OUTPATIENT EMERGENCY DEPARTMENT UTILIZATION: MEASUREMENT AND PREDICTION

A Dissertation Presented

By

Lisa M. Lines, MPH

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences,

Worcester, MA

In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSPHY

April 30, 2014

Clinical and Population Health Research

OUTPATIENT EMERGENCY DEPARTMENT UTILIZATION:
MEASUREMENT AND PREDICTION

A Dissertation Presented
By

LISA MARIE LINES

The signatures of the Dissertation Defense Committee signify completion and approval as to style and content of the Dissertation

____
Arlene S. Ash, PhD, Thesis Advisor

____
Allison B. Rosen, MD, MPH, ScD, Member of Committee

____
Amy K. Rosen, PhD, Member of Committee

____
Jerry Cromwell, PhD, Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee

____
Becky A. Briesacher, PhD, Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the school

____
Anthony Carruthers, PhD,
Dean of the Graduate School of Biomedical Sciences

Clinical and Population Health Research Program

April 30, 2014

Dedicated to Chase, my one and only

ACKNOWLEDGEMENTS

This work would not have been possible without all the support I received from so many wonderful people.

I would like to thank my advisor, Arlene Ash for your patience, candid advice, guidance, support, timely feedback, and commitment to supporting me through this process. I am forever grateful! I have learned so much, and you have been an inspiration as well as a great mentor. This work would never have been possible without your thoughtful insights and direction.

To my dissertation committee—Becky Briesacher, Amy Rosen, Allison Rosen, and Jerry Cromwell—my sincere gratitude for your active participation, guidance, and the opportunity to work with you. Your suggestions and advice have been invaluable to me during this process. I am eternally grateful for all of your assistance with this work, from the proposal to this point and beyond. Your expert opinions and thoughtful comments and critiques were instrumental in helping me to complete my dissertation research. I am a truly grateful to you all for your support and guidance throughout this process, and I know I am a better researcher because of your influences.

Special thanks to the world-class faculty of the Clinical and Population Health Research Program for sharing your expertise and advice, and to the faculty and staff of the Quantitative Health Sciences department for including me in your seminars and celebrations and providing a supportive, collegial home base. Thank you to Kate LaPane, Carole Upshur, Catarina Kiefe, and Jeroan Allison for your mentorship and inspiring leadership of CPHR and QHS; to Robin Clark and Allison Rosen for guiding me through my first research rotations in the CPHR program; and to Stephenie Lemon for your guidance through the first year and helpful support with my proposal. Thank you to Terry Field, Pat Franklin, Lee Hargraves, George Reed, Molly Waring, Judy Ockene, Charles Lidz, Beth Dugan, Jeroan Allison, and Stephenie Lemon for your rigorous and thorough instruction and critique during the coursework phase. Your instruction and guidance have provided me with skills I will use in all my future work.

Thank you to Tom Houston for early insight into the possibilities of the Allscripts data; to Ed Boudreaux, who lent support to my AHRQ grant application; to Bruce Barton, Dave Hoaglin, and Hassan Fouayzi for speedy responses to vexing statistical questions; to Wenjun Li, Jeff Chang, Kevin Kane, and Nisha Kini for assistance with geocoding and geospatial analysis; and to all the dedicated grants and administrative staff, especially Deborah Wiggin, Joyce Barrett, James Thompson, Karen Del'Olio, Paula Nims, Margherita Altobelli, and Bernadette Winberg. Thank you all so much.

Special thanks to Edward Westrick, David Fairchild, and Elaine Fontaine for providing me the opportunity to work with the MCN data, and to Joanna Famadas, Alan Krinsky, Tracey Wilkie, and Qiyao Zhang for assistance with pulling all the data sources together.

Special thanks to Katherine Toro, Ron Russell, Randy Ellis, and Zoe Zhao at Verisk Health for providing me the opportunity to work with the MarketScan data, providing insight about the DxCG software and Verisk processes, and helping with the statistical programming.

I am especially grateful to my colleagues at RTI International, who supported me with both a flexible schedule and a professional development award to cover some of my time, without which this process would have been much harder. My fellow RTI researchers were incredibly patient, understanding, and supportive, and for that I am so thankful.

I would also like to express my appreciation to my fellow CPHR students, who helped make this journey enjoyable through interesting conversations, shared experiences, fun events, good friendships, and mutual assistance and encouragement.

And finally, I would like to thank my parents, Mary Lou Veal and Martha Sayles, who have always loved, inspired, and supported me; and my wonderful partner, Chase, for her love, patience, encouragement, and understanding – I could not have done it without you!

ABSTRACT

Approximately half of all emergency department (ED) visits are primary-care sensitive (PCS) – meaning that they could potentially be avoided with timely, effective primary care. Reducing undesirable types of healthcare utilization (including PCS ED use) requires the ability to define, measure, and predict such use in a population.

In this retrospective, observational study, we quantified ED use in 2 privately insured populations and developed ED risk prediction models. One dataset, obtained from a Massachusetts managed-care network (MCN), included data from 2009-11. The second was the MarketScan database, with data from 2007-08. The MCN study included 64,623 individuals enrolled for at least 1 base-year month and 1 prediction-year month in Massachusetts whose primary care provider (PCP) participated in the MCN. The MarketScan study included 15,136,261 individuals enrolled for at least 1 base-year month and 1 prediction-year month in the 50 US states plus DC, Puerto Rico, and the US Virgin Islands.

We used medical claims to identify principal diagnosis codes for ED visits, and scored each according to the New York University Emergency Department algorithm. We defined primary-care sensitive (PCS) ED visits as those in 3 subcategories: nonemergent, emergent but primary-care treatable, and emergent but preventable/avoidable.

We then: 1) defined and described the distributions of 3 ED outcomes: any ED use; number of ED visits; and a new outcome, based on the NYU algorithm, that we call PCS ED use; 2) built and validated predictive models for these outcomes using administrative claims data; 3) compared the performance of models predicting any ED use, number of ED visits, and PCS ED use; 4) enhanced these models by adding enrollee characteristics from electronic medical records, neighborhood characteristics, and

payor/provider characteristics, and explored differences in performance between the original and enhanced models.

In the MarketScan sample, 10.6% of enrollees had at least 1 ED visit, with about half of utilization scored as PCS. For the top risk group (those in the 99.5th percentile), the model's sensitivity was 3.1%, specificity was 99.7%, and positive predictive value (PPV) was 49.7%. The model predicting PCS visits yielded sensitivity of 3.8%, specificity of 99.7%, and PPV of 40.5% for the top risk group.

In the MCN sample, 14.6% (±0.1%) had at least 1 ED visit during the prediction period, with an overall rate of 18.8 (±0.2) visits per 100 persons and 7.6 (±0.1) PCS ED visits per 100 persons. Measuring PCS ED use with a threshold-based approach resulted in many fewer visits counted as PCS, discarding information unnecessarily. Out of 45 practices, 5 to 11 (11-24%) had observed values that were statistically significantly different from their expected values. Models predicting ED utilization using age, sex, race, morbidity, and prior use only (claims-based models) had lower R^2 (ranging from 2.9% to 3.7%) and poorer predictive ability than the enhanced models that also included payor, PCP type and quality, problem list conditions, and covariates from the EMR, Census tract, and MCN provider data (enhanced model R^2 ranged from 4.17% to 5.14%). In adjusted analyses, age, claims-based morbidity score, any ED visit in the base year, asthma, congestive heart failure, depression, tobacco use, and neighborhood poverty were strongly associated with increased risk for all 3 measures (all $P<.001$).

TABLE OF CONTENTS

PREFACE xvi

ABBREVIATIONS xvii

CHAPTER I. INTRODUCTION 1

Background 2

Why Do Patients Go to the Emergency Department? 4

Problems with the Emergency Department from the Patient's Perspective 5

Overview: Measuring Emergency Department Use 6

Primary Care Sensitive Emergency Department Use 7

Concerns about Reducing Emergency Department Use 8

Primary Care Payment Reform 9

Conceptual Framework 10

Literature Review 12

Methods Used to Categorize Primary Care Sensitive Emergency Department Use 18

Diagnosis-based Classification 19

Procedure-based Classification 22

Triage-based Classification 23

Predictors of Primary Care Sensitive Emergency Department Use 24

Differentiating Between Predictors of General, Frequent, and Primary Care Sensitive Use 28

Proposed Study 28

Datasets 29

Managed Care Network 29

MarketScan 31

Specific Aims 34

Format of this Dissertation 35

Chapter Summary 35

CHAPTER II. PREDICTIVE MODELING OF PRIMARY CARE SENSITIVE AND OVERALL EMERGENCY DEPARTMENT UTILIZATION USING ENHANCED ADMINISTRATIVE DATA: A COMPARISON OF ALTERNATIVE MEASURES AND METHODS 37

Abstract 38

Introduction 40

Methods 41

Conceptual Framework 41

Data Sources and Study Sample 42

Measures 44

Statistical Analyses 48

Results 52

Practice Observed-to-Expected Ratios 54

Discussion 59

Limitations 61

CHAPTER III. RISK-ADJUSTED PREDICTIVE MODELS OF EMERGENCY DEPARTMENT UTILIZATION BASED ON ADMINISTRATIVE DATA 67

Preface 68

Abstract 69

Introduction 71

Methods 73

Conceptual Framework 73

Data and Sample Selection 73

DxCG Clinical Classification Systems 74

Predictor Variables and Outcome Measures 75

Predictor Variables 75

Outcome Measures 75

Statistical Analyses 77

Results 80

Outcome Measure 1: Any ED Use 80

Outcome Measure 2: PCS ED Use 81

Factors Related to Overall and PCS ED Use 83

Discussion 84

CHAPTER IV. USING ENHANCED ADMINISTRATIVE DATA TO PREDICT EMERGENCY DEPARTMENT UTILIZATION: THE ROLE OF NEIGHBORHOOD POVERTY 90

Abstract 91

Introduction 93

Methods 94

Data Sources 94

Study Sample 94

Measures 95

Practice and Provider Characteristics 96

Problem List Variables 96

Morbidity Measurement 98

Small-area Analysis 98

Poverty Analysis 100

Statistical Analyses 101

Results 103

Baseline Characteristics 103

Enrollees 103

Small-area Analysis 104

Unadjusted Effects of Poverty and Morbidity on ED Utilization 107

Prediction Models and Predictors 108

Discussion 116

Limitations 117

Conclusions 119

CHAPTER V. DISCUSSION AND CONCLUSIONS 121

Summary 122

Strengths and Limitations 125

Implications 127

In Conclusion 132

APPENDIX A. VISITS FLAGGED AS UNCLASSIFIABLE BY THE NYU ED ALGORITHM 143

APPENDIX B: DETAILED DESCRIPTION OF MCN PROVIDER QUALITY SCORE 145

APPENDIX C. TOP-CODED VARIABLES 148

APPENDIX D. CONCORDANCE BETWEEN PROBLEM LISTS AND CLAIMS 149

APPENDIX E. CENSUS TRACT VARIABLES, SOURCE FILES, AND EXPLORATORY FACTOR ANALYSIS 155

LIST OF TABLES

Table 1 - 1. Characteristics of studies included in the literature review 16

Table 1 - 2. Significant predictors of primary care sensitive emergency department visits from the refereed literature 25

Table 1 - 3. PCP and enrollee eligibility criteria, MCN data 29

Table 2 - 1. Examples of NYU ED algorithm classifications for selected enrollees, MCN development data, 2011 46

Table 2 - 2. Examples of outcome measure calculations for selected enrollees, MCN development data, 2011 (prediction year) 48

Table 2 - 3. Sociodemographic characteristics, MCN data, 2009-10 (base years) 53

Table 2 - 4. Emergency department utilization using 5 different measures, MCN development data, 2011 (prediction year) 54

Table 2 - 5. Comparison of two hypothetical practices with different distributions of eligibility months in the target year 63

Table 3 - 1. Formulas for model performance metrics 78

Table 3 - 2. Descriptive characteristics of the MarketScan sample, 2007 (base year) 80

Table 3 - 3. Model 1 (any ED use) performance metrics, MarketScan data, 2008 prediction period 81

Table 3 - 4. R^2 results for different estimation methods of Model 2 - PCS ED Use, MarketScan data, 2008 82

Table 3 - 5. Model 2 (PCS ED use) performance metrics, MarketScan data, 2008 prediction period 82

Table 3 - 6. Coefficients, standard errors, z-scores, and *P*-values from models of both outcome measures, MarketScan data, 2008 83

Table 4 - 1. Demographic and clinical characteristics in the base year, by sample, MCN data, 2009-10 104

Table 4 - 2. Improvements in R^2 values for each outcome measure when additional sets of predictors are included 109

Table 4 - 3. Coefficients, standard errors, z-scores, and *P*-values from models of all 3 outcome measures, MCN data, 2010-11 112

Table A - 1. Diagnosis codes associated with 10 or more ED visits that were flagged as unclassifiable by the CHIA-updated version of the NYU ED Algorithm, 2010 143

Table B - 1. Quality measures included in the MCN provider quality score 145

Table C - 1. Top-coded variables and values, MCN data, 2009-11 148

Table D - 1. Algorithms for creating condition indicators from problem list entries in the electronic medical record 150

Table D - 2. Prevalence of selected conditions in problem lists and claims 152
Table E - 1. Census variable definitions, files, variable names, and mean values 155

LIST OF FIGURES

Figure 1 - 1. Conceptual model of emergency department utilization 11
Figure 1 - 2. Flow chart of article selection process 15
Figure 1 - 3. NYU ED Algorithm decision tree 21
Figure 2 - 1. Distribution of Total PCS ED visits, MCN data 51
Figure 2 - 2. Outcome 1: mean observed and expected overall emergency department visit rates (top) and O/E ratios (bottom) by practice 56
Figure 2 - 3. Outcome 2: mean observed and expected total number of emergency department visits (top) and O/E ratios (bottom) by practice 57
Figure 2 - 4. Outcome 3: mean observed and expected number of PCS emergency department visits (top) and O/E ratios (bottom) by practice 58
Figure 4 - 1. Map of primary care practices, hospital emergency departments, and enrollees 106
Figure 4 - 2. Emergency department use varies by neighborhood income category, MCN development data 107
Figure 4 - 3. Mean morbidity score using base-year claims by number of ED visits in the subsequent year 108
Figure 4 - 4. Model predictions vs. actual utilization in the validation sample by predicted quantile of risk 110
Figure 4 - 5. Practice-level observed-to-expected ratios for total PCS ED visits when poverty is included as a predictor 114
Figure 4 - 6. Practice-level observed-to-expected ratios for total PCS ED visits when poverty is NOT included as a predictor 115
Figure D - 1. Concordance between claims and problem lists for 10 priority conditions 154

PREFACE

Portions of this dissertation were presented at the following seminars and meetings:

Lines, L. M. (2012, February). Predictors of potentially avoidable emergency department visits: A systematic review. Presented at UMass Medical School, Worcester, MA.

