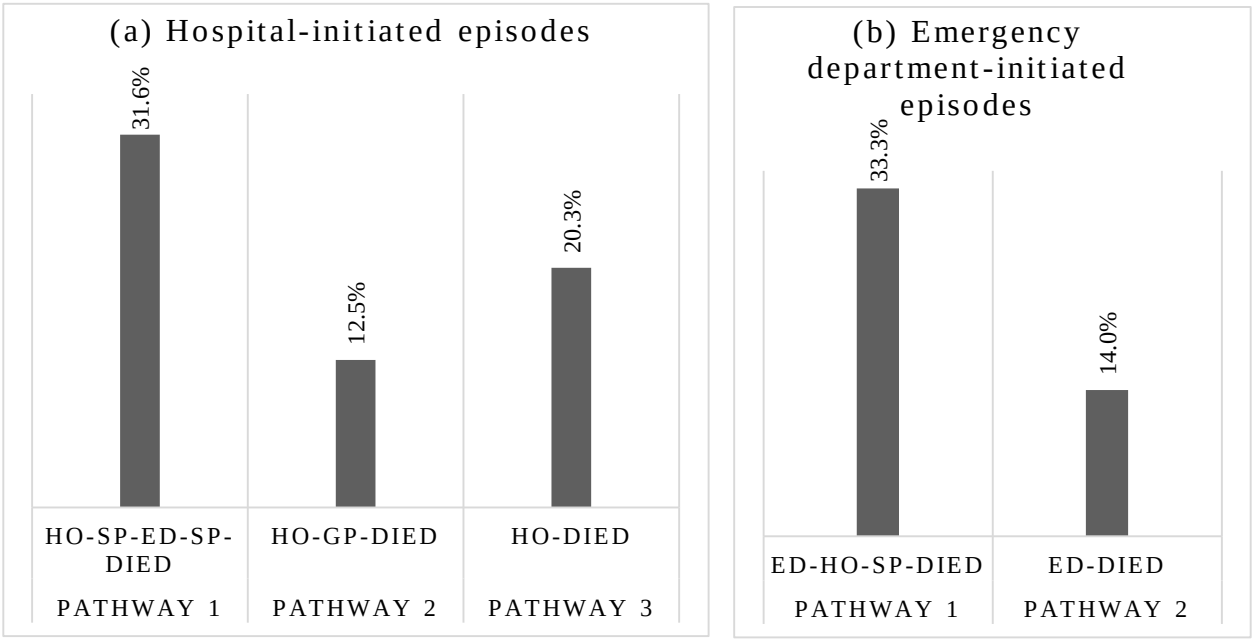
Figure 3-2: Frequent healthcare contact sequences by care pathway for hospital-initiated (a) and emergency department-initiated (b) episodes





Notes: ED = emergency department; GP = general practitioner; HO = hospitalization; SP = specialist

Figure 3-3: Percentage of individuals who died in the most frequent healthcare contact sequences for hospital-initiated (a) and emergency department-initiated (b) episodes



Notes: ED = emergency department; GP = general practitioner; HO = hospitalization; SP = specialist

CHAPTER 4: HIGH-COST PERSISTENCE IN EPISODES OF CHRONIC OBSTRUCTIVE PULMONARY DISEASE EXACERBATIONS

4.1 Chapter overview

Studies have consistently reported that very few individuals account for more than half of healthcare costs in various populations. Recent analyses also revealed that these high-cost patients are a heterogeneous sub-group, with some patients persistently incurring high cost whilst others occasionally incurring high cost. Building on the previous study (Chapter 3), which identified patients who followed complex pathways and incurred high costs, this study examined whether high-cost patients in one episode also incurred high costs in subsequent episodes. Given that the prevalence of chronic obstructive pulmonary disease (COPD) is projected to increase in the future and place even higher economic burden on the healthcare system, identifying models which could predict high-cost groups in the early episodes would be important for timely interventions.

Publication details:

Kuwornu JP, Lix LM, Quail JM, Forget E, Muthukumarana S, Wang XE, Osman M, Teare GF. High-cost persistence in episodes of chronic obstructive pulmonary disease exacerbations. *Canadian Respiratory Journal* [To be submitted]

4.2 Abstract

Objective: To estimate healthcare costs associated with episodes of COPD exacerbations, and compare the ability of different models to predict whether a patient will incur high costs persistently.

Methods: Linked health administrative databases from Saskatchewan, Canada were used to identify a cohort of newly-diagnosed COPD patients and their episodes of healthcare use for disease exacerbations. We computed total costs during index (i.e., first) and follow-up episodes and categorized patients into persistently high-cost, occasionally high-cost and persistently low-cost groups based on cumulative cost distribution ranking and using the 75^{th} percentile cutoff for high cost status. Average healthcare costs in the two episodes were compared between the three groups. The *c*-statistic was used to compare the discrimination ability of logistic regression models containing only socio-demographic and disease characteristics with models containing these characteristics in addition to healthcare utilization characteristics.

Results: Of the n = 1,182 incident cohort members, 8.5% were classified as persistently high-cost, 26.1% as occasionally high-cost, and the remainder (65.4%) as persistently low-cost. The persistently high-cost and occasionally high-cost patients incurred about 10 times ($ 12,449 vs $ 1,263) and seven times ($ 9,334 vs $ 1,263) more costs in the index episode than the persistently low-cost patients, respectively. The *c*-statistic of the models predicting persistently high-cost patients ranged from 0.74 to 0.88.

Conclusion: The costs associated with episodes of COPD exacerbations are substantial; with a small number of patients incurring high healthcare expenditures persistently. The results indicate that patient demographic, disease, and healthcare encounter characteristics, which are all

routinely collected in administrative health databases, could accurately identify patients who are more likely to become persistent high-cost users.

4.3 Introduction

Patients with 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 of patient symptoms warranting additional treatment,[2] are major contributors to the total healthcare cost of treating the disease; with annual costs estimated to be 10 times higher among people with COPD who experienced exacerbations than those who did not.[3]

Treatment of COPD exacerbations often requires an ED visit or hospitalization.[4,5] Patients may also receive follow-up care from their primary care provider or a specialist physician, and might require additional medications.[6] Accordingly, an exacerbation episode may require contacts with a number of different healthcare providers and services. A comprehensive description of costs associated with COPD exacerbations requires an episode-of-care data system, which collects all healthcare services related to the treatment of a specific health condition.[7] The episode of care provides a clinically meaningful unit for measuring healthcare costs,[8] and allows for a detailed analysis of the treatment processes that generate costs.[9]

The phenomenon of a very few individuals, usually the top five percent to 10 percent of healthcare users, accounting for more than 50% of healthcare costs has been consistently reported in the literature.[10] However, recent analyses have also revealed that these high-cost patients are a heterogeneous sub-group, with some patients persistently incurring high cost while others occasionally incur high cost.[11] Understanding this dynamic nature of healthcare expenditures would be important for developing cost-management strategies. Given that the

prevalence of COPD is projected to increase in the future and place even higher economic burden on the healthcare system,[12] identifying models which could predict high-cost groups in the early episodes would be important for timely interventions. Given this background, the objectives of this study were to estimate healthcare costs associated with episodes of COPD exacerbations, and compare the ability of different models to predict whether a patient will incur high costs persistently.

4.4 Methods

4.4.1 Data sources

We used population-based 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. The province maintains multiple administrative health databases in electronic format, which can be anonymously linked via a unique personal health number.[13]

Episodes of care for COPD were constructed using databases that capture the vast majority of care, including primary, emergency, and acute care service, for provincial health insurance beneficiaries: 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 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, costs, 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.

4.4.2 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 twelve health regions in Saskatchewan and the only ones for which ED data were available. Both SHR and RQHR are the only health regions which 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: (1) one or more hospitalizations with a diagnosis of COPD in any diagnosis field, or (2) one or more physician visits with a diagnosis of COPD.[14] This case definition has a sensitivity of 85.0% and a specificity of 78.4% when compared with clinical evaluation by a physician.[1] The index date of COPD diagnosis was the earliest hospitalization or physician visit date for COPD. Cases were identified from hospital discharge abstracts using the following ICD-10-CA codes: J41, J42, J43

or J44; whilst cases in the 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 or not 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. We restricted the cohort to an incident cohort in order to study changes in healthcare utilization and costs as the condition progresses. The cohort was also restricted to only those who had continuous provincial health insurance coverage between April 1, 2002 and March 31, 2012, so that all insured healthcare contacts could be captured. 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 as explain in the next section.

4.4.3 Defining episodes of care for COPD exacerbations

All episodes of care for COPD exacerbations following the index diagnosis were defined using the healthcare services which initiated, continued, and ended them. We identified episodes of care based on a method developed by the CIHI, a national non-profit organization that provides standardized methods and data sources for health services research, for ascertaining exacerbations. [16] Hospital- or ED-initiated episodes had: (1) a COPD diagnosis in the most responsible diagnosis field, or (2) 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 2 days of a physician visit.

An episode continued if there were respiratory-related hospitalizations, or ED, GP, 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 on-going episodes at the end of the observation period were excluded to ensure that we had complete information to estimate cost of all episodes included in the study.

4.4.4 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. Hospital costs were estimated based on a standard methodology developed by CIHI.[17] 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 prescription drug 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 2011/12 constant dollars.

4.4.5 Study measures

4.4.5.1 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 in the 75th percentile and above in the two episodes), occasionally high-cost (i.e., those whose costs were in 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 the two episodes). The choice of a cutoff point is largely empirically driven,[19] and a number of 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 samples that were too small to provide stable estimates in regression model.[22]

4.4.5.2 Health services utilization measures

For each patient, we tracked the number and duration of use of various healthcare services in each episode (please see Table 4-1 for the definitions of these utilization variables). Among patients who were hospitalized, the number of hospital admissions and the total number of days they spent in hospitals were computed. To further examine the intensity of healthcare utilization, we calculated the number of times patients were admitted to special care units (SCUs) and the total days they spent in these SCUs during hospitalizations. We computed the number of visits and total number of days a patient spent in EDs during episodes, as well as the number of visits to GPs and specialists during episodes. The number of different types of drugs used to treat respiratory-related conditions were counted in each episode. These were identified based on drug identification numbers, and included drugs such as oral corticosteroids and antibiotics. Finally,

we calculated the total number of days in an episode of care. Only the healthcare utilization measures counting the number of separate encounters, such as number of hospital admissions, were included in the prediction models as discussed in the statistical analysis section. Measures of duration of utilization, such as total number of days spent in hospitals, were only included in the descriptive analysis.

4.4.5.3 Patient and disease characteristics

The patient and disease characteristics included in the analysis were based on the Andersen healthcare utilization model.[28] 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 was the Charlson comorbidity index,[29] which was based on diagnoses in the hospital discharge abstract and the physician billing claims data. The index was categorized as 0, 1, 2, or ≥ 3. We also included the fiscal year of COPD diagnosis (i.e., 2007/08, 2008/09, 2009/10, or 2010/11). All variables were defined as of the index date of COPD diagnosis except for the comorbidity index, which was calculated using data from the year prior to the index date.