Lines, L. M., & Ash, A. (2013, June). What is the right outcome measure for emergency department visits? Presented at AcademyHealth Annual Research Meeting, Baltimore, MD.

Lines, L. M., & Ash, A. (2013, October). Does neighborhood income predict emergency department visits? Presented at American Public Health Association Annual Meeting, Boston, MA.

ABBREVIATIONS

AOR	Adjusted odds ratio
ACS	Ambulatory care sensitive
CDC	Centers for Disease Control and Prevention
CI	Confidence interval
ED	Emergency department
EMR	Electronic medical record
ESI	Emergency severity index
FPT	Federal poverty threshold
GLM	Generalized linear model
HMO	Health maintenance organization
IOM	Institute of Medicine
MCN	Managed-care network
MEPS	Medical Expenditure Panel Survey
NC	No change
NYU	New York University
O/E	Observed/expected
OLS	Ordinary least-squares
PCMH	Patient-centered medical home
PCP	Primary care provider
PCS	Primary-care sensitive
USC	Usual source of care
ZINB	Zero-inflated negative binomial

CHAPTER I. INTRODUCTION

The purpose of this chapter is to provide a broad overview of the issues associated with emergency department (ED) utilization from the perspectives of researchers, policymakers, payors, and ED users. We introduce the concept of primary-care sensitive (PCS) ED visits, defined as visits for nonemergencies and conditions that are treatable in primary care settings or potentially avoidable with timely, effective primary care. We discuss why patients go to the ED and why policymakers and payors seek to reduce ED use. We discuss our conceptual framework for this research and include a comprehensive review of the published literature on methods used to categorize PCS ED use and the predictors of such use that have been previously identified.

Background

The Institute of Medicine (IOM) described emergency medicine as "at the breaking point" in 2007.[1] Several converging factors contribute to the problem. Primarily, ED utilization is growing as the number of EDs nationwide is shrinking. From 1999 to 2009, ED visits increased by 32%, while the number of EDs decreased by 2%.[2] Further, as the "safety net for the safety net" and the only source of care guaranteed to all Americans regardless of ability to pay, EDs face a steady demand for uncompensated care.[3] *Growth* in ED visits is not driven by the uninsured, however, whose visit rates have remained steady for years;[4,5] nor to undocumented immigrants, who typically have very low rates of ED use.[6] Instead, non-elderly, insured patients appear to be driving the increase.[3,7] Barriers to timely access to primary care are associated with increased ED use,[8-10] and areas of the country with the longest waiting times for ambulatory care appointments have the highest ED use rates.[6]

In 2010, the number of ED visits in the US was 129.8 million, that is, 42.8 visits per 100 person-years (up from 34.2 in 1999).[11] These rates are highest among nursing home residents, children under 1 year of age, African Americans, the homeless, and persons over age 75.[11] In 2010, about 23% of all ED visits were injury-related, and 18% resulted in a hospital admission.[11]

Why is reducing unnecessary ED use important? First, EDs have experienced severe overcrowding in recent years, which degrades care and harms patients.[1,12] More than 90% of EDs report that overcrowding is a serious problem, and 40% report that the problem occurs daily.[13] ED overcrowding and long wait times result from many problems, including a lack of available beds in other hospital departments, staff shortages, an aging population, loosening of managed care controls, and patients' perception of EDs as comprehensive diagnostic centers, among other factors.[14,15]

Second, care in the ED may signal or contribute to poor coordination among providers, potentially resulting in unnecessary procedures and worse care.[16] With only 23% of EDs completely transitioned to using electronic medical records, and with understaffing common, communication and follow-up are often challenging.[1,11]

Third, obstacles to accessing primary care often lead to unnecessary ED visits, suggesting an underlying problem that, if mediated, could reduce unnecessary ED use. In surveys, as many as half of patients visiting the ED for nonurgent reasons (such as a sore throat) cited not being able to get a timely appointment with their healthcare provider as a reason.[17-21]

Finally, care in the ED is more expensive than care in other settings. Many studies have found that the costs to Medicare, Medicaid, and other third-party payors, as well as patient out-of-pocket costs, are considerably higher (320%-728% in one study[22]) for the same services provided in other, less-acute settings.[22-25] Reducing unnecessary ED use represents an opportunity to save as much as $38 billion per year.[26]

Why Do Patients Go to the Emergency Department?

Patients use emergency departments for many reasons. Most importantly, the majority of patients and/or their caregivers believe their condition is an emergency. As many as 60% of ED visits take place outside of normal business hours,[11] and surveys report that EDs are often the destination when a patient cannot get time off work to go to their PCP during business hours or cannot get an evening or weekend PCP appointment.[20] In about 40-50% of cases, patients go to the ED because their PCP or another provider referred them there.[20,27,28] Some patients have more confidence in the ED's ability to diagnose problems, believe they could receive better care at the ED, or prefer the ED because it offers comprehensive evaluation and diagnostic services at one location.[6,19,29] Some patients cannot get to their PCP's office as easily as the ED – either because of transportation issues, or because the ED is closer.[19,30,31] In other cases, patients choose the ED over an urgent care center because they would have to pay more out-of-pocket to go to an urgent care center[25] (because it may not accept their insurance, particularly if they have Medicaid), or there is no urgent care center nearby. In Massachusetts, 55% of individuals who visited the ED for a nonemergency reported going because they were unable to get a PCP appointment as soon as one was needed.[32]

Problems with the Emergency Department from the Patient's Perspective

From the patient's perspective, the ED has many disadvantages, beginning with the often long and stressful wait times.[33] At the ED, patients have to see a provider they do not know, leading to a potential lack of trust. Moreover, the ED provider does not know – and usually cannot access – the patient's full medical history. This has two potential consequences: their regular doctor might not get test results, leading to potential duplicate testing; and patients might get unnecessary tests and treatments, because of defensive medicine practices.[34] After the ED visit, there is often no follow-up, or the follow-up is not coordinated with other providers. Most patients leaving the ED do not fully understand their diagnosis, the care they received in the ED, or post-ED self-care instructions.[35] Finally, there is a higher risk of serious medical errors in the acute setting, and the risk is greater in fast-paced emergency rooms where providers must juggle several critically ill patients whom they are often meeting for the first time.[33]

Nonetheless, it is important not to discourage patients from going to the ED in a true or suspected emergency. Patients should not be blamed for going to the ED for chest pain that turns out to be indigestion or even for going to the ED for something relatively minor if there are no viable alternatives. In addition, presenting to an ED quickly after the onset of many acute conditions can be crucial; for example, treating patients with a thrombolytic within hours of a stroke greatly decreases the likelihood of later disability.[36] Moreover, discouraging ED use by raising copayments carries a risk of serious unintended consequences (particularly for low-income patients), including inpatient admissions and mortality.[37,38]

Overview: Measuring Emergency Department Use

Researchers disagree on how best to measure ED visits, and the measurement used is not always best suited to accomplishing a particular purpose. While overall use (e.g., number of visits per 100) is a standard, simple metric, it does not distinguish between appropriate and potentially avoidable visits and may not be sensitive to differences in the quality of care provided. One alternative is to focus on number of visits and frequent visitors to the ED.[39] However, there is no standard definition of a "frequent" visitor,[39-41] and this measure does not distinguish between "appropriate" visits versus potentially inappropriate visits (those that could have taken place in a less-acute setting or could have been avoided with better primary care). Moreover, many frequent ED visitors are also heavy utilizers of other types of care, including ambulatory visits and inpatient stays,[40,42-44] suggesting both a high level of medical need and the ability to access care, which may not always be the case for other ED utilizers.[41]

An emergency medical condition is defined as active labor for pregnant women or acute conditions for any patients that could cause death, serious bodily organ harm or serious bodily function impairment if not treated immediately.[45] The 1986 Emergency Medical Treatment and Active Labor Act (EMTALA) requires all hospitals that participate in the Medicare program to evaluate any patient who comes to the ED and provide necessary stabilizing treatments for *an emergency condition*, regardless of the ability to pay.[46] Because of this legal requirement (and ever-increasing cost pressures), appropriately and efficiently distinguishing between emergency and nonemergency conditions has become an important goal in ED care.

Primary Care Sensitive Emergency Department Use

One of the problems with distinguishing between emergencies and nonemergencies in the ED is the lack of agreement on definitions of each, depending on the perspective: 1) patients' own perceptions of the acuity of their conditions, 2) admitting nurses' perspective of patients' acuity at the ED triage station, and 3) final determination of patients' acuity after evaluation. The discharge diagnoses reflect this last definition. As an example, among all ED visits in the 2009 National Hospital Ambulatory Medical Care Survey categorized as nonurgent, primary-care treatable, or preventable/avoidable using the primary discharge diagnosis, 11% were triaged as needing immediate care and 13% resulted in an inpatient admission.[47] In contrast, among 287 respondents to a survey of women who visited a specialty ED in Rhode Island for a condition that triage nurses considered not to be an emergency, 36% felt that their condition was a true emergency.[28] Another survey of adults (n=279) presenting at the ED with low-acuity conditions as judged by a triage nurse found that 74% of respondents believed their condition was urgent.[29]

A primary care sensitive (PCS) ED visit is an outpatient ED visit for either a nonemergency condition or an emergent condition that could have been prevented by good primary care or treated in a primary care setting. The term "PCS" has been used by State officials in Utah,[48] Blue Cross-Blue Shield in Michigan,[49] and in the District of Columbia, among others. We use this term, rather than "inappropriate" or "unnecessary," to highlight the connection between these kinds of visits and primary care. PCS ED visits can be seen as a failure of the healthcare system to provide high-quality, coordinated care for chronic conditions, including timely access to care in a more appropriate setting (e.g., extended hours, seeing urgent cases quickly, etc.).[3]

One national study suggests that PCS visits may be responsible for much of the recent increase in ED use; using data from the National Center for Health Statistics, the increase in total ED visits between 1997-98 and 1999-2000 was attributed to visits classified as semi-urgent, nonurgent, or no/unknown triage.[6] In Massachusetts, nearly half of outpatient ED visits in 2008 were deemed potentially preventable or avoidable.[50] In pre-reform Massachusetts (Fall 2006), 34% of adults age 18-64 visited the ED in the prior year, and 16% said that their most recent ED visit was for a nonemergency condition. Post-reform (Fall 2009), there were no significant differences.[51]

Analysts of PCS ED use often focus only on outpatient, or ambulatory, ED visits (i.e., visits by those who were not admitted to the hospital after their ED visit), since ED visits on the path to an inpatient stay are considered unavoidable.[52,53] In 2010, approximately 18% of ED visits ended in a hospital admission nationwide.[11] On the other hand, if certain hospitals admit much higher fractions of their ED visits than others, excluding these visits from analyses could bias the results of studies of ED use.

Concerns about Reducing Emergency Department Use

Although most policymakers and payors believe that reducing potentially avoidable ED use is a worthy goal, many in the emergency medicine (EM) community disagree. They argue that because EDs have fixed costs and must remain staffed at certain levels to be prepared for all types of unscheduled acute care, traumas, and pandemics, the marginal “total cost to society” of providing care to patients with sore throats and headaches is minimal.[54] However, the evidence for this hypothesis is mixed: some studies have found economies of scale, but others have not, and marginal costs to payors, per case, ranged from $150 to $638 in 2010 dollars.[2] Also, consider the similar case of fire departments: they must also be staffed sufficiently to meet emergency needs, but having firefighters rescue kittens from trees is generally seen as an inefficient use of resources.

Opponents of efforts to reduce ED use also point out that care in the ED accounts for only a fraction of overall health expenditures in the US: about 5-10% in recent estimates.[2] However, this is huge, and comparable to the 9% of healthcare expenditures attributed to pharmaceutical spending in 2012.[55] Considering that EM physicians make up only 4% of the workforce, yet manage 28% of all acute encounters and influence half of all inpatient admissions, it is clear that the ED plays a large role in the US healthcare system.[12]

Primary Care Payment Reform

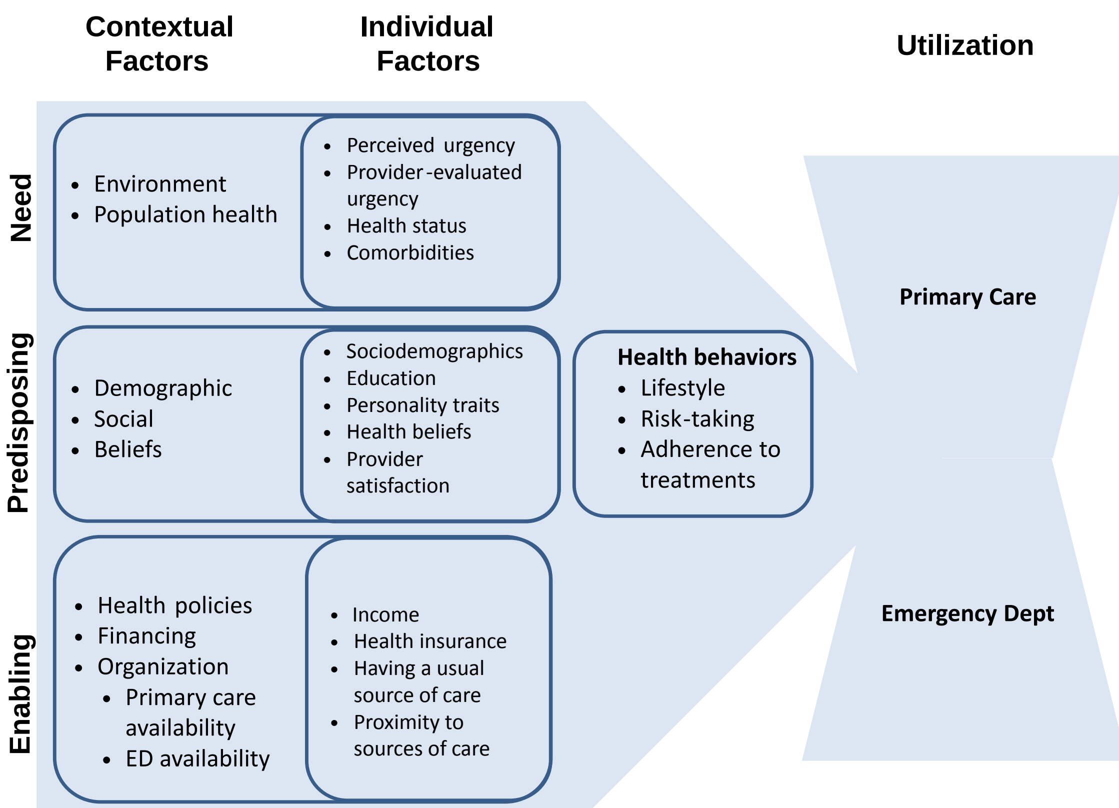
In response to the IOM recommendation to realign financial incentives to produce better medical care rather than more care,[56] different models of providing and paying for primary care have evolved. One is the patient-centered medical home (PCMH), which typically has the following components: a personal physician for every patient; a holistic approach to caring for each patient; coordinated and integrated care across all aspects of the healthcare system; a focus on quality and safety (including evidence-based medicine, shared decision-making, and performance measurement); enhanced access to care; and payment reforms ranging from paying care management fees to reimbursing for specific components of the PCMH.[57] The PCMH is intended to provide patients with a stable and consistent relationship with a healthcare team that provides timely access to coordinated care, including same-day and after-hours appointments.[58] It is hoped that the PCMH will reduce undesirable types of healthcare utilization, including unnecessary ED visits. Indeed, PCMH implementations have reduced overall ED use in at least 10 different evaluation studies, in a wide variety of populations and settings, with reductions ranging from 12-50%.[59]

Conceptual Framework

The conceptual framework for this study was adapted from Andersen's Behavioral Model of Health Services Use, in which health services utilization reflects the combined effects of contextual and individual need, predisposing, and enabling factors, as well as health behaviors.[60] Contextual refers to the context, or setting, in which an individual exists—the time and place, as well as the social, economic, political, and natural environment. Need refers to medical need, both perceived and actual. Predisposing factors include sociodemographic factors (age, sex, race, education, etc.) and health beliefs. Enabling factors either encourage or discourage access to care, such as income, health insurance, having a usual source of care, physician supply, and the patient's proximity to sources of ED or alternative care.

Applying the conceptual framework to an imaginary patient, consider Maria, a 31-year-old woman with less than a high school education (predisposing factors) living in a low-income neighborhood near Route 9 (enabling factors). She has asthma, which is exacerbated by the diesel exhaust from Route 9 (need factors). However, she has never been diagnosed, because she has not told her PCP, who is not fluent in Spanish, that she sometimes experiences wheezing (predisposing and health behaviors). She has a high deductible health insurance plan through her employer, and lives very near UMass Memorial Hospital (enabling factors). All of these factors could contribute to Maria's higher-than-average likelihood of going to the ED for a PCS visit.

Figure 1 - 1. Conceptual model of emergency department utilization



Adapted from McCusker et al.[78] and Andersen and Davidson[61]

Some modifications are needed to apply the Andersen model to ED use (**Figure 1-1**). The original model predicted health services utilization in general, whereas a model of ED use should account for the relationship between ED utilization and other types of utilization – particularly primary care. If ED utilization is a result, in part, of barriers to timely access to primary care, then enabling factors that increase the use of primary care should reduce the number of ED visits.[62] For example, if a patient's primary care provider offers extended (evening/weekend) operating hours, that could reduce ED visits (especially PCS ED visits). Similarly, greater numbers of primary care providers might be expected to increase primary care utilization and reduce ED utilization.[63]

Literature Review

To understand the current state of the art on classifying, measuring, and predicting ED utilization, we conducted a systematic review of the published literature, focusing on the predictors of PCS ED use. The following sections describe the methods used to search and summarize the literature and the results of the search. We then discuss findings and conclusions.

Search Terms and Sources

To identify relevant journal articles for this study, we systematically reviewed English-language articles published through November 2013. Search terms used in searches of PubMed (free text and MeSH terms) included: Emergency Medical Services/utilization, Emergency Service, Hospital/utilization, Patient Admission/statistics & numerical data* AND emergency [tiab], Health Services Misuse/statistics & numerical data* AND emergency [tiab], Health Services Needs and Demand AND emergency [tiab]. Keywords combined with the term "emergency" included: nonurgent OR non-urgent, nonemergent OR non-emergent, avoidable, primary care treatable, ambulatory care sensitive, low complexity OR low-complexity, lower acuity OR low acuity OR low-acuity, appropriateness, appropriate use, and inappropriate.

We hand-searched key journals, including *Academic Emergency Medicine*, *Annals of Emergency Medicine*, *Medical Care,* and *HSR: Health Services Research*, for relevant articles. We also reviewed bibliographies of relevant articles and conducted internet searches using Google Scholar to locate articles not indexed in PubMed.

Inclusion and Exclusion Criteria

To be included, we required articles to: be written in English, have an abstract available, be published in a peer-reviewed journal, and provide quantitative data on the predictors or determinants of primary care sensitive (PCS) emergency department (ED) use among adults in the United States. We focused on adults because they are the most policy-relevant population under health reform, since many children are already covered by Medicaid and State Children's Health Insurance Programs (SCHIP).

We excluded articles that provided data only on predictors of frequent use or general predictors of ED visits (i.e., articles that did not define PCS, preventable, inappropriate, or unnecessary visits). We also excluded literature reviews, commentary/opinion articles, letters to the editor, and editorials.

Data Abstraction and Quality Assessment

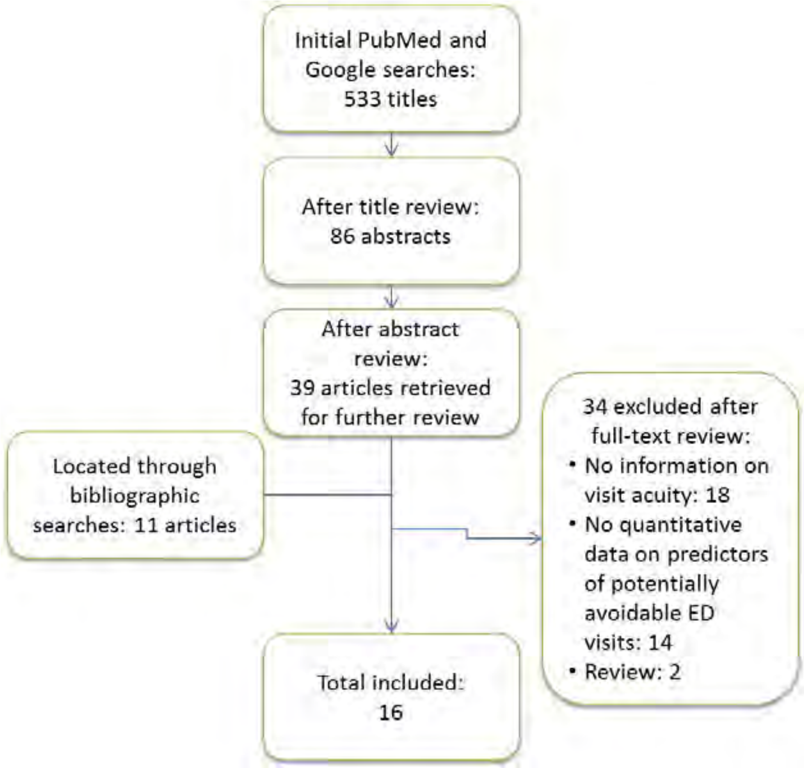
We abstracted the following data from each study: author and year, setting, sample characteristics and patient population, study design and statistical methods, outcome measures, definition of PCS use, results and conclusions, accuracy of the algorithm or model used to predict visits, and strength of the evidence (i.e., quality rating). We rated the quality of reviewed studies using a modified Downs & Black checklist.[64] The following 12 criteria were used: 1) clear descriptions of aims; 2) clear descriptions of outcomes; 3) clear descriptions of patient characteristics; 4) clear descriptions of principal confounders; 5) clear descriptions of main findings; 6) random variability for the main outcome provided; 7) actual *P* value reported; 8) appropriate statistical methods used; 9) accurate outcome measures used; 10) participants recruited from the same population; 11) participants recruited at the same time; and 12) adequate adjustment for confounders performed.

We classified the method of analysis in each study as descriptive (univariate or bivariate associations only) or multivariable. We noted variables that were statistically significant predictors ($P<.05$) of ED utilization. If an article contained both descriptive and multivariable results, we have reported the multivariable results only. We categorized significant factors derived from multivariable analyses as representing need, predisposing, or enabling factors, according to Andersen's behavioral model of health services utilization.[60]

Articles Retrieved and Descriptive Characteristics

We identified 533 articles using the search strategy described above. After eliminating 447 studies that were clearly not relevant based on their titles, we reviewed 86 abstracts for relevance. We eliminated 47 studies based on their abstracts and retrieved 39 studies for full-text review. Another 11 full-text articles were retrieved after reviewing the bibliographies of these 39 articles. From the 50 articles reviewed in full, 16 articles met all inclusion criteria (**Figure 1-2**). All 16 articles described observational studies and were published between 1993 and 2010. Seven studies reported only descriptive results; these studies' quality ratings ranged from 50% to 83% (mean: 63%). The other 9 studies, which reported multivariable estimates, had quality ratings ranging from 83% to 100% (mean: 94%).

Figure 1 - 2. Flow chart of article selection process



Four studies drew from nationally representative, population-based surveys (see **Table 1-1**). One article published results from a national study at 56 EDs across the US, which surveyed patients who were triaged as nonurgent by an ED nurse regarding their reasons for seeking care in the ED. Another 5 studies relied on regional multi-site data. For example, Wharam et al. used data from nonelderly enrollees in the Harvard Pilgrim Health Plan in Massachusetts, comparing nonurgent ED use between those in high-deductible health plans with those in more traditional plans.[65] Finally, 6 articles reported on single-site studies; all were single-hospital surveys of nonurgent patients regarding their reasons for seeking care in the ED.[65]

Sample sizes (or number of ED visits studied) ranged from 94 to 135,723 (median: 3,003). **Table 1-1** summarizes the characteristics of each included study, listing studies that used multivariable analyses first, followed by those that used descriptive methods only; within each section, studies are sorted from highest quality rating to lowest.

Table 1 - 1. Characteristics of studies included in the literature review

Citation	Population & Setting; Time Period	Sample Size & Sampling Method	Definition of PCS ED Visits	Outcome Measure(s)	Quality Rating
			Multivariable Studies		
Liu et al. 1999	All ED visits reported in the National Hospital Ambulatory Medical Care Survey (~400 different hospitals); 1992-1996	n=135,723 ED visits; 4-stage probability sampling used to generate nationally representative estimates	Nonurgent visits were defined as ones in which the patient "does not require attention immediately or within a few hours"	Nonurgent ED visits	100%
Lowe et al. 2005	Nonelderly Medicaid patients assigned to 353 primary care practices affiliated with a Medicaid HMO in Pennsylvania; August 1998 to July 1999	n=57,850; practices randomly selected from database provided by HMO, and patients included if assigned to one of the eligible practices and under age 65	"Potentially avoidable": there was a high probability that a prompt appointment in a primary care practice could have averted the ED visit (using an early version of the NYU ED algorithm by Billings et al)	ED use overall, potentially avoidable ED use	100%
Sarver et al. 2002	Noninstitutionalized civilian adult respondents to the Medical Expenditure Panel Survey with a usual source of care other than the ED who had 1+ health system contact; 1996	n=9,146; nationally representative sample	Using modified Cunningham et al criteria, a visit was considered urgent if 1) it resulted in an admission; 2) it included any imaging or surgical procedure and it was reported as for an accident or injury, diagnosis, or treatment and not the result of a referral; or 3) the reason for the visit was reported as accident/injury, diagnosis, or treatment and the visit was within 3 days of the accident/injury or symptom onset. All other visits were considered nonurgent.	Nonurgent ED visits	100%
Wharam et al. 2007	Nonelderly Massachusetts enrollees in Harvard Pilgrim Health Plan; March 1, 2001 to June 30, 2005	n=68,281; high-deductible health plan group included those with 1+ years' continuous enrollment in traditional HMO followed by 6+ months in a high-deductible health plan, while control group included those with traditional HMO plans (groups randomly matched 8:1)	NYU ED Algorithm (Billings et al): visits were classified as low severity if the probability of needing ED care was less than 25% using the algorithm, which assigns probabilities based on ICD-9 codes	Total ED visits, first visits, and repeat visits, as well as low-, indeterminate-, and high-severity first and repeat visits, comparing the two study groups to determine the effect of high-deductible health plans on each type of visit	100%