4.5 Statistical analysis

We described overall and individual components of episodes of care costs with means (standard deviations; SDs). The χ^2 statistic was used to test for differences in patients' healthcare encounters in the three cost groups. We also plotted the durations (in days) of various healthcare utilizations during episodes.

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). Studies [21] have shown that including the number of previous healthcare services utilization would enhance a model's ability to predict 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 GP visits (model 3), number of specialist visits (model 4), number of drugs dispensed (model 5), and all five healthcare utilization measures (model 6).

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 for the data 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 are accepted as having adequate classification accuracy.[30] To evaluate the models' discrimination ability, we conducted two pairwise logistic regressions 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 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.

4.6 Results

4.6.1 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%]) the remaining 7,099 individuals were eligible for study inclusion. During follow-up, 2,659 individuals experienced at least one COPD episode of care, which resulted in a total of 5,348 episodes. The final cohort (n = 1,182) was comprised of all individuals who experienced at least two COPD episodes of care during the follow-up period. The average follow-up time was 1,358.2 days (SD = 320.9 days).

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.

4.6.2 Episode costs

Average episode costs are summarized in Table 4-2. 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 $ 1,263.45). Similarly, the occasionally high-cost patients incurred a little over seven times more costs than the persistently low-cost patients in the index episode ($ 9,334.61 vs $ 1,263.45). 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 the highest costs compared to the other two cost groups. Similarly, patients in the persistently high-cost group incurred higher average total costs in the follow-up episode than those in the other groups.

4.6.3 Health services utilization during episodes

The number of hospital admissions, SCU admissions, ED visits, GP visits, and specialist visits were significantly different among the three cost groups in the two episodes (*P* value <0.0001 for all services) (Table 4-3). 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 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 than patients in the other two groups. However, a higher percentage of patients in the persistently low-cost group utilized more GP 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 4-1). The average number of days in EDs did not differ significantly 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.

4.6.4 Multinomial logistic regression results

In the multinomial models (Table 4-4), compared to 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 to those with no comorbid condition, patients with Charlson comorbidity score of 1 (OR = 2.68, 95% CI 1.51 – 4.77), score of 2 (OR = 2.28, 95% CI 1.17 – 4.42), or score ≥ 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.

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

With regards to healthcare utilization, one more hospital admission in the index episode significantly increased the likelihood of a patient being in the persistently high-cost group (OR = 37.76, 95% CI 21.50 – 66.35) and occasionally high-cost group (OR = 10.21, 95% CI 7.45 – 14.01) than in the persistently low-cost group.

4.6.5 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 4-5). 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 again 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.

4.7 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 disease-related costs.

The findings indicate that 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 [10,11,31] revealed that older patients are more likely than younger ones to be in the persistently high-cost group. We found that older age (75+ years) is 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 (data not include, but available upon request). Thus, it is likely that the rest 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.[32] Previous studies have developed models to predict patients who might become high-cost users in the future, with *c*-statistic ranging from 0.81 to 0.85.[33,34] Our models which predicted persistently high-cost patients had *c*-statistics ranging between 0.74 and 0.88, whilst those which 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 to 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.[35] 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.[36] Scheduled visits beyond 30 days may be counted as part of a new episode. However, recommended practice [6] 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 per diem rates and charges to estimate some cost components. However, hospitalization, which was the major component of episode costs, was based on a standard methodology developed by CIHI to reflect actual resource utilization. Lastly, the prediction accuracy of the models compared in this study were based on the model building data set only. There is the need to validate these models in independent data sets.

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.[37,38] Understanding healthcare costs based on the episodes of care is important for adopting these innovative provider payments schemes.

4.8 Conclusions

The costs associated with episodes of COPD exacerbations are substantial; with a small number of patients incurring high healthcare expenditures persistently. Our results indicate that patient demographic, disease, and healthcare encounter characteristics, which are all routinely

collected in administrative health databases, could be used to accurately identify patients who are more likely to become persistent high-cost users. Being able to identify persistently high-cost patients is important for implementing strategies to manage costs and improve quality of life.

4.9 References

1. 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.

2. Burge S, Wedzicha JA. COPD exacerbations: definitions and classifications. *Eur Respir J Suppl*. 2003;41:46s–53s.

3. Blanchette CM, Gross NJ, Altman P. Rising costs of COPD and the potential for maintenance therapy to slow the trend. *Am Heal Drug Benefits*. 2014;7(2):98–106.

4. Strassels SA, Smith DH, Sullivan SD, Mahajan PS. The costs of treating COPD in the United States. *Chest*. 2001;119(2):344–352.

5. Seemungal TA, 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):1608–1613.

6. Sharifabad MA, Parsons JP. COPD. *BMJ Best Practice*. Oct 2014. Available at: http://bestpractice.bmj.com/best-practice/monograph/7/follow-up/recommendations.html. Accessed January 20, 2015.

7. Hornbrook MC, Hurtado AV, Johnson RE. Health care episodes: definition, measurement and use. *Med Care Res Rev*. 1985;42(2):163–218.

8. 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):319–325.

9. Miller HD. From volume to value: better ways to pay for health care. *Health Aff*. 2009;28(5):1418–1428.

10. 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):347–369.

11. Riley GF. Long-term trends in the concentration of medicare spending. *Health Aff*. 2007;26(3):808–816.

12. 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.

13. Roos LL, Gupta S, Soodeen R-A, Jebamani L. Data quality in an information-rich environment: Canada as an example. *Can J Aging / La Rev Can du Vieil*. 2005;24(S1):153–170.

14. Gershon A, Wang C. Identifying individuals with physcian diagnosed COPD in health administrative databases. *COPD J Chronic Obstr Pulm Dis*. 2009;6(October):388–394.

15. 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.

16. 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.

17. 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.

18. Statistics Canada: Comsumer price index. Ottawa, ON. 2014. Available at: http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/econ161a-eng.htm. Accessed April 20, 2015.

19. Coughlin TA, Long SK. Health care spending and service use among high-cost medicaid beneficiaries, 2002-2004. 2013;46(4):405–417.

20. 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–E118.

21. 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.

22. Fleishman JA, Cohen JW. Using information on clinical conditions to predict high-cost patients. *Health Serv Res*. 2010;45(2):532–552.

23. Weir S, Aweh G, Clark RE. Case selection for a Medicaid chronic care management program. *Health Care Financ Rev*. 2008;30(1):61–74.

24. Baser O, Burkan A, Baser E, et al. High cost patients for cardiac surgery and hospital quality in Turkey. *Health Policy (New York)*. 2012;109(2):143–149.

25. 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.

26. Freeborn DK, Pope CR, Mullooly JP, McFarland BH. Consistently high users of medical care among the elderly. *Med Care*. 1990;28(6):527–540.

27. Leeb K, Jokovic A, Sandhu M, Zinck G. Intensive care in Canada. *Healthc Q*. 2006;9(1):32–33.

28. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? *J Health Soc Behav*. 1995;36(1):1–10.

29. 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):373–383.

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.

31. Coughlin TA, Long SK. Health care spending and service use among high-cost medicaid beneficiaries, 2002-2004. *Inquiry*. 2009;46:405–417.

32. Monheit AC. Persistence in health expenditures in the short run: prevalence and consequences. *Med Care*. 2003;41(7 Suppl):III53–III64.

33. Meenan RT, Goodman MJ, Fishman PA, et al. Using risk-adjustment models to identify high-cost risks. *Med Care*. 2003;41(11):1301–1312.

34. 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):440–447.

35. Doupe M, Kozyrskyj A, Soodeen R-A, et al. An initial analysis of emergency departments and urgent care in Winnipeg. 2008. Available at: http://mchp-appserv.cpe.umanitoba.ca/reference/EDFull.pdf. Accessed November 15, 2014.

36. 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.

37. Davis K, Guterman S. Rewarding excellence and efficiency in Medicare payments. *Milbank Q*. 2007;85(3):449–68.

38. 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.

Table 4-1: Definitions of healthcare utilization variables

Variable	Definition	Analysis in which the 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 GP visits	The number of times a patient visited GPs during episode	√	√
No. of Specialist visits	The number of times a patient visited specialist physicians during episode	√	√
No. of drugs dispensed	The number of different types of out-patient drugs dispensed during the episode, used for treating respiratory-related conditions	√	√
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: ED = emergency department; GP = general practitioner; SCU = special care units; √ = variable was included in the specified analysis

[a] These variables were defined for only those who had hospital admission

Table 4-2: Episode of care cost components by cost group and episode of care following COPD diagnosis

	Index episode					Follow-up episode				
	Hospital	ED	Physician	Drugs	Total episode cost	Hospital	ED	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)	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)	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)	4538.24 (835.73)	339.94 (100.79)	125.27 (169.87)	91.73 (150.30)	1263.45 (2028.50)	4259.17 (971.)	385.63 (260.49)	110.40 (188.08)	90.88 (129.28)	714.40 (1450.62)

Notes: ED = emergency department; costs are reported in CAD constant dollars for 20011/12

Table 4-3: 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 GP 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 drugs dispensed						
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

Notes: Notes: SCU = special care unit; ED = emergency department; GP = general practitioner; SD = standard deviation

* Utilization distributions in the three cost groups are significantly different using a χ^2 test at p <.0001

Table 4-4: Odds ratios (ORs) and 95% confidence intervals (CIs) for multinomial logistic regression models

Characteristic	All (*n*=1182)	Persistently high-cost patients (*n*=100) [a]	Occasionally high-cost patients (*n*=309) [a]
	n (%)	ORs (95% CIs)	
Demographic and disease characteristics			
Age group, years			
35 - 54	304 (25.7)	0.19 (0.09 - 0.41) *	0.35 (0.23 - 0.52) *
55 - 74	483 (40.9)	0.53 (0.33 - 0.85) *	0.71 (0.52 - 0.96) *
75+	395 (33.4)	ref	ref
Sex			
Female	575 (48.7)	1.25 (0.81 - 1.93)	1.03 (0.78 - 1.36)
Male	607 (51.4)	ref	161 (52.1)
Residence location			
Urban	893 (75.5)	1.33 (0.79 - 2.25)	1.05 (0.76 - 1.45)
Rural	289 (24.5)	ref	ref
Charlson comorbidity index			
0	764 (64.6)	ref	ref
1	183 (15.5)	2.68 (1.51 - 4.77) *	2.43 (1.68 - 3.52) *
2	121 (10.2)	2.28 (1.17 - 4.42) *	1.83 (1.17 - 2.86) *
≥ 3	114 (9.6)	4.29 (2.30 - 8.00) *	2.67 (1.69 - 4.22) *
Fiscal year of COPD diagnosis			
2007/08	383 (32.4)	ref	ref
2008/09	310 (26.2)	1.22 (0.68 - 2.16)	1.18 (0.82 - 1.69)
2009/10	266 (22.5)	1.45 (0.79 - 2.64)	1.56 (1.07 - 2.25)
2010/11	223 (18.9)	1.12 (0.60 - 2.11)	0.85 (0.56 - 1.29)
Healthcare utilization in index episode [b]			
No. of hospital admissions	---	37.76 (21.50 - 66.35) *	10.21 (7.45 - 14.01) *
No. of ED visits	---	3.32 (2.39 - 4.60) *	2.70 (2.08 - 3.50) *
No. of GP visits	---	1.06 (0.99 - 1.13)	1.06 (1.01 - 1.11) *
No. of specialist visits	---	1.60 (1.45 - 1.77) *	1.49 (1.36 - 1.63) *
No. of drugs dispensed	---	1.04 (1.02 - 1.07) *	1.02 (1.00 - 1.04)