Citation	Population & Setting; Time Period	Sample Size & Sampling Method	Definition of PCS ED Visits	Outcome Measure(s)	Quality Rating
Petersen et al. 1998	Adults with chest pain, abdominal pain, or asthma presenting to 1 of 5 urban EDs in the Northeast; 1 month in 1993	n=1696; convenience sample	Triage criteria developed by Baker et al. Patients were classified as urgent if they had abnormal vital signs, urgent chest pain (taking risk factors into account), asthma symptoms present for less than 1 week, abdominal pain present for less than 48 hours AND patient was either older than 65, pregnant, or was experiencing bleeding. All other presentations were classified as nonurgent.	Nonurgent ED visits	92%
Chiou et al. 2010	Type 2 diabetics enrolled in a disease management program in Louisiana; 1999-2006	n=8596; all patients meeting inclusion criteria were selected	ICD-9 codes for visits occurring on weekdays were classified by 2 expert coders as either urgent or less urgent.	Inappropriate use of the ED	92%
Cunningham et al. 1995	Civilian non-institutionalized US respondents to the National Medical Expenditure Survey (NMES); 1987	n=35,000; all respondents were included in the analysis	Visits were classified as urgent if they 1) resulted in a hospitalization, 2) occurred within 3 days of an injury/accident, 3) included any surgical procedures, 4) involved a physician's referral, 5) involved an ambulance, or 6) were associated with a self-reported "very serious" condition. All other ED visits were considered nonurgent.	Nonurgent ED visits	92%
Grumbach et al. 1993	All patients in the ED waiting area at San Francisco General Hospital who were not assigned to the immediate care triage category; one week in July 1990	n=700; all patients meeting inclusion criteria were asked to participate	Acuity score assigned by ED triage nurse: 1 - needs immediate care 2 - needs urgent care 3 - needs care within 3 hours (possibly inappropriate) 4 - needs nonurgent care (inappropriate)	Appropriate ED use	92%
Wolinsky et al. 2008	Elderly (age 70+) respondents to the Survey on Assets and Dynamics Among the Oldest Old; 1991-1996	n=4,310; nationally representative sample	Current Procedural Terminology (CPT) codes 99281-99282 were considered "low intensity" visits.	Low-intensity, mixed-intensity, and high-intensity ED visits	83%
Young et al. 1996	Ambulatory patients at 56 EDs across the US; 1 day in 1994	n=6187; all ambulatory patients presenting during the 24-hour period were eligible	A triage nurse performed a brief, directed examination to determine the urgency of each patient's condition. Nonurgent was defined as "treatment can be safely delayed" 12-24 hours.	Reasons for seeking care in the ED	83%
			Descriptive Studies		
Gill and Riley 1996	Adults and children at an urban teaching hospital; one week in January 1993	n=268; convenience sample	Considered nonurgent by the ED triage nurse	Patient-perceived urgency, self-reported reasons for using the ED	67%

Citation	Population & Setting; Time Period	Sample Size & Sampling Method	Definition of PCS ED Visits	Outcome Measure(s)	Quality Rating
Matteson et al. 2008	Women who visited a specialty OB/GYN ED in Rhode Island for a nonemergency; May-Oct. 2005	n=287; convenience sample of women with nonemergent complaints visiting the ED during the study period	Nurses assessed patients according to the Emergency Severity Index (Wuerz et al), with patients in categories 3-5 considered low acuity.	Reasons for seeking care in the ED	67%
Redstone et al. 2008	Adults with a primary care provider presenting with a nonurgent complaint to the University of Colorado Hospital ED; June-Nov. 2006	n=240; convenience sample with 60 surveys collected during 4 different time frames	Nurses assessed patients according to the Emergency Severity Index (Wuerz et al), with patients in categories 3-5 considered low acuity.	Reasons for seeking care in the ED; comparison between weekday daytime visitors and non-weekday daytime visitors	67%
Pilossoph-Gelb et al. 1997	Ambulatory, noncritical patients at a university-based private ED and a public county hospital ED in the Los Angeles area; Dec. 1994-Dec. 1995	n=700; convenience sample, with attempts made to survey patients at all hours to be representative of ambulatory triage patients	Three ED physicians rated each complaint as 1) life- or limb-threatening if not immediately treated; 2) neither life- nor limb-threatening, but appropriate for ED treatment; 3) neither life- nor limb-threatening and appropriate for treatment in a primary care setting. The majority opinion was used to classify patients.	Occurrence of psychosocial difficulties among emergent/ nonemergent ED visitors	58%
Northington et al. 2005	Adults presenting at the University of North Carolina Hospital between 9am and 1am; June 23, 1999, to August 8, 1999	n=279; convenience sample excluding intoxicated, pregnant, mentally impaired, non-English speakers, suspected abuse victims, those referred by their physicians, and those who refused participation	Low-acuity patients in Emergency Severity Index (Wuerz et al) triage categories of 4 or 5 as assessed by triage nurse. These patients were responsive, oriented, in no acute distress, had stable vital signs, and were estimated to require no more than one resource (lab, test, or consult).	Reasons for seeking care in the ED	50%
Schwartz 1995	Patients at the Family Practice Center, Augusta, GA, with non-life-threatening illnesses who either sought care at the ED or in the clinic during a 1-month period; early 1990s	n=94; all eligible patients invited to participate	Visits for non-life-threatening illnesses, such as bronchitis, cold, flu, sprains	Reasons for seeking care in the ED vs. the clinic	50%

ED: Emergency department; HMO: health maintenance organization

Methods Used to Categorize Primary Care Sensitive Emergency Department Use

The methods described in the literature for categorizing the acuity of ED visits fall into three categories: diagnosis-based, procedure-based, and triage-based. In this section, we describe each method and provide examples of studies that used it.

Diagnosis-based Classification

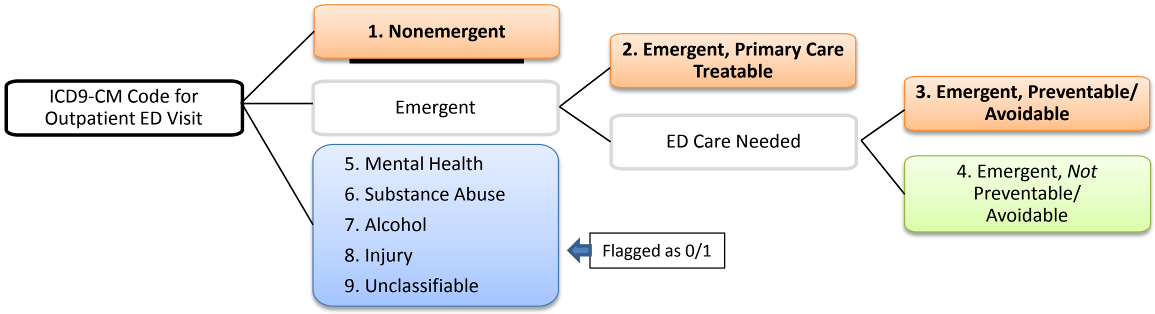
Retrospective classification based on diagnosis codes reflects a judgment as to the probability of a patient's underlying reason for the visit being emergent or nonemergent. There is no consensus on criteria for using administrative data to judge whether a particular visit was potentially avoidable.[54] However, this type of system is best exemplified by the New York University (NYU) ED algorithm.[66]

The developers of the algorithm included emergency medicine and primary care providers. They sought insight into the ED utilization patterns of a population – for example, what proportion of ED visits are nonemergencies or could be treated in a primary care setting. To answer these questions, the researchers, with funding from the Commonwealth Foundation, conducted a detailed chart review of nearly 6,000 medical records from patients seen at 6 New York City hospital EDs in 1994 and 1999. The records examined included patients' initial complaints, vital signs, age, medical history, procedures performed and resources used in the ED, and the final discharge diagnosis.

For each case, the developers determined whether patients were emergent or nonemergent, based on whether the data in the chart suggested that they needed medical care within 12 hours. Next, based on procedures performed and resources used during the visit, they classified each case according to whether care could have been provided in a primary care setting. For example, patients who had a CAT scan during the visit were classified as "emergent/ED care needed." All emergent/ED care needed cases were then evaluated as to whether the complaint could have been prevented or avoided with timely and effective outpatient care. For example, acute exacerbations of asthma may be emergent and require treatment in the ED, but such episodes can be avoided with better management and care. These discharge diagnoses were based on the ambulatory care sensitive conditions previously developed by researchers at NYU and the United Hospital Fund for analyzing hospital discharges.[67]

As shown in **Figure 1-3**, the algorithm assigns the probability that the principal ICD-9 diagnosis code associated with an ED visit falls into 1 of 4 categories: 1) non-emergent (immediate care not required within 12 hours); 2) emergent/primary care treatable (care required within 12 hours that could have been provided in a primary care setting); 3) emergent, ED care needed, possibly preventable/avoidable (ED required, but visit could possibly have been prevented with good primary care); and 4) emergent, ED care needed, not preventable/avoidable. The first three categories represent PCS visits. The algorithm also flags visits with a principal diagnosis code related to injury, mental health, substance abuse, or unclassifiable, and does not consider those visits any further. Unclassifiable visits are generally those that were infrequent in the original data on which the algorithm was developed (please see APPENDIX A for further details). The algorithm uses standardized diagnosis and payment codes, is nonproprietary, and is easily downloaded in SAS, SPSS, or ACCESS from NYU's Center for Health and Public Service Research's website.[66]

Figure 1 - 3. NYU ED Algorithm decision tree



Adapted from NYU Center for Health and Public Service Research.
Primary care sensitive categories are shown in orange (numbered 1, 2, and 3).

The algorithm has been adapted for use by the CDC to describe the characteristics of high safety-net burden EDs, has been validated as accurate in predicting future hospitalizations and mortality, and has been used by several states and municipalities to track ED visit patterns.[48,50,68,69] The algorithm has been validated in its ability to distinguish cases with a higher risk of mortality or subsequent hospital admission from less acute cases.[68] However, some studies have suggested that it is not sensitive to changes in access to care.[70,71] Although it has limitations, the algorithm remains the only validated tool for classifying ED visits using ICD-9-CM diagnosis codes.[72] The algorithm has not been updated by the original developers since 2003,[73] but a version of the algorithm updated in 2009 by the Massachusetts Center for Health Information and Analysis (CHIA) is available from the authors. CHIA's version of the algorithm built on the original to incorporate new codes with input from the original developer and an emergency medicine physician, but did not involve new data abstractions.

This algorithm was used by 2 studies in this review.[65,74] Lowe et al (2005) modified the algorithm by collapsing its 4 emergent/nonemergent categories into 2 (potentially avoidable and probably unavoidable). Wharam et al., used the algorithm to classify visits as high severity (at least 75% likelihood of being emergent), low severity (25% or less likelihood of being emergent), or indeterminate (26-74% chance of being emergent).

Another study by Chiou et al., used a diagnosis-based classification system. In it, two experts in medical coding used ICD-9-CM codes to classify ED visits as urgent or nonurgent.[75] No other information (or citations) about the methods were provided.

Procedure-based Classification

Procedure-based systems look primarily at what occurs during the ED visit, such as whether patients receive imaging tests or are admitted to the hospital. The method of Cunningham et al. classifies visits as urgent or nonurgent based on procedures and utilization that occur during and after the visit as well as the patient-reported reason for the visit.[52] Applied to data from the 1987 National Medical Expenditure Survey, visits were classified as "urgent" if they 1) resulted in a hospitalization, 2) occurred within 3 days of an injury/accident, 3) included any surgical procedures, 4) involved a physician's referral, 5) involved an ambulance, or 6) were associated with a self-reported "very serious" condition.

Sarver et al. modified these criteria slightly for use in the 1996 Medical Expenditure Panel Survey and validated the criteria by also applying them to outpatient hospital and physician office visits, finding that they only classified 6% of visits in each setting as urgent.[53] In contrast, about 60% of ED visits were classified as urgent in both Cunningham et al. and Sarver et al.[52,53]

The American Medical Association's Current Procedural Terminology (CPT) codes are used by providers when submitting claims to insurance companies and can be used in a procedure-based classification system. One study, by Wolinsky et al., used CPT codes to classify ED visits, categorizing visits with the CPT codes of 99281 (self-limited problem) and 99282 (low to moderate severity problem) as "low intensity" visits.[76] The authors validated the method against diagnosis codes and against the NYU ED algorithm and reported good criterion validity.

Triage-based Classification

In most emergency departments in the US, patients are assigned by a nurse or physician to 1 of between 3 and 5 triage categories based on the clinician's judgment on how soon the patient needs to be seen. This prospective classification is part of the medical record, subjectively reflects degree of urgency, and is done prior to a definitive diagnosis. Unfortunately, ED triage acuity systems in the United States are not standardized [77] and their reliability varies widely.[78,79] A version of this system is currently used by the National Hospital Ambulatory Medical Care Survey (NHAMCS), in which survey respondents assign visits to urgency categories based on the triage category used by their hospital, which are then recoded to one of five categories: immediate (should be seen within 1 minute), emergent (should be seen in 1-14 minutes), urgent (should be seen in 15-60 minutes), semiurgent (should be seen in 61-120 minutes), and nonurgent (should be seen in 121 minutes to 24 hours). The study by Liu et al. relied on this NHAMCS classification system.[80] Five other studies in this review used triage-based methods of classifying visit acuity developed by that study's authors.[21,81-84]

Another 3 studies used the Emergency Severity Index (ESI) to classify ED visits.[20,28,29] The ESI is a triage system developed by Wuerz et al for use by ED nurses and physicians; it has been validated against patients' subsequent resource needs (such as diagnostic testing and hospitalization) and for inter-rater reliability between nurses and physicians (weighted k = .80 [95% CI = 0.76 to 0.84]). The flowchart-based algorithm sorts patients into 5 categories (ESI-1 being most acute) based on patient medical condition (including acuity, stability of vital functions, and degree of distress), expected resource intensity (such as cardiac monitoring, specialty consultation, or diagnostic tests), and timeliness (expected staff response, time to disposition). Vital signs are used to move patients from ESI-3 to ESI-2, but are not used in assignment to other categories.[85]

Predictors of Primary Care Sensitive Emergency Department Use

In **Table 1-2**, we show reported adjusted odds ratios (AORs) from the 16 studies included in our review for factors associated with PCS ED use in multivariable analyses, arranged by whether they represent contextual or individual need, predisposing, or enabling factors. Since different studies controlled for different factors, odds ratios are only roughly comparable across studies. However, to facilitate comparisons between studies and factors, we have transformed AORs that were originally reported as negative (below 1) by reporting their inverse, that is, 1/AOR, and use a diesis (‡) to indicate this in the table.

Table 1 - 2. Significant predictors of primary care sensitive emergency department visits from the refereed literature

Factor	Reference Group	AOR*	Population (n)	Study
Need				
More than 5 bed days	5 or fewer	1.03	General (n=30,038)	Cunningham 1995
More than 5 reduced activity days	5 or fewer	1.02	General (n=30,038)	Cunningham 1995
Fair health	Excellent/very good/good	2.12	Adults with USC (n=9,146)	Sarver 2002
Poor health		2.94		
Poor health	Excellent health	2.17‡	General (n=30,038)	Cunningham 1995
	Good health	1.52‡		
No prior year hospitalization	Hospitalization in prior year	1.19‡	Type 2 diabetics (n=8,596)	Chiou 2010
Predisposing factors				
Age 18-24	Age 45+	2.79	Adults with USC (n=9,146)	Sarver 2002
Age 25-44		1.66		
Age 16-30	Age > 60	4.80	Adults (n=1,696)	Petersen 1998
Age 31-40		6.50		
Age 41-50		2.40		
Age 51-60		2.00		
Younger age	N/A (continuous)	1.05‡	General (n=30,038)	Cunningham 1995
Under age 65	Age 65 or older	1.79‡	General (n=135,723 visits)	Liu 2003
Black	White	1.22‡	Type 2 diabetics (n=8,596)	Chiou 2010
Black	White	1.68	General (n=30,038)	Cunningham 1995
African American	White	1.08	General (n=135,723 visits)	Liu 2003
Female	Male	1.30	Adults (n=1,696)	Petersen 1998
Female	Male	1.44	Adults with USC (n=9,146)	Sarver 2002
Female	Male	1.12‡	General (n=135,723 visits)	Liu 2003
Fewer years of education	N/A (continuous)	1.03‡	General (n=30,038)	Cunningham 1995
Lower immediate word recall score	Higher immediate word recall score	1.55	Elderly (n=4.135)	Wolinsky 2008
Smaller family size	N/A (continuous)	1.09‡	General (n=30,038)	Cunningham 1995
Enabling factors				
Dissatisfaction with USC score	N/A (continuous)	1.13	Adults with USC (n=9,146)	Sarver 2002
No regular doctor	Regular doctor	1.60	Adults (n=1,696)	Petersen 1998
Household income < 125% of FPT	Income 400%+ of FPT	1.70	Adults with USC (n=9,146)	Sarver 2002
Household income 125-399% of FPT		1.39		
Large facility (100+ beds)	Small facility	1.44	Type 2 diabetics (n=8,596)	Chiou 2010
Commercial insurance	Uninsured	1.28	Type 2 diabetics (n=8,596)	Chiou 2010
Medicaid insurance	Uninsured	1.28	Type 2 diabetics (n=8,596)	Chiou 2010
	Uninsured all year	1.47	General (n=30,038)	Cunningham 1995
	Privately insured	1.54	Adults with USC (n=9,146)	Sarver 2002
	Privately insured	1.14	General (n=135,723 visits)	Liu 2003

Factor	Reference Group	AOR*	Population (n)	Study
Clinic had more Medicaid patients	N/A (ordinal)	1.04	Nonelderly HMO enrollees (n=57,850)	Lowe 2005
Medicare insurance	Uninsured	1.32	Type 2 diabetics (n=8,596)	Chiou 2010
Medicare + other insurance	Uninsured all year	1.61	General (n=30,038)	Cunningham 1995
Living in an area with more EDs	N/A (continuous)	1.37	General (n=30,038)	Cunningham 1995
Living in an urbanized non-metro area	Rural area	1.53	General (n=30,038)	Cunningham 1995
Living in a small city	Major city	2.92	Elderly (n=4.135)	Wolinsky 2008
Living in a rural county		2.29		
Living in the Northeast	South	1.45	General (n=30,038)	Cunningham 1995
Living in the Midwest	Northeast	1.25	General (n=135,723 visits)	Liu 2003
South		1.27		
West		1.28		
Less time in disease management program	N/A (continuous)	1.02‡	Type 2 diabetics (n=8,596)	Chiou 2010
Poor	Middle income High income	1.20‡ 1.39‡	General (n=30,038)	Cunningham 1995
Living in area with lower per-capita income	N/A (continuous)	1.35‡	General (n=30,038)	Cunningham 1995
Working fewer weeks in the year	N/A (continuous)	1.04‡	General (n=30,038)	Cunningham 1995
Lacking a regular source of care	Having a regular source of care	2.39‡	Adults waiting for care in the ED (n=489)	Grumbach 1993
Living in an urban area	Rural area	1.11‡	General (n=135,723 visits)	Liu 2003
For-profit hospital	Nonprofit hospital	1.12‡	General (n=135,723 visits)	Liu 2003
Medicare insurance	Private	1.33‡	General (n=135,723 visits)	Liu 2003
PCP had 5-7 weekday evening hours PCP had 8-11 weekday evening hours PCP had 12+ weekday evening hours	No evening hours	1.22‡ 1.20‡ 1.25‡	Nonelderly HMO enrollees (n=57,850)	Lowe 2005

*All adjusted odds ratios (AORs) in this table were significant at *P* < .05. ‡ AOR has been estimated as 1/original AOR reported by study authors

ED: Emergency department; FPT: Federal poverty threshold; HMO: health maintenance organization; PCP: primary care provider; USC: usual source of care

Enabling factors were studied most. The individual enabling factor most frequently reported to be associated with increased PCS ED use was having Medicaid coverage (4 studies).[52,53,75,80] This may be related to the fact that Medicaid enrollees have low or no copayments when visiting the ED. In contrast, there were mixed findings for Medicare-only coverage: one study found a positive association;[75] one, a nonsignificant association (not shown in table);[52] and one, a negative association between nonurgent visits and Medicare-only coverage.[80] These mixed findings are most likely due to different comparison groups and populations: comparing Medicare beneficiaries to the uninsured might be expected to lead to a different conclusion than comparing them to those with private insurance.

Among predisposing factors, age was significant in 5 out of 8 studies that tested for an association, with all studies finding an increased risk for persons under age 65.[52,53,76,80,82] Women had significantly higher risk in 3 of the 7 studies that tested for an association.[53,80,82] African Americans had a higher risk in 3 of 6 studies.[52,75,80] Impaired cognitive function, as measured by immediate word recall, was also associated with higher risk in 1 study,[76] as was fewer years of education.[52]

Only 2 studies examined need factors (including number of days spent in bed, number of reduced activity days, self-reported health), both mostly confirming the expected association between poorer health and increased risk of PCS ED visits.[52,53] However, among Type 2 diabetics, prior-year hospitalization was associated with a reduced risk of PCS ED visits in one study.[75]

Differentiating Between Predictors of General, Frequent, and Primary Care Sensitive Use

To determine how predictors of PCS ED use differ from predictors of frequent or any use, we consulted two systematic reviews—one of frequent use,[39] and one of determinants of any ED visits in elderly adults[62]—and compared their results with findings from this review of the predictors of PCS ED use. In general, the literature on ED use in elderly adults and frequent users has found that "need" is the driver. Frequent ED users tend to be sicker than occasional users, with greater overall health services utilization. Frequent ED visitors are about 6 times more likely to have been hospitalized in the preceding 3 months[39] and, in the elderly, previous hospital or ED use, or both, were significant determinants of ED utilization.[62]

LaCalle and Rabin found that the risk of frequent ED use is higher among women and African Americans. However, frequent use has a bimodal age distribution, with peaks around age 25 to 44 and over 65 years. In contrast, in three studies, McCusker et al found that older age independently predicted any ED utilization.[62] As with PCS ED use, frequent ED use is more common among those with public insurance.[39]

Proposed Study

To gain further insight into ED utilization and develop improved performance measures for PCPs related to their patients' use of the ED, we proposed a retrospective, observational study in two datasets, described in detail in the next section. We proposed to measure the prevalence of ED use in these populations, develop a set of ED risk prediction models using only administrative data in each dataset, and develop and validate enhanced ED risk prediction models for the MCN, including additional predictors from the neighborhoods in which enrollees resided, provider and payor characteristics, and clinical data from the network's electronic medical record (EMR).

Datasets

Managed Care Network

There were two sources of data for this study. The first was a managed care network (MCN) in Massachusetts. In partnership with the MCN, we obtained and merged several sources of data into a final, deidentified analytic dataset.

Table 1-3 summarizes the eligibility criteria we used to select PCPs and enrollees in the MCN study, showing the original sample and the number of enrollees excluded for each criterion.

Table 1 - 3. PCP and enrollee eligibility criteria, MCN data

PCP eligibility criteria				
Primary care provider affiliated with the MCN in all or part of 2009-11				
Transitioned to AllScripts before 2009				
Enrollee eligibility criteria				
1+ months of coverage in base year (either 2009 or 2010) AND 1+ months of coverage in subsequent year (2010 or 2011)				
Matched to one eligible PCP (per enrollment file)				
Home address in Massachusetts				
	Development		**Validation**	
	N	%	N	%
Original	57,805		59,254	
After excluding enrollees with out-of-state addresses	56,530	97.8%	57,712	97.4%
After excluding enrollees who could not be matched to one PCP	53,112	91.9%	54,337	91.7%

All included persons were enrolled in one of 4 commercial insurance plans for at least 1 month in either 2009 or 2010 and at least one month in the subsequent year (2010 or 2011, respectively). The four plans are among the largest in Massachusetts by market share. Details on the types of coverage (HMO, PPO, etc.) were not available, but all were commercial (private) plans, either employer-sponsored or individually purchased. We included all ages. The enrollees were split into two groups: development (for those enrolled in 2010-11) and validation (2009-10).

Enrollees were all affiliated with one of the 235 primary care providers (PCPs) in the MCN who had transitioned away from paper medical records to using an EMR system. Each enrollee's affiliation with a PCP was determined based on the PCP listed in that enrollee's EMR. Among enrollees who had more than one PCP in their records, we selected the match with greatest number of encounters first, then the last visited provider in the appropriate base year. We excluded 5 enrollees who could not be matched to a single provider. After excluding individuals who lived outside of Massachusetts and those who could not be matched to a participating PCP, the dataset included 64,623 unique enrollees (107,449 observations).

The data included claims for all inpatient, outpatient, ED, and office visits made by each eligible enrollee. Each paid claim, submitted by providers to one of the 4 insurance plans, contained up to 4 diagnostic (ICD-9-CM) codes. The MCN routinely merges all claims from these 4 plans into a single, harmonized data warehouse for quality improvement monitoring on a quarterly basis.

The EMR system used in the MCN is Allscripts (Allscripts Healthcare Solutions, Inc., Chicago, IL), which was implemented in the MCN in 2007. The Allscripts system contains provider-entered information for each patient, including problem lists (conditions and complaints, history of illness, and presence or absence of risk factors such as tobacco use) and measured BMI and blood pressure. To ensure comparable data and minimize missing EMR data, we restricted the providers included in the study to those who had transitioned away from paper medical records to using the EMR by 2009, the first year of our study period.

Data on practice and provider characteristics were obtained from MCN administrative records. Practice characteristics included location, practice specialty, and a provider quality score. The specialties were family practice, internal medicine, maternal/pediatrics, and multi-specialty/other. The provider quality scores were developed by the MCN as part of their internal rewards program, and were based on 21 quality and efficiency measures (primarily HEDIS measures, such as well-child visits; all measures are described in APPENDIX B).[86] The purpose of the internal rewards program was to reward physicians across domains of quality and efficiency with a payout based on their contribution to the network.

Data on neighborhood characteristics were obtained by first mapping each enrollee’s address to a Census tract, then merging the enrollee data with the Census-tract-level data elements, such as median income, percent under the poverty line, and percent homeowners. After linking addresses to Census tracts and enrollment data to clinical and provider data, all records were deidentified.

MarketScan

Our second dataset was the MarketScan (Truven Health Analytics, Ann Arbor, MI) Commercial Claims and Encounters database from 2007 and 2008, a proprietary dataset that we accessed through a partnership with Verisk Health, Inc. Truven Health Analytics compiles this database from claims submitted to health plans that contract with large private employers, public agencies, and public organizations in the United States. The database includes employer-sponsored, private, fee-for service, and capitated insurance plans for employees and covered dependents. MarketScan includes data from approximately 45 large employers who self-insure employees and their dependents.[87] This nationwide claims database is widely used by researchers to examine health services utilization and costs, with over 550 peer-reviewed articles published since 1990.[88] The MarketScan data are validated to ensure that claims and enrollment data are complete, accurate, and reliable, and the data are fully HIPAA compliant.

The MarketScan Commercial Claims and Encounters database includes enrollees of all ages in participating private health plans and large self-insured employer plans, including comprehensive, HMO, POS, and PPO plans. The database contains residents of the 50 US states, the District of Columbia, and the US Virgin Islands; the 2007 database contains 35 million unique enrollees, and the 2008 database contains 49 million. To be included in our study, individuals were required to have drug benefits and to have at least 1 month of enrollment in 2007 and at least 1 month of enrollment in the subsequent 6-month period (January-June, 2008). Our final dataset included 15,136,261 unique individuals.

For both datasets, our eligibility inclusion criteria required only 1 month of eligibility in the base period and 1 month in the prediction period. This approach has several advantages: it includes those who were born or died during either period, it retains all observations that potentially contribute information, and it is consistent with an implementation-oriented approach that seeks to include as much information as is available on as wide a cross-section as feasible of the kinds of people who will need to be managed. Others have found that risk adjustment models improve predictive ability (for the whole population) when enrollees with partial observations are modeled with whatever data are available, as opposed to only modeling outcomes for those with full years of eligibility and defaulting to a simple demographic model for the others.[89]

Nonetheless, persons with missing eligibility months in the base year are at risk of having fewer problem list entries in their EMRs and less utilization (office visits, inpatient stays, and ED utilization) and fewer diagnoses listed on their claims data, than if they were fully observed in that year. Thus, a model that treats everyone the same, regardless of months of eligibility in the base year, would likely underpredict their ED use in the subsequent period. Future research on how best to adjust for this during model-building is needed, but was beyond the current scope of work. The problem is complex, both because people who transfer in or out of a plan mid-year are different from those with full-year eligibility, and because information on the presence of serious disease arrives in a non-linear way during the year, so that no simple adjustment (such as doubling what is seen during 6 months to impute what might have been seen during 12) will produce better results than what we did, which is to treat all observations the same, regardless of how many months of data are observed in the base year.

In contrast, one way to address partial-year eligibility when predicting and interpreting a utilization or cost outcome in the target year is to assume that each month of utilization represents 1/12th of what would have been seen had we observed for 12 months. The model is first used to calculate the expected outcome, E, for each person, using only data in the base year. After observing the target year, we define for each person: 1) a weight w = # of months eligible/12, and 2) an annualized outcome, in which observed utilization is multiplied by 1/w. That is, people with one ED visit during 6 months of eligibility would be treated as ½ of a person-year of observation, whose observed utilization reflects a utilization rate of 2 ED visits per year. Their data would contribute ½*2 to the numerator and ½*E to the denominator when constructing an observed-to-expected (O/E) ratio for the physician panels to which they belong. Unfortunately, we could not explore the consequences of partial-eligibility issues empirically in either of our datasets, since person-level eligibility fractions were not available.