Notes**:** COPD = chronic obstructive pulmonary disease; ED = emergency department; GP = general practitioner; * Statistically significant at $\alpha = 0.05$;

[a] Reference group was persistently low-cost patients; [b] Each of the ORs (95% CIs) reported were separately adjusted for the above demographic and disease characteristics

Table 4-5: Comparison of goodness-of-fit and classification accuracy between multinomial logistic regression models

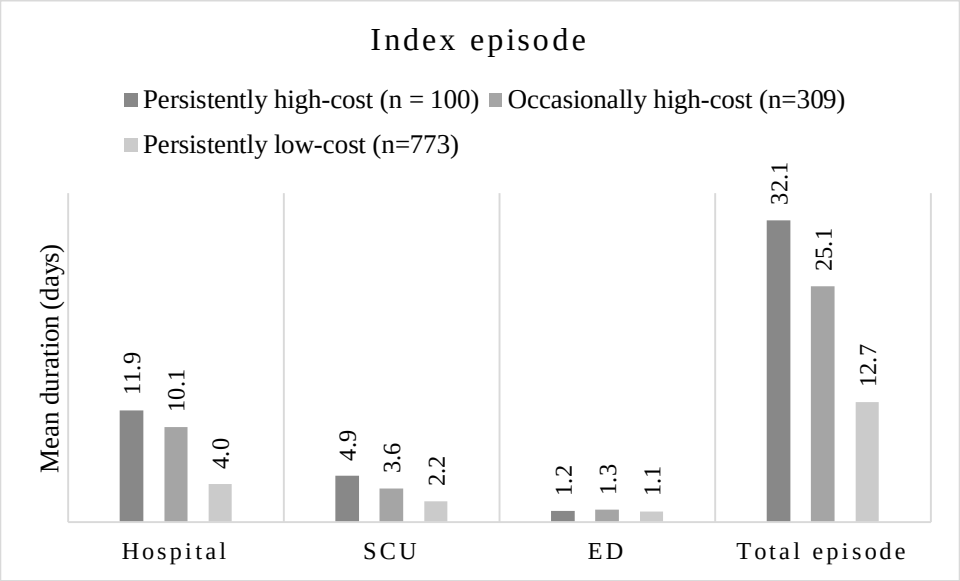
Performance metric	Base model	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Goodness-of-fit, Multinomial models							
-2 Log-likelihood	1842.98	1455.33	1762.55	1835.90	1711.69	1833.55	1429.88
BIC	1998.64	1625.13	1932.33	2005.70	1881.49	2003.35	1656.28
Classification accuracy, Multinomial models							
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	32.3	43.9	32.0	33.2	37.4	32.5	41.4
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

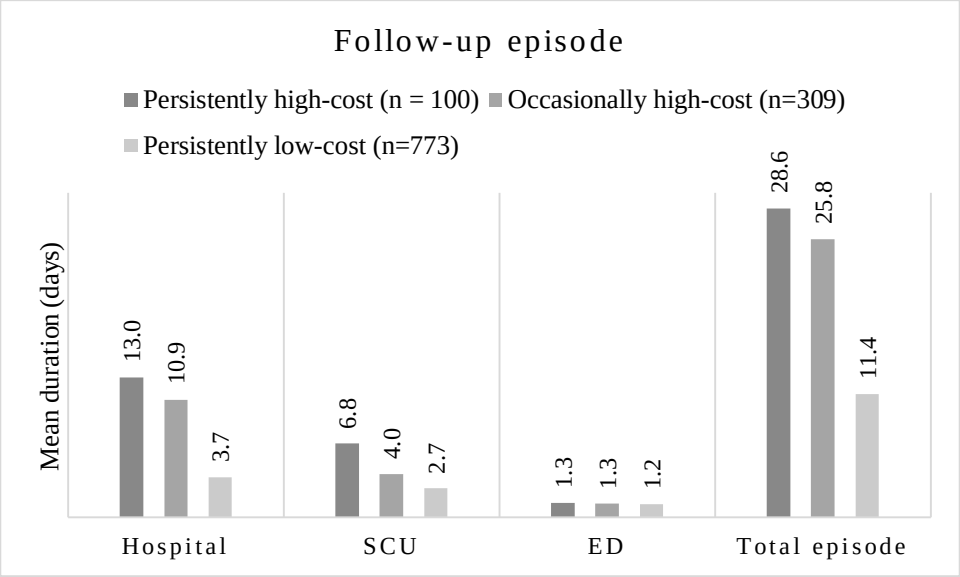
Notes: BIC = Bayesian information criterion; PBCAR = proportional-by-chance accuracy rate

[a] PBCAR = $(100/1,182)^2 + (309/1,182)^2 + (773/1,182)^2 = 0.50$)

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 general 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 of the above healthcare services in index episode

Figure 4-1: Average health services utilization and episode durations by cost group





Notes: ED = emergency department; SCU = special care unit

CHAPTER 5: A COMPARISON OF STATISTICAL MODELS FOR ANALYZING EPISODES OF CARE COSTS FOR CHRONIC OBSTRUCTIVE PULMONARY DISEASE EXACERBATIONS

5.1 Chapter overview

Episode-based remuneration schemes are receiving favourable evaluations, but the methods to facilitate their implementation are still under development. Accurate prediction of episode-based costs is important for fair remuneration of care providers. This manuscript compares the prediction accuracy of eight regression models to identify the best model(s) for predicting episode-based costs. The cohort included in the study differed slightly from the previous studies. This study included both incident and prevalent chronic obstructive pulmonary disease (COPD) cases and covered the entire province of Saskatchewan. Also, we did not include the emergency department (ED) data in constructing episodes of care because these data were only available for the two largest health regions in the province. These modifications ensured that the cohort was large enough to undertake the various statistical models investigated.

Publication details:

Kuwornu JP, Lix LM, Quail JM, Wang E, Osman M, Teare GF. A comparison of statistical models for analyzing episodes-of- care costs for chronic obstructive pulmonary disease exacerbations. *Health Services and Outcomes Research Methodology*. 2013;13(2-4):203–218.

5.2 Abstract

Objective: Accurate predictive models of costs for episodes of healthcare utilization associated with acute and chronic conditions can be used to develop non-fee-for-service provider remuneration systems. This study examined the performance of eight predictive models for costs associated with episodes of care for COPD exacerbations.

Method: Administrative health data including hospital separations, physician billing claims, prescription drug records, and population registration files from Saskatchewan, Canada was used to identify a cohort (35+ years) with diagnosed COPD and define all episodes of healthcare utilization and costs for COPD exacerbations over a nine-year period (fiscal years 2001/02 to 2009/10). Using cross-validation, we examined the performance of eight predictive models for episode of care costs for patients' first episode during the follow-up period: (1) ordinary least squares (OLS) regression, (2) OLS regression on log-transformed costs with normal retransformation, (3) OLS regression on log-transformed costs with heteroscedastic retransformation, (4) OLS regression on log-transformed costs with Duan's retransformation, (5) robust regression (6) generalized linear model (GLM) with Poisson distribution and log link function, (7) GLM with a Gamma distribution and identity link function, and (8) GLM with a Gamma distribution and log link function. The predictors included were age, sex, income group, region of residence and Charlson comorbidity score.

Results: A total of 17,480 individuals with a hospital-initiated episode (n = 7,910) and physician visit-initiated episode (n = 9,570) were identified; of which 51.7% were males and the average age was 71.3 (SD = 12.1). Half of the costs were below $595 CAD, while the 95th percentile was $ 13,934 CAD. Cross-validation results showed that none of the models consistently resulted in the best prediction of episode costs; the GLM Poisson model had the highest R^2, but the OLS

model on the actual cost and OLS model on log-transformed cost had the best prediction accuracy in terms of the root mean square error and mean absolute prediction error respectively. *Conclusions:* The study findings suggest that researchers can consider adopting one of these three models for predicting costs of healthcare use in episodes of care, but also emphasize that further comparisons of model performance are warranted.

5.3 Background

Fee-for-service reimbursement systems that remunerate providers based on the provision of individual healthcare services can result in a fragmented approach to patient-oriented care. Accordingly, there is increased emphasis on remuneration systems based on episodes of care,[1] which pay providers for the cluster of services associated with the treatment or management of a condition or event. Episodes of care reimbursement systems have received favourable evaluations because they are more likely to reward care coordination and enhance efficiency than fee-for-service reimbursement models.[2] An episode of care approach also captures the dynamics of the processes that generate healthcare utilization, such as the timing of services,[3] which may influence healthcare costs. Accurate predictive models of healthcare costs for episodes of care are central to the development and implementation of alternative payment systems.

Scitovsky [4] pioneered methods for the analysis of costs based on episodes of illness by developing a cost-per-episode of illness measurement method; it was used to compare changes over time in costs for five illnesses. Recently, Foster and Xuan [3] demonstrated the reward renewal model for modeling costs of mental health services in an episode-based framework. While neither study examined the performance of predictive models for episode costs, such comparisons have occurred in other related studies. Austin et al.[5] considered different models for predicting length of stay after coronary artery bypass graft surgery and Dodd et al.[6] evaluated

predictive models of healthcare costs for inflammatory bowel disease. Both authors found that different models could result in dissimilar conclusions about (a) the patient characteristics associated with utilization and costs and (b) predicted values of utilization or costs.

Analyses of episodes of care costs for COPD exacerbations (i.e., periods in the disease course that is characterized by a worsening of patient symptoms including shortness of breath, cough, and/or sputum production) are important for developing innovative provider reimbursement models because these resource-intensive events are potentially avoidable if care is well-coordinated between providers. In Canada for example, COPD is the fourth and the sixth most common cause of hospitalization among men and women, respectively.[7] It is estimated that COPD will be the third leading cause of death worldwide by 2020.[8]

The objective of the study was to identify the optimal statistical model(s) for predicting the mean cost of episodes of care for COPD exacerbations. We used measures of prediction accuracy [9] to compare eight regression models.

5.4 Methods

5.4.1 Data sources

Administrative health data from the province of Saskatchewan, Canada, which has a population of approximately 1.1 million,[10] were used to conduct the research. Like all Canadian provinces, Saskatchewan has a universal healthcare program. The province maintains multiple healthcare databases in electronic format and these can be anonymously linked via a unique personal health number.[11] Ethics approval for database access was received from the University of Saskatchewan Biomedical Research Ethics Board. Data were accessed and analyzed at the Health Quality Council in accordance with a standing data sharing agreement between the organization and the health ministry.

COPD episodes of care were constructed using the following databases that capture acute, primary, and supportive care services for provincial health insurance beneficiaries: hospital discharge abstracts, physician billing claims, prescription drug dispensation records, and home care service records. A hospital discharge abstract is completed when a patient is discharged from an acute care facility. Diagnoses are recorded using the International Classification of Disease (ICD) codes; ICD-9 codes were used up to and including the 2000/01 fiscal year (a fiscal year extends from April 1 to March 31). Beginning in 2001/02, ICD-10-CA codes were introduced. Between 3 and 16 diagnoses are captured in the data prior to the introduction of ICD-10-CA and up to 25 diagnoses are captured subsequently. Physician billing claims contain information submitted by physicians who are paid on a fee-for-service basis. A single diagnosis is recorded on each claim using ICD-9 codes. Prescription drug dispensation records contain information on outpatient drugs, including the date of dispensation and national drug identification number (DIN); the latter are linked to codes in the American Hospital Formulary System (AHFS) Pharmacologic-Therapeutic Classification System (www.ashp.org). The AHFS is used to group drugs with similar pharmacologic, therapeutic, and/or chemical characteristics. Home care services are provided based on the assessed needs of individuals to remain independent at home and include nursing care, homemaking, personal care, respite and home management, meal service, and physiotherapy services. The home care database contains records of the annual number of services provided in each of these categories and the date that service was first initiated. The population registry system and vital statistics registry were also used in this study; they provide demographic and socio-economic information, as well as dates of healthcare coverage and death.

5.4.2 Study design and cohort selection

The study adopted a retrospective cohort design. The cohort was composed of individuals 35 years of age or older with: (a) a hospital discharge abstract with a diagnosis of COPD in any diagnosis field or (b) at least two physician visits for COPD within 365 days.[12] The cohort was defined using data from fiscal years 1996/97 to 2009/10. All subjects without continuous insurance coverage within this period were excluded. We excluded those without continuous insurance coverage to ensure that all of a patient's healthcare utilization was captured. The index date of COPD diagnosis was the earliest of either the hospitalization or the date of the first of the two physician visits.

5.4.3 Defining episodes of care for COPD exacerbations

We identified both hospital-initiated and physician visit-initiated episodes of care for COPD exacerbations (see Table 5-1) during the nine-year period from fiscal years 2001/02 to 2009/10; the latter was the most recent available year. Hospital-initiated episodes had: (a) a COPD diagnosis in the most responsible (i.e., first) diagnosis field, or (b) 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 another diagnosis field.[13] Physician-visit initiated episodes were identified by ICD-9 codes for COPD or respiratory infection 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 the relevant physician visit.[13]

Continuation of an episode was determined by identifying respiratory-related hospitalizations or physician visits that followed the initiating event within a 30-day period. An

episode of care for a COPD exacerbation ended after either the occurrence of a 30-day clean period, in which there were no respiratory-related hospitalizations or physician visits, or death.

5.4.4 Defining episode costs for COPD exacerbations

An episode cost was the sum of respiratory-related costs for acute, primary, and supportive healthcare services incurred between the episode start and end dates. Hospital costs reflect hospital resource utilization during an inpatient stay and are estimated based on a standard methodology developed by the CIHI, a national non-profit organization that provides standard methods and data sources for health services research (www.cihi.ca). Full details on CIHI's inpatient hospital cost calculation can be found elsewhere.[14] Briefly, CIHI computes the cost of each inpatient hospitalization by multiplying a resource intensity weight (RIW), which is based on the most responsible diagnosis or intervention, and the cost per weighted case (CPWC). An RIW is a relative value that describes the expected resource consumption of a patient based on: (a) patient case mix group (CMG), (b) factors known to affect resource utilization and length of stay, including age, comorbidity, hospital-based interventions, number of interventions, and out-of-hospital interventions, and (c) atypical length of stay or level of care.[15] Atypical cases include patients who are transferred between facilities, patients who sign themselves out against medical advice, and palliative cases. The CPWC represents the average cost of one patient receiving healthcare services in a hospital. We used cost figures developed specifically for Saskatchewan.

The cost of a physician visit was the amount billed by the physician to the provincial Ministry of Health. In Saskatchewan, the Saskatchewan Medical Association negotiates physician fee schedules with the provincial government on a semi-annual basis. Prescription drug costs were based on prices of the active substance plus a dispensing fee. The cost of home care services was based on per diem values based on average daily costs, which were obtained

by dividing annual costs for home care service delivery by the number of days home care was received during an episode. Only costs for new home care services (i.e., services initiated on or after the date of the COPD episode) were included.

Episode costs were adjusted for inflation using the health and personal care component of the Saskatchewan monthly consumer price indices.[16] They were expressed in 2002/03 constant dollars.

5.4.5 Study variables

The outcome variable was the total episode cost for the COPD exacerbation. For patients with multiple episodes of care during the observation period, only the first episode was selected for analysis. Episodes that were right censored (i.e., ending after March 31, 2010) or left censored (i.e., starting before April 1, 2001) were excluded.

The predictors of interest were age, sex, income group, region of residence and comorbidity. Age was categorized as 35 – 44 years, 45 – 59 years, 60 – 74 years, 75 – 84 years and 85 years and above. Income quintiles were calculated using average household income from the 2006 Statistics Canada Census.[17] Each individual's postal code was assigned to a dissemination area (DA), the smallest geographic unit in the census data. Income ranges were determined such that the entire Saskatchewan population was divided into five approximately equal groups. Residents were assigned an income quintile according to their DA average household income. Residence location was defined using postal codes and dichotomized as rural or urban. Urban residents were those whose postal codes were in a census metropolitan or agglomeration area (i.e., 10,000+ population). All socio-demographic variables were defined as of the index date for COPD case ascertainment. Comorbidity was measured using the Charlson

comorbidity index [18] and dichotomized as ≤ 1 and 2+; the index score was based on diagnoses in hospital and physician data in the year prior to the episode start date.

Additional variables included in the predictive models were episode initiation (i.e., whether an episode was initiated via hospital or a physician visit), death during episode, recency of COPD diagnosis (i.e., whether or not a patient was diagnosed with COPD before April 1, 2001) and fiscal year.

5.4.6 Statistical analysis

We selected the following eight models for predicting episode costs based on previous studies [19, 20, 9, 6, 21] : (a) ordinary least squares (OLS) regression, (b) OLS regression on log-transformed costs with normal retransformation, (c) OLS regression on log-transformed costs with heteroscedastic retransformation, (d) OLS regression on log-transformed costs with Duan's retransformation, (e) robust regression (f) generalized linear model (GLM) with Poisson distribution and log link function, (g) GLM with a Gamma distribution and identity link function, and (h) GLM with a Gamma distribution and log link function. While the OLS regression model on untransformed cost data has been shown to perform well in predicting healthcare costs,[9, 22] the distribution of error terms for untransformed cost data tends to be non-normal.[9] For these data, the logarithmic transformation resulted in a distribution closest to the normal distribution. Retransformation was used to facilitate interpretation of predicted values on the original cost scale. We adopted: (a) normal retransformation, which is applied when the residuals are normally distributed and homoscedastic,[23] (b) Duan's retransformation,[24] which is used when the residuals are homoscedastic but non-normally distributed, and (c) heteroscedastic retransformation,[25] which is applicable when the residuals are normally distributed but heteroscedastic. Briefly, Baser proposed the following: (a) fit a regression model to the natural

logarithm of cost with covariates and output the residuals and fitted values, (b) fit a regression model to the natural logarithm of the squared residuals with covariates and output the fitted values, (c) exponentiate the fitted values and multiply each by half (i.e., 0.5), and (d) define the heteroscedastic retransformation by exponentiating the summation of the fitted values obtained in (a) and the results from (c). The robust regression model uses a Huber-type estimator.[26]

Predictive performance was evaluated using R^2, mean absolute prediction error (MAPE), and root mean square error (RMSE).[9] To ensure the comparability of the R^2 values between models, we used the predicted costs from each model regressed on observed costs. MAPE is the mean of the absolute differences between predicted and observed values. RMSE is the square root of the mean of the squared differences between the predicted and observed values. Good model performance is indicated by high values for R^2 and low values for the MAPE and RMSE.

Measures of prediction accuracy were based on 10-fold cross-validation. Specifically, the data was randomly apportioned into 10 equal-sized folds, in which 9 folds were designated as the training dataset and used to estimate a regression equation. This equation was used to predict costs for the remaining fold, which was denoted as the out-of-sample prediction fold. Measures of prediction accuracy were then calculated for this out-of-sample fold. This process was repeated 10 times.

Finally, we plotted the mean of the observed and predicted costs by deciles to facilitate visual comparison of each model's prediction accuracy across the entire cost distribution. These plots were based on the out-of-sample data for each of the repetitions in the 10-fold cross-validation. SAS™ version 9.2 was used for all analyses.[27]

5.5 Results

There were a total of 49,791 episodes. Of this number, 1,318 (2.7%) were censored and another 32,242 (60.7%) were second or subsequent episodes. A few episodes (n = 751; 1.5%) that were missing a total cost were excluded; they contained only a single outpatient hospitalization. Hence the total number of first episodes included in the study was 17,480. Table 5-2 describes the characteristics of cohort members with a hospital-initiated episode (n = 7,910) and physician visit-initiated episode (n = 9,570). There were similar percentages of both sexes and income quintiles in the two groups. However, there was a greater percentage of individuals in older age groups, rural residence location, and higher number of comorbid conditions with an episode initiated by hospitalization.

As expected, the cost distribution was positively skewed (Table 5-3). Half of the costs were below $595 CAD, while the 95th percentile was $ 13,934 CAD. This wide variability in cost also reflects the diversity of the contents of episodes of care for COPD. While some episodes could be composed of only few physician visits, others could extend across periods of extended hospitalization and physician visits.

The characteristics of the total cost distribution in each of the 10 equal-sized folds are reported in Table 5-4. Generally, the total cost distribution was fairly similar except for fold number 3, which had slightly higher mean cost and standard deviation.