Specific Aims

The Specific Aims of this study were to:

1. Calculate prevalence rates for overall and PCS ED use in two commercially insured populations. Identify associations between outcomes and enrollee and practice-level factors.

2. Create predictive models using administrative claims data (using both datasets) and calculate ED risk scores. Evaluate the performance of models predicting overall and PCS ED use.

3. Expand these models by adding enrollee characteristics from EMRs, neighborhood characteristics, and practice characteristics (MCN dataset only). Compare the performance of models predicting overall and PCS ED use. Determine additional predictors of each measure of ED use from the additional data sources.

Based on the review of the literature and the conceptual framework, we expected that greater perceived and actual medical need (as determined by prior-year medical claims, conditions in the problem list, and prior ED, PCP, and inpatient utilization), black race (from the MCN eligibility file), lower income (based on median income in the enrollee's Census tract), and proximity to the ED (based on the enrollee's home address and the address of the nearest hospital ED) would result in greater ED utilization.

Format of this Dissertation

This dissertation is presented in five chapters, plus appendices. Chapter I is the introduction, Chapters II – IV are standalone research papers, and Chapter V consists of discussion and conclusions, including strengths, limitations, and directions for future research. The first research paper, presented in Chapter II, focuses on the methods we used to explore different ED utilization outcomes in the MCN dataset, including the results of models predicting ED utilization at the practice level. The second research paper, Chapter III, focuses on the results of ED risk prediction models developed using the MarketScan administrative data. The third research paper, Chapter IV, describes the results of the enrollee-level ED risk prediction models developed using the MCN data, including the effects of adding additional characteristics from the EMR, Census tract, and payor/provider.

Chapter Summary

Patients are visiting EDs more often, for a wide range of reasons. Since care in the ED is usually uncoordinated, lacks follow-up, and is costly, and because EDs are overcrowded, reducing nonessential ED visits is important. The problem of predicting and measuring PCS ED visits has multiple dimensions. Methods for predicting such utilization have not been well defined, and it is not clear which variables and models predict best. A few characteristics have been found in more than one study to predict this type of use, including being female, over 65, African American, and covered by Medicaid. Studies have used a wide variety of methods for defining what constitutes a PCS ED visit, which complicates the question of how best to prevent such visits. Reducing undesirable types of healthcare utilization (including PCS ED use) requires the ability to define, measure, predict, and manage such use in a population.

CHAPTER II.
PREDICTIVE MODELING OF PRIMARY CARE SENSITIVE AND OVERALL EMERGENCY DEPARTMENT UTILIZATION USING ENHANCED ADMINISTRATIVE DATA: A COMPARISON OF ALTERNATIVE MEASURES AND METHODS

Abstract

BACKGROUND: Because of a demonstrated association between access to primary care and emergency department (ED) utilization, reducing patients' use of the ED has been proposed as a performance measure for primary care providers (PCPs).

OBJECTIVES: To examine alternative performance measures for primary care physicians based on their patients' use of the ED and evaluate practice-level observed-to-expected (O/E) ratios for each measure.

METHODS: In this retrospective, observational study, we included 64,623 individuals enrolled for at least 1 base-year month and one prediction-year month (in either 2009-10 or 2010-11) in 1 of 4 commercial insurance plans in Massachusetts who were assigned to a participating PCP in a managed care network. We used the NYU ED algorithm to assign a probability of being primary care sensitive (prob_PCS) to each ED visit, based on its principal diagnosis. Using claims data, we defined 5 ED-based outcome measures: 1) any ED visit; 2) total number of ED visits; 3) sum of prob_PCS over all ED visits; 4) any ED visits whose prob_PCS equals or exceeds .50; 5) same as 4), but using a 0.75 threshold. We compared these outcomes, examining the fraction of non-zeroes, the fraction of visits meeting or exceeding the threshold (0.50 or 0.75), the mean volume of visits counted, and other features of their distributions, such as skewness and suitability for analysis using linear or other regression models for establishing PCP-level benchmarks. We evaluated 45 practices (each with at least 100 patients) on their panel-average observed (O) and expected (E) outcomes, and calculated O/E ratios to identify practices with significantly higher or lower ED use than expected based on multiple regression models.

RESULTS: The practice-level analysis included 45 practices and 205 PCPs, with an average of 4.8 PCPs and 1,259 enrollees per practice (total n= 64,623). About 14.6% (±0.1%) of the sample had 1 or more ED visits during the prediction period, with an overall mean ED visit rate of 18.8 (±0.2) visits per 100 persons and 7.6 (±0.1) PCS ED visits per 100 persons. Measuring PCS ED use with a threshold-based approach resulted in many fewer visits counted as PCS, discarding information unnecessarily. Among 45 practices, 5 (11%) had observed values that were statistically significantly different from their expected values, based on the model predicting any ED visit. For the second outcome, number of ED visits, 11 (24%) practices had significant O/E ratios. For the third outcome, number of PCS ED visits, 9 (20%) practices had significant O/E ratios.

CONCLUSIONS: We have proposed and explored the characteristics of a new PCP performance measure based on the NYU algorithm. This measure addresses the concern that PCP penalties and rewards should be based on events that PCPs can influence, while retaining more information than previous algorithm-based measures that simply count the number of visits whose probability of being PCS exceeds an arbitrary threshold. In our data, the assessment method based on our PCS measure flagged fewer outliers than the one that counted all ED visits.

KEYWORDS: emergency department, claims analysis, risk drivers, utilization

Introduction

Most health care reform models (such as patient-centered medical homes and accountable care organizations) aim to reduce avoidable emergency department (ED) visits,[90] and many experts believe that the safest and surest way to do so is to improve access to primary care.[12] Thus, reducing ED visits has been proposed as a performance measure for primary care providers (PCPs). It has also been used as a measure of the success of patient-centered medical homes (PCMHs).[91] Better tools are needed to help PCPs and practices understand their population's use of the emergency department (ED) and identify ways to reduce visits to the ED that could have been prevented.

Generally, those seeking to evaluate a population's ED use have used a simple measure, such as percent of the population with any ED visit or the rate of ED visits per 100 persons. However, this measure does not distinguish between ED visits that were necessary and visits that could have been prevented, avoided, or handled in a less-acute setting, such as a PCP office. One alternative measure that has been used is frequent visits to the ED.[39] However, there is no agreement on the number of visits that defines a "frequent" user, few individuals are long-term frequent ED visitors, and frequent ED use is not necessarily an indication of problems with access to care.[39,41-43]

Another alternative is to focus on a subset of ED visits that may be preventable or avoidable. Researchers have used many different methods to categorize ED use, but most rely on triage or medical record data sources that are hard to access and analyze.[72] Only one validated method of categorizing ED visits using administrative data is widely available: the NYU ED algorithm, developed by Billings and colleagues.[17,68,92,93] This algorithm calculates a probability of being primary-care sensitive (PCS) for each ED visit, based on its principal ICD-9-CM diagnosis code. In prior published studies, analysts have counted a visit as “PCS” when this probability meets or exceeds a threshold value, such as 0.50, 0.75, or even 1.00.[13,47,65,68] Others – particularly emergency medicine practitioners – deny the validity of PCS determinations entirely, asserting that such efforts basically “scapegoat” ED users.[54,71,74,94]

In this study, our overall objective was to develop a measure for ED-related PCP profiling that is appropriate for its intended use. To accomplish this objective, we compared the performance of several alternative outcomes, including any ED use, total ED visits, and total PCS ED visits. We explored several different modeling approaches, evaluated a small set of primary care practices on their panel-average observed (O) and expected (E) outcomes, and calculated O/E ratios to identify practices with significantly higher or lower ED use than expected based on regression models controlling for multiple covariates.

Methods

Conceptual Framework

The conceptual framework for this study was adapted from Andersen’s behavioral model of health services utilization.[60] This model posits that utilization is influenced by need (e.g., number of medical conditions), predisposing (e.g., sociodemographic), and enabling (e.g., insurance and income) factors – both at the individual and societal level – as well as health behaviors. Further, primary care utilization has been shown to influence

ED use. We used this conceptual framework to help guide our selection of covariates in our ED risk prediction models and to aid in interpreting our findings.

Data Sources and Study Sample

Our data on enrollees and their primary care providers was provided by a managed-care network (MCN) in Massachusetts, which requested help in developing a performance measure for PCPs in the network based on practice-level ED use. The study was approved by the University of Massachusetts Medical School institutional review board. Through this partnership with the MCN, we obtained and merged several sources of data into a final deidentified analytic dataset.

The data included claims data for all inpatient, outpatient, ED, and office visits made by each eligible enrollee. Each paid claim, submitted by providers to one of the 4 insurance plans, contained up to 4 diagnostic (ICD-9-CM) codes. The MCN routinely merges all claims from these 4 plans into a single, harmonized data warehouse for quality improvement monitoring on a quarterly basis. The four plans are among the largest in Massachusetts by market share. Details on the types of coverage (HMO, PPO, etc.) were not available, but all were commercial (private) plans, either employer-sponsored or individually purchased.

The EMR system used in the MCN is Allscripts (Allscripts Healthcare Solutions, Inc., Chicago, IL), which was implemented in the MCN in 2007. The Allscripts system contains provider-entered information for each patient, including problem lists (conditions and complaints, history of illness, and presence or absence of risk factors such as tobacco use) and measured BMI and blood pressure. We restricted the providers included in the study to those who had transitioned from using paper medical records to using the EMR by 2009, the first year of our study period, in order to minimize missing EMR data.

Data on practice and provider characteristics were obtained from MCN administrative records. Practice characteristics included location, practice specialty, practice quality score, and provider quality score. The specialties were family practice, internal medicine, maternal/pediatrics, and multi-specialty/other. The practice-level quality scores were developed by the MCN to measure practices' performance under the terms of the Blue Cross-Blue Shield Alternative Quality Contract.[95] The provider quality scores were developed by the MCN as part of their internal rewards program, and were based on 21 quality and efficiency measures (primarily HEDIS measures, such as well-child visits; all measures are described in more detail in APPENDIX B).[86] We averaged these provider-level quality scores across three years (2009-11) to obtain a mean score for each provider and top-coded the resulting mean at the 99.5th percentile (refer to APPENDIX Cfor details on the top-coding procedure).

Data on neighborhood characteristics were obtained by first mapping each enrollee's address to a Census tract, then merging the enrollee data with the Census-tract-level data elements, such as median income, percent under the poverty line, and percent homeowners. After linking addresses to Census tracts and enrollment data to clinical and provider data, all records were deidentified.

All included persons were enrolled in one of 4 commercial insurance plans for at least 1 month in either 2009 or 2010 and at least one month in the subsequent year (2010 or 2011). We included all ages. The enrollees were split into two groups: development (those enrolled in 2010-11) and validation (2009-10).

Enrollees were all affiliated with one of the 235 primary care providers (PCPs) in the MCN who had transitioned to the Allscripts EMR system. Each enrollee's affiliation with a PCP was determined based on the PCP listed in that enrollee's EMR. Among enrollees who had more than one PCP in their records, we selected the match with greatest number of encounters first, then the last visited provider in the appropriate base year. We excluded 5 enrollees who could not be matched to just one provider. After excluding individuals who lived outside of Massachusetts and those who could not be matched to a participating PCP, the dataset included 64,623 unique enrollees (107,449 observations).

Measures

The claims data from each enrollee's base period (2009 for the validation sample and 2010 for the development sample) were used to calculate two morbidity scores using DxCG Intelligence version 4.1 (Verisk Analytics, Jersey City, NJ). Although DxCG models were originally developed to predict total costs, they are often used as a summary morbidity measure that is associated with other health outcomes, including utilization of specific services such as the ED.[96-98] The concurrent morbidity score used current-year claims to predict current-year costs, and the prospective morbidity score used current-year claims to predict next-year costs.

We defined 5 ED-based outcome measures. The first two outcome measures included all ED visits, and were specified as follows: 1) a binary indicator for any ED visit during the prediction period; and 2) the number of ED visits over the 12-month prediction period. We then focused on a subset of ED visits thought to be potentially avoidable.

We used principal ICD-9-CM diagnosis codes and the New York University ED algorithm to classify each outpatient ED visit. Although diagnosis codes change from year to year, the original algorithm has not been updated since 2003. To reduce the number of diagnosis codes that could not be classified, we used a version developed in 2009 by the Massachusetts Center for Health Information and Analysis (CHIA) (personal communication, April 30, 2013). CHIA's version of the algorithm built on the original to incorporate new codes with input from the original developer and an emergency medicine physician, but did not involve new data abstractions.[50] When we applied the CHIA-updated algorithm to our 2010 sample, it reduced the percentage of unclassified claims from 14.8% to 10.1%. The updated algorithm is available from the authors.

The 4 main algorithm categories are: 1) nonemergent, 2) emergent but primary care treatable, 3) emergent but preventable/avoidable, and 4) emergent, not preventable/avoidable **(Figure 1-3)**. To calculate our third outcome, we sum the values of the first 3 categories to create a number between 0 and 1 for each visit, referred to as prob_PCS; for each person, we sum all of the prob_PCS values from all ED visits during the year to derive the total number of PCS visits.

The equation is a weighted sum of ED visits,

$$\text{Total PCS ED visits} = \sum (w_j * ED_j)$$

where *w* is derived from the NYU ED algorithm's probabilities associated with each diagnosis code, and *j* is an individual.

For example, as shown in **Table 2-1**, Enrollee A had 3 ED visits in 2010: 2 for palpitations (ICD-9-CM code 785.10) and 1 for other chest pain (786.59). Palpitations is assigned a prob_PCS of 0.44, and chest pain is 0.61. Summing the prob_PCS across all 3 visits yields a total of 1.49 PCS visits in 2010 for that enrollee.