The 10-fold cross-validation results are reported in Table 5-5. Based on the averages from the 10 replications, the OLS model on log-transformed costs with the normal retransformation had the highest R^2 (18.77%) and the GLM model with a Gamma distribution and log link function had the lowest (14.52%). The robust regression model produced the smallest MAPE ($2,164.72 CAD), while the OLS model on the log-transformed cost with Duan's

retransformation had the highest ($9,213.42 CAD). The linear OLS regression model on the actual costs had the best prediction error based on the RMSE ($8,185.23 CAD), while the linear OLS model on log-transformed cost with Duan's retransformation resulted in the largest value of $15,430.89 CAD. However, model performance differed across the 10 replications. For instance, the linear OLS regression on log-transformed cost with the normal retransformation had the lowest R^2 (6.73%) in the third repetition while it had the highest R^2 (25.30%) in the fifth repetition (Table 5-5).

Figure 5-1 depicts out-of-sample prediction accuracy for each of the models in each of the 10 replications. A consistent pattern of results was observed. Specifically, most of the predicted values were close to the observed cost (actual cost) up to the fifth decile. However, predictive performance of each model differed substantially in the upper deciles of the cost distribution.

5.6 Discussion

Episode-based reimbursement schemes are receiving favourable evaluations for the development of patient-oriented healthcare systems, but the methods to facilitate implementation of these schemes are still under development. Predictive models are used to establish reward systems and evidence-based allocation of funds to providers based on the attributes of their patients. In this study we assessed the prediction accuracy of different models for forecasting episodes of care costs for COPD exacerbations.

Our study showed that the OLS linear regression on the log-transformed cost (with the normal retransformation) and untransformed costs had the best prediction error based on the R^2 and RMSE respectively. Other researchers also found that the OLS linear regression model on the untransformed costs had superior predictive performance compared to other regression models.[20] Examination of residuals in our log-transformed OLS models across the 10 repetitions

revealed fairly normal distributions. This may have accounted for the better performance of the normal retransformation compared with the Duan's and the heteroscedatic retransformations for the R^2 and RMSE. The out-of-range values from Duan's retransformation for both the MAPE and RMSE could be attributed to the fairly normal distribution of the residuals of the log-transformed OLS model. Duan's retransformation will perform better when the residuals are non-normal but homoscesdastic.[24] The robust regression model was best for predicting episodes of care cost for COPD based on the MAPE. Although the robust regression is not often included in the comparators in studies comparing model performance for predicting healthcare cost (see for example Mihaylova et al.[28]), the findings from this study suggest this model is a potential candidate for episodes of care costs.

This study has some limitations. Emergency department costs were not available for the research. Although their exclusion may underestimate the total cost, we also recognized that exacerbations first treated in an emergency department and then admitted to hospital will be included in the hospitalization cost component. We used proxies, such as per diem rates and charges, for some components of the total episodes of care cost for COPD exacerbations. However, hospitalization, which was the major cost components of episode costs, was based on a standard methodology developed by CIHI to reflect actual resource utilization. Population-based studies that use health administrative databases to construct episodes of care often face similar challenges in measuring costs. In spite of this, the episodes of care approach provides a useful framework to build cost prediction models for designing alternative reimbursement schemes that reward care coordination when compared with alternative methods that analyze related health services costs components in isolation.

This study suggests a number of opportunities for future research. Joint models of episode cost components might be adopted for building predictive models of episodes of care cost.[29, 30] Other semi-parametric approaches such as the finite mixture models (FMM) could also be used to model healthcare data.[31] The FMM may be particularly useful given the possible differences that might exist between episodes composed primarily of hospitalizations and those composed primarily of physician visits.

5.7 Conclusions

We compared the performance of eight regression models for predicting costs of episodes of care for COPD. The 10-fold cross-validation revealed that none of the models consistently resulted in the best prediction; the OLS regression model on log-transformed costs with normal retransformation had the highest average R^2, the OLS model on untransformed costs had the lowest average root mean square error and the robust regression model had the lowest average mean absolute prediction error. Based on these findings, we recommend that researchers consider adopting one of these three models for predicting costs of healthcare use in episodes of care. However, we emphasize that further research is needed to explore the relative performance of these models.

5.8 References

1. Averill RF, Goldfield NI, Hughes JS, Eisenhandler J, Vertrees JC. Developing a prospective payment system based on episodes of care. *J Ambul Care Manage*. 2009;32: 241-251.

2. Davis K. Paying for care episodes and care coordination. *N Engl J Med*. 2007;356:1166-1168.

3. Foster EM, Xuan F. An episode-based framework for analyzing healthcare expenditures: An application of reward renewal models. *Health Serv Res*. 2005;40:1953-1971.

4. Scitovsky AA. Changes in the costs of treatment of selected illnesses, 1951-1965. *American Economic Review.* 1967;57:1182-95.

5. Austin PC, Rothwell DM, Tu JV. A comparison of statistical modeling strategies for analyzing length of stay after CABG surgery. *Health Serv Outcomes Res Method*. 2002;3:107-133.

6. Dodd S, Bassi A, Bodger K, Williamson P. A comparison of multivariable regression models to analyse cost data. *J Eval Clin Pract*. 2006;12:76-86.

7. Editorial Board for Respiratory Disease in Canada: Respiratory disease in Canada. Health Canada. Ottawa, ON. 2001. Available at: http://www.phac-aspc.gc.ca/publicat/rdc-mrc01/pdf/rdc0901e.pdf. Accessed 20 June 2012.

8. Murray CJ, Lopez AD. Alternative projection of mortality and disability by cause 1990 – 2020: global burden of disease study. *Lancet.* 1997;349:1498-504.

9. Dunn G, Mirandola M, Amaddeo F, Tansella M. Describing, explaining or predicting mental healthcare costs: a guide to regression models. Methodological review. *Br J Psychiatry*. 2003;183:398-404.

10. Statistics Canada: Population and dwelling counts for Canada, provinces and territories, 2011 and 2006 censuses. Ottawa, ON. 2012. Available at: http://www12.statcan.gc.ca/census-recensement/2011/dp-pd/hlt-fst/pd-pl/Table-Tableau.cfm?LANG=Eng&T=101&S=50&O=A. Accessed 14 July 2012.

11. Downey W, Beck P, McNutt M, et al. Health databases in Saskatchewan. In: Strom B. (eds.) Pharmacoepidemiology, pp 325-345. Wiley, Chinchester (2000).

12. Camp PG, Levy RD. A snapshot of chronic obstructive pulmonary disease in British Columbia and Canada. *BCMJ*. 2008;50:80-84.

13. Canadian Institute for Health Information: Health indicators: Definitions, data sources and rationale. Ottawa, ON. 2012. Available at: http://www.cihi.ca/CIHI-ext-portal/pdf/internet/DEFINITIONS_062011_EN. Accessed 20 October 2012.

14. Canadian Institute for Health Information: Canadian MIS database - Hospital financial performance indicators, 1999/2000 to 2009/2010. Ottawa, ON. 2012. Availabe at: https://secure.cihi.ca/free_products/CMDB_hospita_financial_performance_indicators_meth_notes_2011_en.pdf. Accessed 15 August 2012.

15. Hopkins RB, Tarride JE, Leslie WD. Estimating the excess costs for patients with incident fractures, prevalent fractures, and nonfracture osteoporosis. *Osteoporos Int*. 2012;24:581-93.

16. Statistics Canada: Comsumer price index 2001/02-2009/10. Ottawa, ON. 2012. Available at: http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/econ161a-eng.htm. Accessed 22 June 2012.

17. Roos NP, Mustard CA. Variation in health and healthcare use by socio-economic status in Winnipeg, Canada: the system works well? Yes and no. *Milbank Q.* 1997;75:89-111.

18. 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;40:373-383.

19. McCullagh P, Nelder JA. *Generalized linear models.* Chapman and Hall, London (1989).

20. Diehr P, Yanez D, Ash A, Hornbrook M, Lin DY. Methods for analysing healthcare utilization and cost. *Annu Rev Public Health*. 1999;20:125-144.

21. Larimer ME, Malone DK, Garner MD. Healthcare and public service use and costs before and after provision of housing for chronically homeless persons with severe alcohol problems. *JAMA.* 2009;301:1349-1357.

22. Kilian R, Matschinger H, Löeffler W, Roick C, Angermeyer MC. A comparison of methods to handle skew distributed cost variables in the analysis of the resource consumption in schizophrenia treatment. *J Ment Health Policy Econ.* 2002;5:21-31.

23. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? *J Health Econ*. 2001;20:461-494.

24. Duan N. Smearing estimate: a non-parametric retransformation method. *J Am Stat Assoc*. 1983;78:605-10.

25. Baser O. Modeling Transformed healthcare cost with unknown heteroskedasticity. *App. Econ. Res. Bull*. 2007;01:1-6.

26. Huber PJ. Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann Stat*. 1973;1:799-821.

27. SAS Institute Inc.: SAS/STAT user's guide. Cary, NC: SAS Institute Inc. (2004).

28. Mihaylova B, Briggs A, O'Hagan A, Thompson SG. Review of statistical methods for analysing healthcare resources and costs. *Health Econ*. 2011;20:897-916.

29. Gebregziabher M, Zhao Y, Dismuke CE, et al. Joint modeling of multiple longitudinal cost outcomes using multivariate generalized linear mixed models. *Health Serv Outcomes Res Method*. 2012;13:39-57.

30. Daggy JK, Thomas J, Craig BA. Modeling correlated healthcare costs. *Expert. Rev. Pharmacoecon Outcomes Res*. 2011;11:101-111.

31. Deb P, Holmes AM. Estimates of use and costs of behavioural healthcare: a comparison of standard and finite mixture models. *Health Econ*. 2000;9:475-489.