Table 2 - 1. Examples of NYU ED algorithm classifications for selected enrollees, MCN development data, 2011

Enrollee	ED Visit Principal Diagnosis Codes and Descriptions	Algorithm Classifications NE	PCT	PA	NPA	Prob_PCS	Total Number of PCS Visits
A	785.10 - Palpitations	0	0.44	0	0.56	0.44	1.49
	786.59 - Other chest pain	0	0.61	0	0.39	0.61	
	785.10 - Palpitations	0	0.44	0	0.56	0.44	
B	346.90 - Migraine, unspecified	0.78	0.09	0	0.13	0.87	3.49
	784.00 - Headache	0.78	0.09	0	0.13	0.87	
	729.50 - Pain in limb	0.71	0.17	0	0.13	0.88	
	346.90 - Migraine, unspecified	0.78	0.09	0	0.13	0.87	
C	789.03 - Abdominal pain, right lower quadrant	0	0.67	0	0.33	0.67	0.99
	786.50 - Chest pain, unspecified	0	0.32	0	0.68	0.32	
D	493.90 - Asthma, unspecified	0	0.02	0.98	0	1.00	1.75
	620.20 - Ovarian cyst, other/unspecified	0.25	0.5	0	0.25	0.75	

NE: nonemergent; PCT: primary-care treatable; PA: preventable/avoidable; NPA: not preventable/avoidable; PCS: primary care sensitive; MCN: managed care network

This table shows how the NYU ED algorithm assigns probabilities for each principal diagnosis code associated with an ED visit, and how they can be summed to provide a total probability of being PCS for each visit and for all of a person's ED visits during a period. The sum of the probabilities in algorithm categories NE, PCT, and PA constitute the prob_PCS value. For example, Enrollee B had 2 visits for migraine, each of which were assigned a prob_PCS of 0.78 + 0.09 = 0.87; summing all the Prob_PCS values for Enrollee B results in a total of 3.49 PCS visits. Enrollee D had 1 visit for asthma, which was assigned a prob_PCS of 0.02 + 0.98 = 1; summing all the Prob_PCS values for Enrollee D results in a total of 1.75 PCS visits.

The algorithm also creates indicators for whether a visit's principal diagnosis code was unclassifiable by the algorithm, injury-related, alcohol- or drug-related, or mental-health related; visits in those 4 categories are assigned a 0 or 1, not a prob_PCS value, and are not included in the PCS outcomes (please see APPENDIX A for further details on unclassifiable visits).

The final two outcome measures were as follows: 4) the number of PCS ED visits where prob_PCS equals or exceeds .50; and 5) the same as 4), but using a 0.75 threshold. These thresholds have been recommended in prior studies as potential cutoffs.[13,65,68] An even more restrictive method – only categorizing visits as PCS if the algorithm predicted a 100% probability that the diagnosis related to that visit was PCS[47] – was not evaluated because only 3% of cases met those criteria, leading to an unacceptable loss of information.

Returning to our 4 individual enrollee examples from Table 2-1 above, in **Table 2-2** below we show how each outcome would be calculated for each enrollee. Note that only the number of ED visits and number of PCS ED visits reflect the magnitude of ED use seen with multiple ED visits. Although the threshold methods could also be used to count multiple visits, they would still be less sensitive than the total PCS ED use measure.

Table 2 - 2. Examples of outcome measure calculations for selected enrollees, MCN development data, 2011 (prediction year)

Enrollee	ED Visit Principal Diagnosis Codes and Descriptions	Overall ED Visits: 1 - Any ED Visit	Overall ED Visits: 2 - Total ED Visits	PCS ED Visits: 3 - Total PCS ED Visits	PCS ED Visits: 4 - Any PCS Visit, .50 Threshold	PCS ED Visits: 5 - Any PCS Visit, .75 Threshold
A	785.10 – Palpitations 786.59 - Other chest pain 785.10 – Palpitations	1	3	1.5	1	0
B	346.90 - Migraine, unspecified 784.00 – Headache 729.50 - Pain in limb 346.90 - Migraine, unspecified	1	4	3.5	1	1
C	789.03 - Abdominal pain, right lower quadrant 786.50 - Chest pain, unspecified	1	2	1	1	0
D	493.90 - Asthma, unspecified 620.20 - Ovarian cyst, other/unspecified	1	2	1.8	1	1

Source: Development sample; MCN: managed care network

Among the 3 PCS measures considered, the 2 threshold-based ones have several undesirable properties. Measuring PCS ED use with a threshold-based approach results in many fewer visits counted as PCS, discarding information unnecessarily. The approach also requires the analyst to arbitrarily choose a threshold, and the choice of threshold strongly affects the outcome distribution. Because of the limitations associated with these methods of measuring PCS ED use, in the remainder of analyses, we measured PCS outcomes using only total PCS ED visits, which sums the probabilities associated with PCS ED use and does not discard any visits that the algorithm sees as potentially PCS.

Statistical Analyses

After conducting basic descriptive analyses of the enrollees and practices, we used a two-sample analytic strategy to build and validate predictive models.[99] We used several model development strategies, including forced entry of all potential risk factors, as well as forward and backward stepwise selection. We retained all risk factors that were statistically significant ($P < .05$) in one or more of those approaches in each final model.

Analyses were conducted in Stata/IC version 11.2 (Stata Corp., College Station, TX) and SAS version 9.2 (SAS Institute, Research Triangle Park, NC).

The development sample consisted of enrollees from 2010-11, and the validation sample included enrollees from 2009-10. In the development sample, we examined overall visit rates and compared the 5 outcome measure candidates, examining the fraction of non-zeroes, the fraction of visits that met or exceeded a threshold cusp (0.50 or 0.75), the mean volume of visits counted, and other features of their distributions, such as minimum, maximum, and skewedness, coefficient of variation, and correlation with the other outcome measures. We also calculated correlation statistics among the outcome measures and among the predictor variables to check for potential collinearity.

We constructed multivariable models to predict 3 different measures of ED use. Initial models were specified using **need factors** (whether individuals had any ED, inpatient, and office visits in the base period and a prospective morbidity score derived from diagnosis codes in the base period), **predisposing factors** (age, sex, and race), and an **enabling factor** (payor), all from administrative data. Subsequent models adjusted for additional **need factors** (problem list conditions recorded in the EMR) and **enabling factors** (neighborhood and practice/provider characteristics from the Census and the MCN's administrative records, respectively). We then evaluated improvements in model fit and performance resulting from the addition of these covariates.

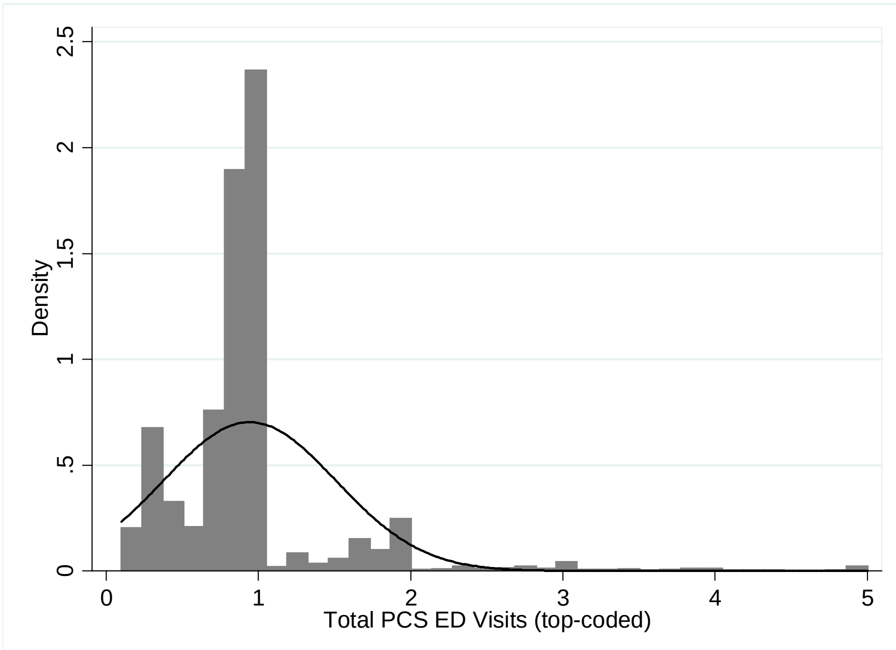
In order to limit the effect of outliers on the analysis, we top-coded all continuous variables, including outcome variables and predictor variables (BMI, distance to the ED, distance to the PCP, diastolic and systolic blood pressure readings, Census tract-level characteristics [median home value, median income, median rent, and travel time to work], morbidity scores, and provider quality scores) at the 99.5th percentile. APPENDIX C provides details on the variables that were top-coded.

We predicted 3 distinct outcomes. For outcome 1—any ED visit, we used a logistic model. We used a zero-inflated negative binomial (ZINB) model for outcome 2—total ED visits. ZINB models are used to model outcomes that are counts of an event where there are many people with a count of zero and the excess zeroes can be modeled separately.[100]

For outcome 3—total PCS ED visits, we evaluated several modeling estimators, including several specifications of generalized linear models (GLMs), as well as a two-part (hurdle) model, consisting of a logit in the first part to predict any ED visit, and an ordinary least squares (OLS) regression model in the second part to predict the number of PCS ED visits among those with any ED visit.[101] GLM regression is an extension of ordinary linear regression that can be used to model outcome variables that have distributions other than the normal distribution; the model specification process requires determining the link function (e.g., identity, log, or negative binomial), and the distribution of the outcome variable (e.g., normal [Gaussian], count [Poisson], or gamma).[102]

Unlike number of ED visits, total PCS ED visits is not integer-valued; also, it is far from normally distributed. In fact, its distribution in our sample was roughly tri-modal (**Figure 2-1**). Despite this, the generalized linear model (GLM) and the two-part (hurdle) logit-OLS model resulted in similar estimates (not shown). Since the modeling approach did not appear to affect the results in any substantive way, we present the results of the hurdle model here.

Figure 2 - 1. Distribution of Total PCS ED visits, MCN data



Source: Development sample (n=53,112); ED: emergency department; MCN: managed care network; PCS: primary-care sensitive. The black line represents the normal distribution curve.

Analysis of Practice-level Outcomes

In this paper, we examine practice-level variations in outcomes. We included practices with at least 100 enrollees in their panels (which reduced the number of practices from 54 to 45). We examined differences among practices in their panel's average observed outcomes, in their expected outcomes based on multivariable regression models, and in their observed-to-expected (O/E) ratios. We created a pooled estimate of the standard deviation (SD) of an outcome from its prediction by first taking the square root of the sum of the variances for each practice using weights based on practice size (number of enrollees in the development sample).[103] We used this to generate standard errors for each practice's expected outcomes.

Results

The final development sample included 53,112 observations, and the validation sample, 54,337; combined, the dataset included 64,623 unique individuals and 107,449 observations. The two samples were similar on most measures, although statistically significant differences were present within some age, race, and practice/payor groups (**Table 2-3**).

Table 2 - 3. Sociodemographic characteristics, MCN data, 2009-10 (base years)

	Development (2010)		Validation (2009)		*P*-value
	N enrollees	%	N enrollees	%	
N	53,112		54,337		
Age (as of Jan. 1 of base year)					<.001
< 1	920	1.7%	749	1.4%	
1-10	5,889	11.1%	5,178	9.5%	
11-17	4,555	8.6%	4,187	7.7%	
18-24	5,075	9.6%	4,978	9.2%	
25-39	11,516	21.7%	10,641	19.6%	
40-64	25,237	47.5%	24,770	45.6%	
65+	1,145	2.2%	2,609	4.8%	
Female	27,983	52.7%	27,199	50.1%	.344
Race/ethnicity					.006
White	38,075	71.7%	36,947	68.0%	
Black	1,362	2.6%	1,323	2.4%	
Other	2,771	5.2%	2,563	4.7%	
Unknown	12,129	22.8%	12,279	22.6%	
Neighborhood income category					.089
Low (<200% FPT)	3,528	6.7%	3,585	6.6%	
Middle (200-399% FPT)	31,947	60.2%	32,365	59.6%	
High (400+ FPT)	17,637	33.2%	18,387	33.8%	
PCP type					<.001
Internal medicine	25,700	48.4%	25,657	47.2%	
Family medicine	14,278	26.9%	15,229	28.0%	
Maternal/pediatrics	13,056	24.6%	13,340	24.6%	
Other	78	0.2%	111	0.2	
Payor					<.001
Plan 1	29,935	55.1%	26,791	50.4%	
Plan 2	8,578	15.8%	9,585	18.0%	
Plan 3	10,265	18.9%	10,819	20.4%	
Plan 4	5,559	10.2%	5,917	11.1%	

FPT: Federal poverty threshold; PCP: primary care provider; MCN: managed care network

About 14.7% of the sample had any ED visit in the prediction year, with a mean visit rate of 18.9 visits per 100 persons (**Table 2-4**). When PCS ED visits were defined using the sum of prob_PCS method, the visit rate decreased to approximately 7.6 per 100; it dropped to 6.8 per 100 when using a .50 threshold to define PCS visits, and 5.7 per 100 when using a .75 threshold. Outcomes 4 and 5, the threshold-based PCS measures, have only 47% and 39% as many non-zero observations as the first outcome measure (any ED visit), respectively.

Each of the outcomes is skewed, with the number of ED visits and number of PCS ED visits being most strongly skewed, especially before top-coding at the 99.5th

percentile (see APPENDIX C for details on the top-coding procedure). Top-coding changed the mean number of ED visits from 0.189 to 0.181 and the mean number of PCS ED visits from 0.076 to 0.072. The correlations between each of the outcomes are shown in **Table 2-4**.