Table 5-1: ICD codes and decision rules to define episodes of care for chronic obstructive pulmonary disease exacerbations

Hospital-initiated episode	Physician visit-initiated episode
Any of these ICD-10-CA codes in the most responsible diagnosis field for hospital separations or ICD-9 codes in the single diagnosis field in physician billing claims	
ICD-10-CA	**ICD-9**
J41 – Simple and mucopurulent chronic bronchitis	491 – Chronic bronchitis
J42 – Unspecified chronic bronchitis	492 – Emphysema
J43 – Emphysema	
J44 – Other chronic obstructive pulmonary disease	496 – Chronic airway obstruction
OR	
Any of the following ICD-10-CA codes in the most responsible diagnostic field for hospital separations or ICD-9 codes in the single diagnosis field in physician billing claims	
J10.0 - Influenza with pneumonia, influenza virus identified	466 - Acute bronchitis and bronchiolitis
J11.0 - Influenza with pneumonia, virus not identified	480 - Viral pneumonia
J12 - Viral pneumonia, not elsewhere classified	481 - Pneumococcal pneumonia
J13 - Pneumonia due to Streptococcus pneumoniae	482 - Other bacterial pneumonia
J14 - Pneumonia due to Haemophilus influenzae	483 - Pneumonia due to other specified organism
J15 - Bacterial pneumonia, not elsewhere classified	484 - Pneumonia in infectious diseases classified elsewhere
J16 - Pneumonia due to other infectious organisms, not elsewhere classified	485 - Bronchopneumonia, organism unspecified
J18 - Pneumonia, organism unspecified	486 - Pneumonia, organism unspecified
J20 - Acute bronchitis	487 - Influenza w/ pneumonia
J21- Acute bronchiolitis	

J22 - Unspecified acute lower respiratory infection		
	PLUS	
J44 in another diagnosis field		At least one dispensation of any of the following prescription drugs: SABA SABA/Anticholinergic Antibiotic Systemic corticosteroid

Note: SABA = short acting beta agonists

Table 5-2: Frequency (%) of cohort members with hospital-initiated and physician visit-initiated episodes of care

Variable	Hospital-Initiated *n* = 7910	Physician Visit-Initiated *n* = 9570
Age (years)		
35 –44	174.0 (2.2)	421.1 (4.4)
45 – 59	1028.3 (13.0)	2076.7 (21.7)
60 – 74	3084.9 (39.0)	4249.1 (44.4)
75 – 84	2499. 6 (31.7)	2201.1 (23.0)
≥ 85	1123.2 (14.2)	622.1 (6.5)
Income quintile		
Q1 (Lowest)	2112.0 (26.7)	2143.7 (22.4)
Q2	1708. 6 (21.6)	2153.3 (22.5)
Q3	1945.9 (24.6)	2306.4 (24.2)
Q4	1241.9 (15.7)	1799.2 (18.8)
Q5 (Highest)	901.7 (11.4)	1167.5 (12.2)
Charlson score		
≤ 1	4097.4 (51.8)	7091.4 (74.1)
≥ 2	3812.6 (48.2)	2478.6 (25.9)
Gender		
Male	4160.7 (52.6)	4976.4 (52.0)
Female	3749.3 (47.4)	4593.6 (48.0)
Patient survival in episode		
Alive	7087.4 (89.6)	9321.2 (97.4)
Dead	822.6 (10.4)	248.8 (2.6)
Recency of COPD diagnosis		
Diagnosis after April 1, 2001	6264.7 (79.2)	6545.9 (68.4)
Diagnosis before April 1, 2001	1645.3 (20.8)	3024.1 (31.6)
Area of residence		
Urban	3709.8 (46.9)	4919.0 (51.4)
Rural	4200.2 (53.1)	4651.0 (48.6)

Table 5-3: Characteristics of COPD episode costs and transformations

Variable	Mean	SD	25^{th} percentile	Median	75^{th} percentile	95^{th} percentile	γ_1	γ_2
Cost	3835	9211	56	595	4603	13,934	13.06	311.85
Log-cost	6.32	2.47	4.04	6.39	8.43	9.54	-0.18	-1.41
$Cost^{0.5}$	43.82	43.76	7.45	24.39	67.85	118.04	1.95	9.34
$-(Cost^{-0.5})$	-0.08	0.09	-0.13	-0.04	-0.01	-0.01	-1.27	0.78

Note: SD = standard deviation; γ_1 = skewness; γ_2 = kurtosis; Costs are reported in CAD constant dollars for 2002/03

Table 5-4: Distributional characteristics of COPD episode costs in the ten folds for cross-fold validation

	Fold Number									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Mean	3977.98	3889.4	4242.85	3984.61	3613.57	3785.27	3682.66	3691.5	3609.13	3875.96
SD	8836.65	10405.72	12746.43	10293.26	7214.65	7781.64	7482.33	7855.43	7977.57	9985.92
25^{th} percentile	59.8	56.19	58.59	61.27	56.23	56.19	47.6	51.28	52.57	61.53
Median	1383.35	535.32	751.8	733.49	445.18	423.37	531.01	730.38	504.54	983.37
75^{th} percentile	4634.92	4635.64	4571.26	4724.38	4502.93	4547.52	4663.28	4512.23	4509.96	4761.76
95^{th} percentile	14435.69	14420.13	15746.33	14236.31	14142.18	14637.88	13821.43	13110.49	12157.25	13295.75

Note: SD = standard deviation; each of the 10 random samples had N = 1,748; costs are reported in CAD constant dollars for 2002/03

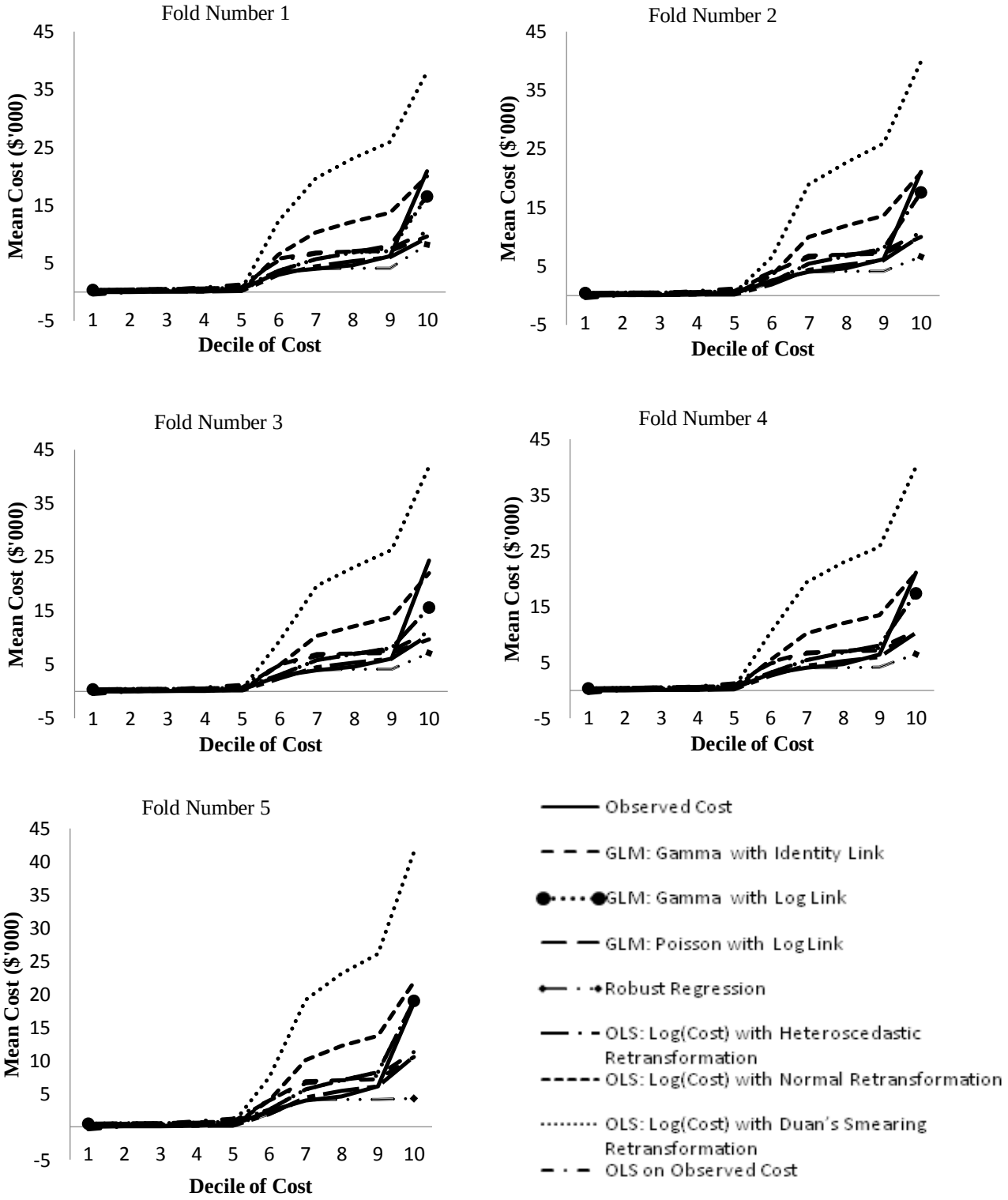
Table 5-5: Measures of prediction accuracy for COPD episode cost prediction models from ten-fold cross-validation

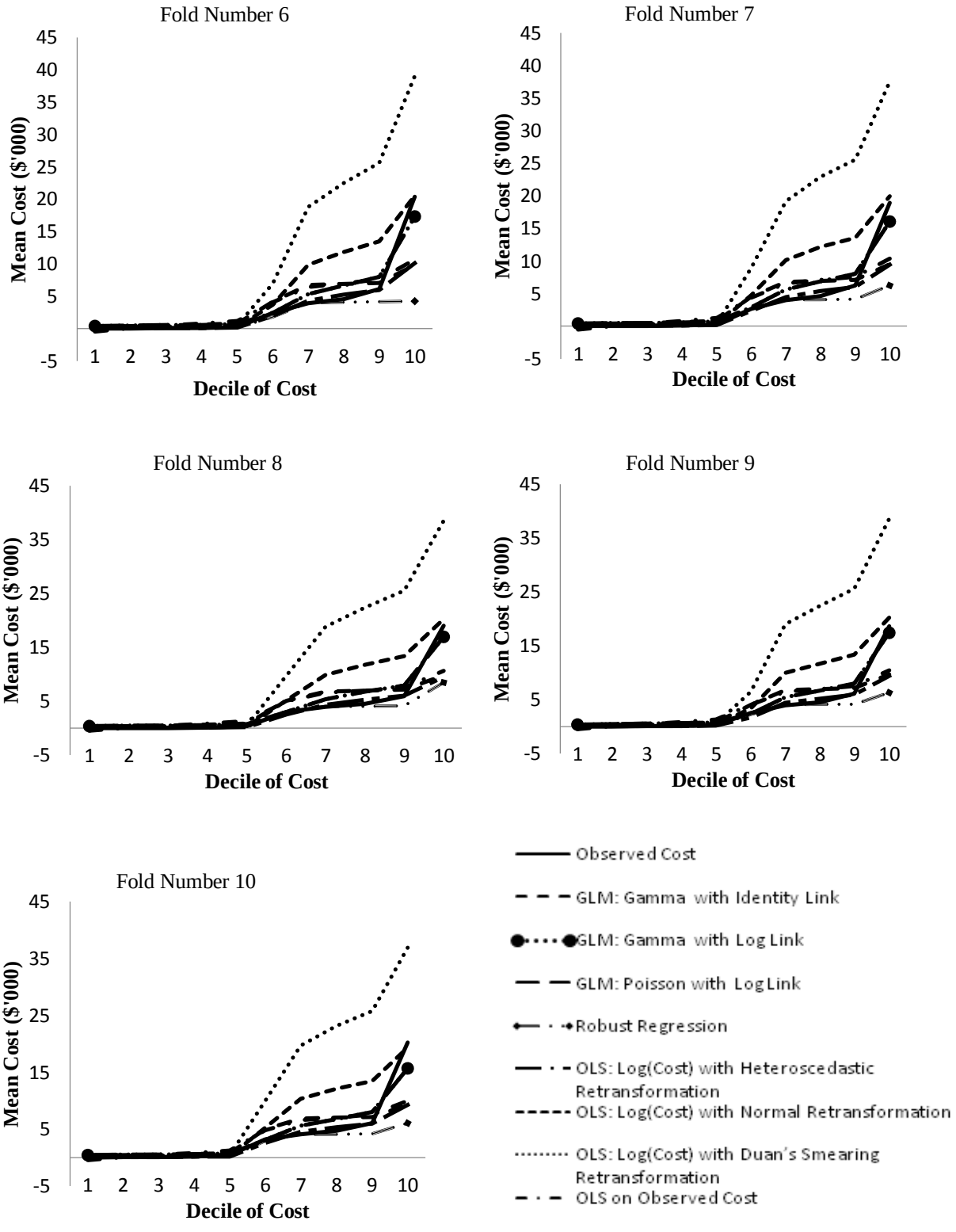
	Replication Number										
Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Average**
						R^2 **(%)**					
1	18.82	16.81	6.73	15.44	25.30	21.95	24.76	19.19	21.69	17.01	**18.77**
2	18.81	16.42	6.85	15.12	24.30	21.15	24.71	19.92	21.02	16.95	18.53
3	19.83	18.01	17.99	17.94	18.01	17.98	18.02	17.97	18.01	18.02	18.18
4	18.45	16.71	7.07	14.03	23.48	19.93	23.92	18.79	21.83	16.90	18.11
5	16.50	17.57	17.57	17.57	17.58	17.58	17.58	17.58	17.58	17.57	17.47
6	18.12	16.96	16.72	16.93	16.93	16.93	16.96	16.97	16.98	16.94	17.04
7	15.56	16.90	16.88	16.88	14.62	14.65	16.88	15.49	16.89	16.89	16.16
8	14.19	14.65	15.71	14.53	14.17	14.37	14.43	14.32	14.18	14.60	14.52
						MAPE ($ CAD)					
1	4472.99	4264.69	4973.47	4421.12	4216.77	4264.73	4142.1	4267.69	4101.94	4310.93	4343.64
2	9469.96	9040.01	10018.41	9378.61	9190.99	8898.68	8875.67	9171.1	8903.71	9187.08	9213.42
3	3049.33	2955.94	3333.52	3047.33	2970.78	3003.8	2798.65	2951.24	2830.13	2821.77	2976.25
4	2265.91	2285.43	4244.32	3986.12	3615.02	3786.77	3684.11	3692.97	3610.6	3877.56	3504.88
5	2903.19	2837.74	3238.81	2885.2	2625.55	2780.07	2671.51	2753.89	2691.23	2796.38	2818.36
6	2799.28	2719.32	3128.42	2773.86	2519.5	2668.05	2528.96	2630.11	2541.04	2693.73	2700.23
7	2300.43	2225.05	2549.66	2244.58	1961.57	2136.12	2010.53	2102.73	1965.94	2150.57	**2164.72**
8	3049.61	2956.29	3333.87	3047.6	2971.12	3004.91	2799.95	2951.53	2830.43	2822.01	2976.73
						RMSE ($ CAD)					
1	8986.40	10244.70	13242.25	10331.87	7898.65	8141.99	7799.62	8394.85	8245.69	9840.20	9312.62
2	15095.08	15756.12	18531.92	16056.07	15098.56	14628.76	14280.84	14899.01	14645.59	15316.92	15430.89
3	8471.30	9750.76	12519.28	9812.24	7129.41	7630.26	7051.32	7777.69	7569.74	9172.01	8688.40
4	8064.17	9577.70	13431.25	11035.73	8068.28	8652.58	8338.72	8678.61	8754.96	10709.97	9531.20
5	7949.45	9520.96	12168.88	9416.15	6226.29	6823.45	6512.42	7044.13	7077.07	9113.51	**8185.23**

6	7992.39	9564.30	12276.30	9472.79	6251.72	6840.75	6509.27	7072.54	7063.06	9195.87	8223.90
7	8259.05	9842.26	12443.98	9738.47	6723.02	7338.71	6827.54	7325.69	7329.48	9467.91	8529.61
8	8471.84	9751.22	12519.58	9812.74	7130.18	7632.98	7054.15	7778.26	7570.38	9172.32	8689.36

Note: Model 1= linear OLS regression on log (cost) with normal retransformation; Model 2 = linear OLS regression on log (cost) with Duan's retransformation; Model 3 = generalized linear model (GLM) with Poisson distribution and log link function; Model 4= linear OLS regression on log (cost) with heteroscedastic retransformation; Model 5 = linear OLS regression on actual cost; Model 6 = GLM with Gamma distribution and identity link function; Model 7= robust regression; Model 8 = GLM with Gamma distribution and log link function. Highest R^2 values and lowest MAPE and RMSE values represent the best models. The best model on each performance indicator is shown in the last column in boldface font. MAPE = Mean absolute prediction error; RMSE = Root mean square error.

Figure 5-1: Observed and predicted COPD episode costs by deciles for ten-fold cross-validation




Fold Number 6
Fold Number 7
Fold Number 8
Fold Number 9
Fold Number 10
Mean Cost ($'000)
Decile of Cost
45
40
35
30
25
20
15
10
5
-5
1 2 3 4 5 6 7 8 9 10
Observed Cost
GLM: Gamma with Identity Link
GLM: Gamma with Log Link
GLM: Poisson with Log Link
Robust Regression
OLS: Log(Cost) with Heteroscedastic Retransformation
OLS: Log(Cost) with Normal Retransformation
OLS: Log(Cost) with Duan's Smearing Retransformation
OLS on Observed Cost

CHAPTER 6: SUMMARY

6.1 Summary of findings

The overall goal of this research was to construct an episode of care data system and use it to describe and predict healthcare resource use and cost; with a focus on chronic obstructive pulmonary disease (COPD) exacerbations. To achieve this research goal, we undertook four related studies. The first study was focused on data quality evaluation whilst the other three dealt with different aspects of healthcare utilization and costs. The following paragraphs describe the key findings from each of these studies.

Although emergency departments (EDs) may play important roles in managing acute exacerbations of COPD, previous studies which assessed the accuracy of Saskatchewan's administrative health databases did not include the ED data. Given that a number of patients who visit EDs for COPD exacerbations may later be hospitalized during the same encounter, we investigated the agreement between ED and hospital discharge abstracts for capturing transitions from the emergency to the acute care settings (Chapter 2). The results indicate that the discharge dispositions recorded in the ED data is not a reliable source of patient transition information to inpatient acute care, particularly for patient who were transferred to different hospitals. Therefore, studies of patient care trajectories that describe transitions from the ED to acute care should not rely on the discharge disposition field recorded in the ED data, but rather be conducted by linking patient-specific records across the two care settings.

We linked the physician billing claims, ED data, hospital discharge abstracts, outpatient drugs data, home care and long-term care databases to construct an episode of care data system. This enabled patients' entire journeys through the primary, emergency, acute and publicly funded supportive care settings to be captured.

Information about healthcare pathways is important for identifying opportunities to redesign healthcare delivery to improve quality and reduce costs. In Chapter 3, using the episode of care data system, LCA was applied to multiple indicators of healthcare use to identify COPD patient groups with distinct healthcare pathways. These pathways were associated with patient demographic and comorbidity characteristics, as well as with healthcare costs. We found a few individuals who followed complex pathways, in which multiple healthcare services were repeatedly used. As expected, these individuals also incurred substantially higher costs than individuals with less complex pathways.

Building on the previous study which identified high-cost patients, the study presented in Chapter 4 examined whether high-cost patients in one episode also incurred high costs in subsequent episodes. We found only a few patients who persistently incurred high costs in the index and subsequent episodes. The findings also show that patient demographic and disease characteristics, as well as healthcare encounter history, all of which are routinely collected in administrative data, could satisfactorily predict patients who become persistent high users.

In Chapter 5, we compared the performance of eight alternative linear regression models for predicting patient's actual costs of episodes of COPD exacerbations. The results were validated using 10-fold cross-validation approach. This study reveals that none of the eight models consistently resulted in the best prediction. The ordinary least squares (OLS) regression model on log-transformed costs with normal retransformation had the highest average R^2, the OLS model on untransformed costs had the lowest average root mean square error and the robust regression model had the lowest average mean absolute prediction error.

6.2 Implications and further research

Many healthcare jurisdictions have proposed to use episode of care as a basis for provider reimbursements and performance measurement.[1] However, many of these proposals are still in their conceptual stages; with little empirical work available to guide their design.[2] One of the reasons of the slow implementation of episode of care research is the complexity involved in constructing a data system. For example, a decision has to be made on whether to construct the episode for a single health condition or a set of related conditions, and how this choice would affect the definition of an episode's start, duration, and end.[3] In this thesis, we demonstrated how population-based administrative health databases could be linked to construct an episode of care data system for a chronic health condition. The methodology can be adopted in other jurisdictions that house repositories of population-based administrative health databases.

With the episode of care data system in place, a variety of topics relevant to the implementation of new performance measures [4] and payment schemes [5] can be investigated. For instance, Peltola et al.[4] used episode of care approach to implement health system-wide resource utilization measures to profile care providers in Finland. Also, a previous study [6] linked various administrative databases to examine the routes to cancer diagnosis and survival in England. In Canada, the Health Quality Ontario is developing a methodology referred to as quality-based procedures (QBPs), which will reimburse healthcare providers using evidence-informed rates that are associated with the quality of care delivered (www.hqontario.ca/evidence/evidence-process/episodes-of-care). The episode of care has been adopted as a unit of analysis in developing the QBPs for selected health conditions including COPD.

In this research, we focused on the methodology of defining healthcare pathways through episodes of care, as well as estimating and predicting episode-based costs. Although information

on patient care pathways is important for anticipating patients' needs and allocating resources to meet them, there has not been any standard method of defining them. Previous research [7] has adopted a deterministic approach, using predefined categories of healthcare contacts, to describe care pathways. However, this approach may result in an inefficiently large number of categories because patients often exhibit heterogeneity in their use of healthcare services.[8] We have shown that latent class analysis, a model-based approach, can be used to identify care pathways for patients with complex health conditions such as COPD. Our method allows for describing outcomes associated with pathways as well as examining the exact sequence in which patients contact care providers. This modeling approach could be applied to other resource intensive and complex conditions, such as diabetes and hypertension.

Again, because an episode of care usually spans multiple healthcare settings (i.e., hospitals, emergency departments, physician visits, prescription dispensations, etc.) an episode cost is an aggregate of costs associated with these services. However, the characteristics of costs of these individuals services differ substantially in terms of their frequency and magnitude.[9] Risk factors may also relate to these cost components differently than to total episode costs. Consequently, existing cost prediction models or risk-adjustment models designed for single services such as hospitalization or prescription drugs may not adequately predict episode-based costs. Our study on the comparison of statistical models for predicting episode costs found that the robust regression model, a model not often considered for cost prediction, was among the best models for prediction accuracy.

Aside from the methodological contributions discussed above, the studies included in this thesis also have direct implications for health policy. Specifically, the studies were conducted in collaboration with researchers from Saskatchewan Health Quality Council (HQC). HQC is an

independent agency that is legislated to measure and report on the quality of healthcare in Saskatchewan, as well as recommend and support innovative ways to improve the quality of care provided in the province's healthcare system. This collaboration ensured that study outcomes/findings remain relevant to the needs of policy makers in the province. Particularly, our care pathways analysis showed the critical link between care processes and outcomes; potentially providing an efficient tool for measuring quality of care. As already mentioned, Ontario is currently designing episode-based bundled payment schemes for selected health conditions including COPD. Given that accurate prediction of episode costs is important for fairly reimbursing providers, the cost prediction models presented in this thesis could complement the studies being conducted in Ontario.

The following are some areas for future research. First, our episode of care data construction relied only on administrative health databases. Although these databases have rich information on patients' healthcare utilization and costs, they do not have detailed clinical and functional information as electronic medical records (EMRs).[10] Future studies can combine EMRs and administrative health data to construct episodes of care. This, for example, will allow for measuring disease severity, and will be useful for enhancing accuracy of prediction models.[11] Particularly, variables such as forced expiratory volume in one second (FEV_1) and forced vital capacity (FVC) are useful for measuring COPD severity. Future studies should include these variables to be able to tease out the effects of disease severity on utilization and costs.[12]

Second, the US-based Patient-Centered Outcomes Research Institute (http://www.pcori.org/) is currently promoting patient-centered cost-effectiveness studies. These studies usually compare the costs and outcomes associated with alternative healthcare interventions/policies. Future studies could conduct cost-effectiveness studies using the episode of care as the unit of analysis.

For example, we found that a small percentage of COPD patients persistently incur high costs, and might be suitable candidates for interventions. To identify the most appropriate intervention, cost-effectiveness studies could be conducted using the episode of care as the unit of analysis.

References

1. Davis K. Paying for care episodes and care coordination. *N Engl J Med*. 2007;356(11):1166–8.

2. Hussey PS, Sorbero ME, Mehrotra A, Liu H, Damberg CL. Episode-based performance measurement and payment: making it a reality. *Health Aff (Millwood)*. 2009;28(5):1406–17.

3. Hornbrook MC, Hurtado AV, Johnson RE. Health care episodes: definition, measurement and use. *Med Care Res Rev*. 1985;42(2):163–218.

4. Peltola M, Juntunen M, Häkkinen U, et al. A methodological approach for register-based evaluation of cost and outcomes in health care. *Ann Med*. 2011;43 Suppl 1(January):S4–13.

5. Averill RF, Goldfield NI, Hughes JS, Eisenhandler J, Vertrees JC. Developing a prospective payment system based on episodes of care. *J Ambul Care Manage*. 2009;32(3):241–51.

6. Elliss-Brookes L, McPhail S, Ives A, et al. Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. *Br J Cancer*. 2012;107(8):1220–6.

7. Dobson A, DaVanzo JE, El-Gamil A, et al. Clinically appropriate and cost-effective placement study. *Final Report by Dobson DaVanzo & Associates, LLC*. 2012. Available at: http://ahhqi.org/images/pdf/cacep-report.pdf. Accessed December 22, 2014.

8. Lafortune L, Béland F, Bergman H, Ankri J. Health state profiles and service utilization in community-living elderly. *Med Care*. 2009;47(3):286–294.

9. Gebregziabher M, Zhao Y, Dismuke CE, et al. Joint modeling of multiple longitudinal cost outcomes using multivariate generalized linear mixed models. *Heal Serv Outcomes Res Methodol*. 2012;13(1):39–57.

10. Yu AP, Yang H, Wu EQ, et al. Incremental third-party costs associated with COPD exacerbations: a retrospective claims analysis. *J Med Econ*. 2011;14(3):315–323.

11. Maciejewski ML, Liu C-F, Fihn SD. The performance of comorbidity, risk adjustment, and functional status measures in expenditure prediction for patients with diabetes. *Diabetes Care*. 2009;32(1):3–8.

12. Make BJ, Eriksson G, Calverley PM, et al. A score to predict short-term risk of COPD exacerbations (SCOPEX). *Int J Copd*. 2015;10:201–209.

APPENDIX A: ETHICS APPROVAL FROM UNIVERSITY OF MANITOBA

UNIVERSITY OF MANITOBA | BANNATYNE CAMPUS Research Ethics Board

P126 - 770 Bannatyne Avenue
Winnipeg, Manitoba
Canada R3E 0W3
Telephone 204-789-3255
Fax 204-789-3414

HEALTH RESEARCH ETHICS BOARD (HREB)
CERTIFICATE OF FINAL APPROVAL FOR NEW STUDIES
Delegated Review

PRINCIPAL INVESTIGATOR:	INSTITUTION/DEPARTMENT:	ETHICS #:
Mr. John Paul Kuwornu	U of M /Community Health Sciences and Saskchewan Health Quality Council	HS 18948 (H2015:365):
APPROVAL DATE: September 15, 2015	**EXPIRY DATE:** **September 15, 2016**	
STUDENT PRINCIPAL INVESTIGATOR SUPERVISOR (If applicable): Dr. Lisa Lix		

PROTOCOL NUMBER:	PROJECT OR PROTOCOL TITLE; Visioning with Health Administrative Data (V-HAD): A Partnership for the Development and Use of an Episodes-of-Care Data System
SPONSORING AGENCIES AND/OR COORDINATING GROUPS: Saskatchewan Health Quality Council - Grant and U of Saskatchewan Centennial Research Chair Program	

Submission Date of Investigator Documents:	HREB Receipt Date of Documents:
September 8, 2015	September 8, 2015

THE FOLLOWING ARE APPROVED FOR USE:

Document Name	Version(if applicable)	Date
Protocol: Proposal		September 8, 2015
Consent and Assent Form(s):		
Other: List of Variables		September 8, 2015

CERTIFICATION
The above named research study/project has been reviewed in a ***delegated manner*** by the University of Manitoba (UM) Health Research Board (HREB) and was found to be acceptable on ethical grounds for research involving human participants. The study/project and documents listed above was granted final approval by the Chair or Acting Chair, UM HREB.

HREB ATTESTATION
The University of Manitoba (UM) Research Board (HREB) is organized and operates according to Health Canada/ICH Good Clinical Practices, Tri-Council Policy Statement 2, and the applicable laws and regulations of Manitoba. In respect to clinical trials, the HREB complies with the membership requirements for Research Ethics Boards defined in Division 5 of the Food and Drug Regulations of Canada and carries out its functions in a manner consistent with Good Clinical Practices.

QUALITY ASSURANCE
The University of Manitoba Research Quality Management Office may request to review research documentation from this research study/project to demonstrate compliance with this approved protocol and the University of Manitoba Policy on the Ethics of Research Involving Humans.

CONDITIONS OF APPROVAL:

1. The study is acceptable on scientific and ethical grounds for the ethics of human use only. ***For logistics of performing the study, approval must be sought from the relevant institution(s).***
2. This research study/project is to be conducted by the local principal investigator listed on this certificate of approval.
3. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to the research study/project, and for ensuring that the authorized research is carried out according to governing law.
4. **This approval is valid until the expiry date noted on this certificate of approval. A Bannatyne Campus Annual Study Status Report** must be submitted to the HREB within 15-30 days of this expiry date.
5. Any changes of the protocol (including recruitment procedures, etc.), informed consent form(s) or documents must be reported to the HREB for consideration in advance of implementation of such changes on the **Bannatyne Campus Research Amendment Form.**
6. Adverse events and unanticipated problems must be reported to the HREB as per Bannatyne Campus Research Boards Standard Operating procedures.
7. The UM HREB must be notified regarding discontinuation or study/project closure on the **Bannatyne Campus Final Study Status Report**.

Sincerely,

John Arnett, PhD. C. Psych.
Chair, Health Research Ethics Board
Bannatyne Campus

APPENDIX B: ETHICS APPROVAL FROM UNIVERSITY OF SASKATCHEWAN

UNIVERSITY OF SASKATCHEWAN

Biomedical Research Ethics Board (Bio-REB)

Certificate of Re-Approval

PRINCIPAL INVESTIGATOR	DEPARTMENT	Bio #
Gary Teare	Health Quality Council	11-35

INSTITUTION(S) WHERE RESEARCH WILL BE CARRIED OUT
Health Quality Council
Atrium Building, Innovation Place
241 - 111 Research Dr.
Saskatoon SK S7N 2R2

SUB-INVESTIGATOR(S)
Lisa Lix, Jacqueline Quail

STUDENT RESEARCHER(S)
Paul Kuwornu

FUNDER(S)
CANADIAN INSTITUTES OF HEALTH RESEARCH (CIHR)

TITLE
Visioning with Health Administrative Data (V-HAD): A Partnership for the Development and Use of an Episodes-of-Care Data System

RE-APPROVED ON	EXPIRY DATE
13-Jan-2015	12-Jan-2016

Delegated Review ☒ Full Board Meeting ☐

CERTIFICATION
The study is acceptable on scientific and ethical grounds. The principal investigator has the responsibility for any other administrative or regulatory approvals that may pertain to this research study, and for ensuring that the authorized research is carried out according to governing law. This re-approval is valid for the specified period provided there is no change to the approved protocol or consent process.

FIRST TIME REVIEW AND CONTINUING APPROVAL
The University of Saskatchewan Biomedical Research Ethics Board reviews above minimal studies at a full-board (face-to-face meeting. Any research classified as minimal risk is reviewed through the delegated (subcommittee) review process. The initial Certificate of Approval includes the approval period the REB has assigned to a study. The Status Report form must be submitted within one month prior to the assigned expiry date. The researcher shall indicate to the REB any specific requirements of the sponsoring organizations (e.g. requirement for full-board review and approval) for the continuing review process deemed necessary for that project. For more information visit http://www.usask.ca/research/ethics_review/.

REB ATTESTATION
In respect to clinical trials, the University of Saskatchewan Research Ethics Board complies with the membership requirements for Research Ethics Boards defined in Part 4 of the Natural Health Products Regulations and Division 5 of the Food and Drug Regulations and carries out its functions in a manner consistent with Good Clinical Practices. This re-approval and the views of this REB have been documented in writing. The University of Saskatchewan Biomedical Research Ethics Board has been approved by the Minister of Health, Province of Saskatchewan, to serve as a Research Ethics Board (REB) for research projects involving human subjects under section 29 of The Health Information Protection Act (HIPA).

Ildiko Badea, Chair
University of Saskatchewan
Biomedical Research Ethics Board

Please send all correspondence to:

Research Ethics Office
University of Saskatchewan
Box 5000 RPO University
1607 - 110 Gymnasium Place
Saskatoon, SK Canada S7N 4J8