Table 2 - 4. Emergency department utilization using 5 different measures, MCN development data, 2011 (prediction year)

	Overall ED Visits		PCS ED Visits		
	1 - Any ED Visit	2- Total ED Visits	3 - Total PCS ED Visits	4 - Any PCS Visit, .50 Threshold	5 - Any PCS Visit, .75 Threshold
Original ED Visit Data					
Number with Y > 0	7,783	7,783	4,266	3,627	3,005
Mean	0.15	0.19	0.08	0.07	0.06
Standard deviation (SD)	0.35	0.56	0.32	0.25	0.23
Skewness	2.00	8.07	8.75	3.42	3.84
Minimum	0	0	0	0	0
Maximum	1	30.0	16.2	1	1
Visit rate per 100 persons	14.65	18.86	8.03	6.83	5.66
Coefficient of variation (CV)	2.41	2.99	4.15	3.69	4.08
Correlation coefficients					
With outcome 1	--				
With outcome 2	0.81	--			
With outcome 3	0.58	0.76	--		
With outcome 4	0.64	0.62	0.84	--	
With outcome 5	0.59	0.58	0.83	0.90	--
Top-coded ED Visit Data					
Number with Y > 0	NC	NC	NC	NC	NC
Mean	NC	0.18	0.07	NC	NC
Standard deviation (SD)	NC	0.48	0.27	NC	NC
Skewness	NC	3.11	4.13	NC	NC
Minimum	NC	NC	NC	NC	NC
Maximum	NC	3	1.88	NC	NC
Visit rate per 100 persons	NC	18.09	7.20	NC	NC
Coefficient of variation (CV)	NC	2.67	3.73	NC	NC
Correlation coefficients					
With outcome 1	--				
With outcome 2	0.91	--			
With outcome 3	0.65	0.73	--		
With outcome 4	NC	NC	0.94	--	NC
With outcome 5	NC	NC	NC	NC	--

Source: Development sample (n=53,112). NC: No change; MCN: managed care network

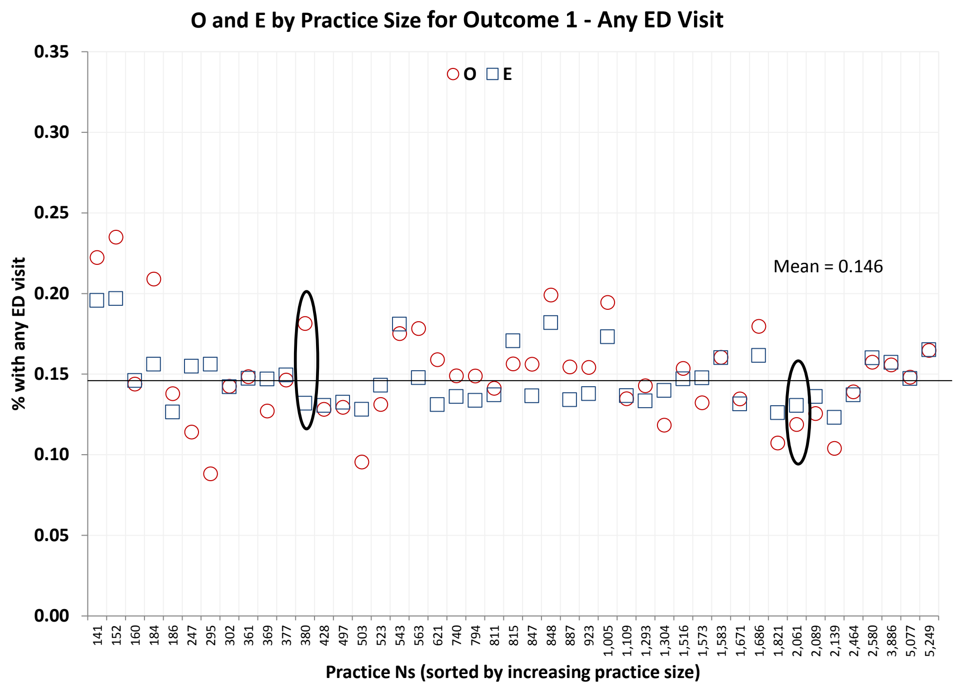
Practice Observed-to-Expected Ratios

The final practice-level sample included 45 practices (each with at least 100 enrollees) and 205 PCPs, with an average of 4.8 PCPs and 1,259 enrollees per practice. **Figures 2-2, 2-3**, and **2-4** show, in **Panel A** (top)**,** mean observed (O) and expected (E) values by practice, sorted from smallest to largest and, in **Panel B** (bottom), O/E ratios by practice, sorted from smallest to largest, for each of the 3 outcomes, respectively. We sorted by practice size to draw attention to the fact that predicted and actual values tend to converge as practice size increases.

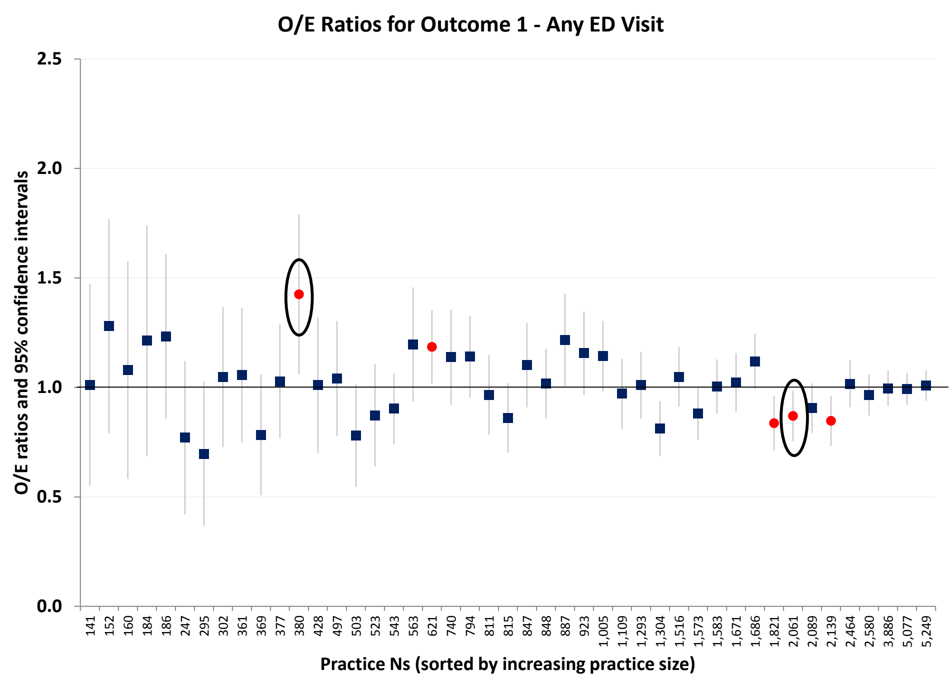
Two practices are highlighted with black ovals in the figures: Practice X, with 380 enrollees; and Practice Y, with 2,061 enrollees. Both are family medicine practices. On the first outcome, any ED visit, Practice X had an observed ED visit rate of 0.181 per 100 persons (higher than average, as indicated by the black horizontal line at 0.146). Based on enrollee characteristics and our regression model, their expected ED visit rate was 0.132. The O/E ratio (95% CI) was 1.425 (1.061 to 1.790). Practice Y's observed rate was 0.119 (lower than average), the expected was 0.130, and the O/E ratio (95% CI) was 0.870 (0.755 to 0.985). Thus, both practices had significant O/E ratios, as indicated by the fact that the 95% confidence intervals do not cross 1. In other words, Practice X's observed ED visit rate was significantly higher than would have been expected based on its enrollees' characteristics, whereas Practice Y's was significantly lower than expected. Out of 45 practices, 5 (11%) had observed values that were statistically significantly different from the expected value, based on the model predicting any ED visit.

Figure 2 - 2. Outcome 1: mean observed and expected overall emergency department visit rates (top) and O/E ratios (bottom) by practice

Panel A



Panel B

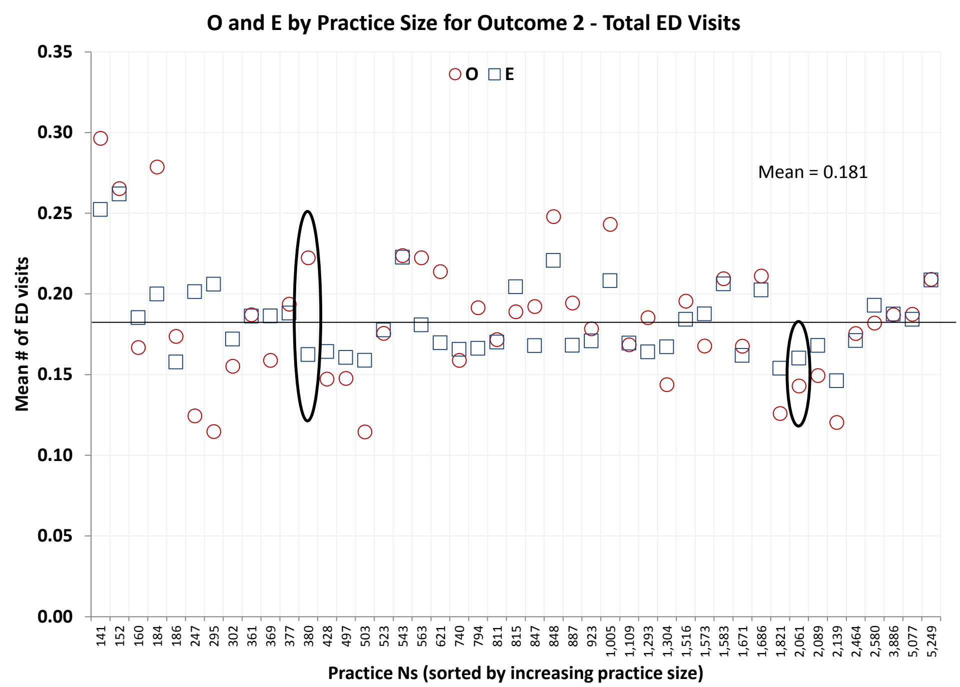


Source: Development data (n=53,112). Includes 45 practices with ≥ 100 enrollees.

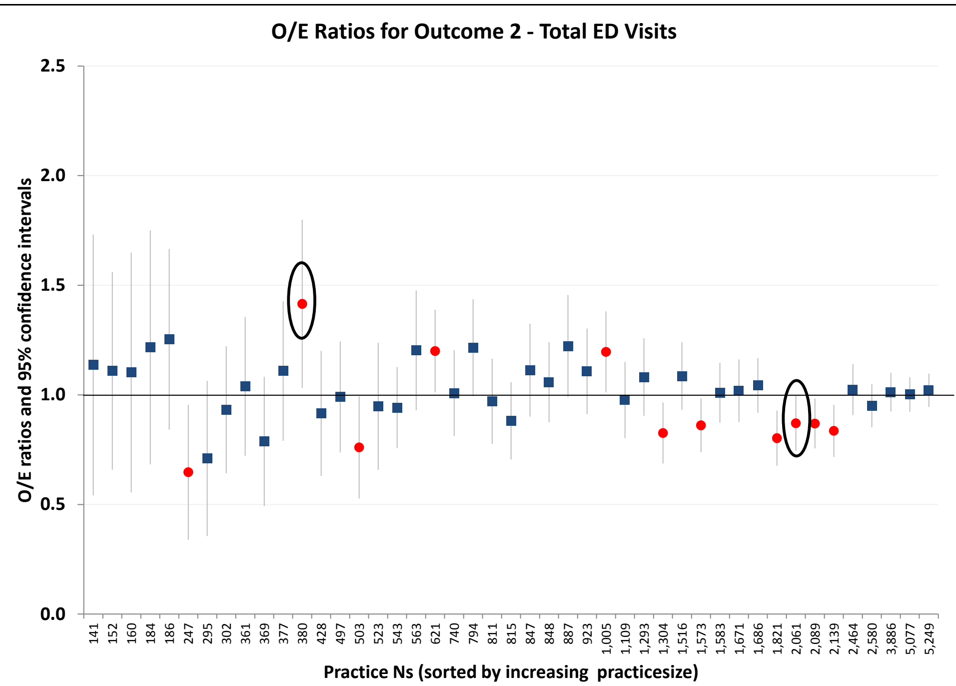
Note: Red circles indicate practices that differ significantly from 1. Hollow ovals highlight the two family medicine practices discussed in the text.

Figure 2 - 3. Outcome 2: mean observed and expected total number of emergency department visits (top) and O/E ratios (bottom) by practice

Panel A



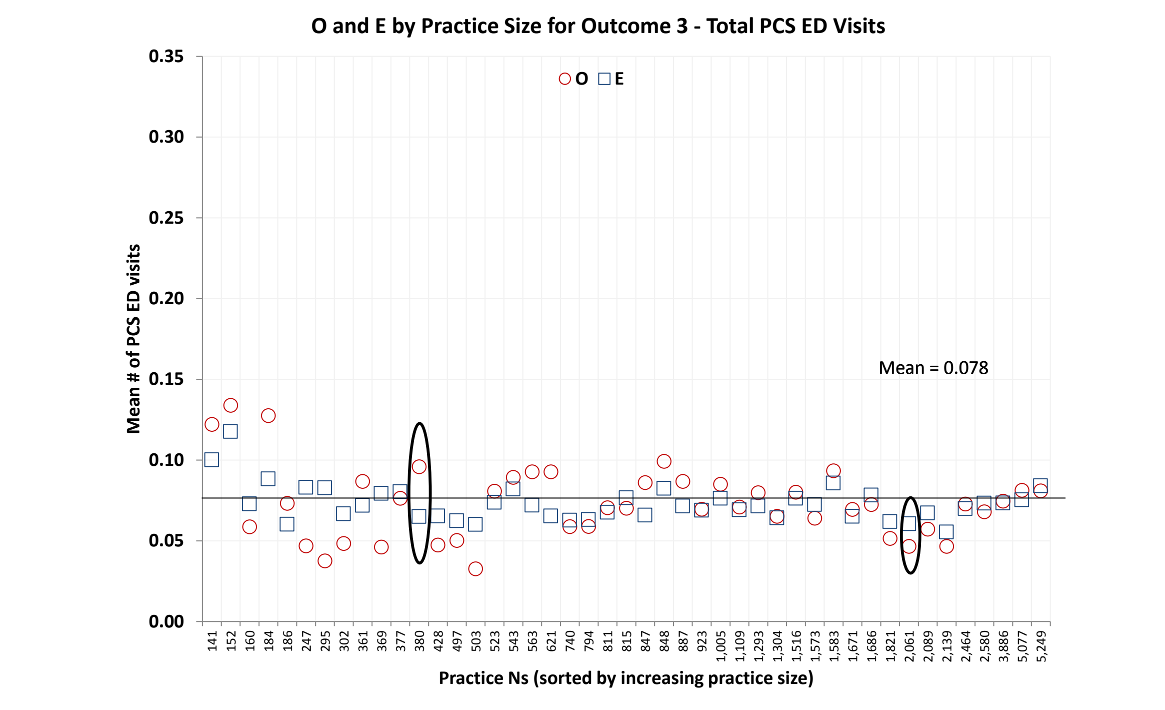
Panel B



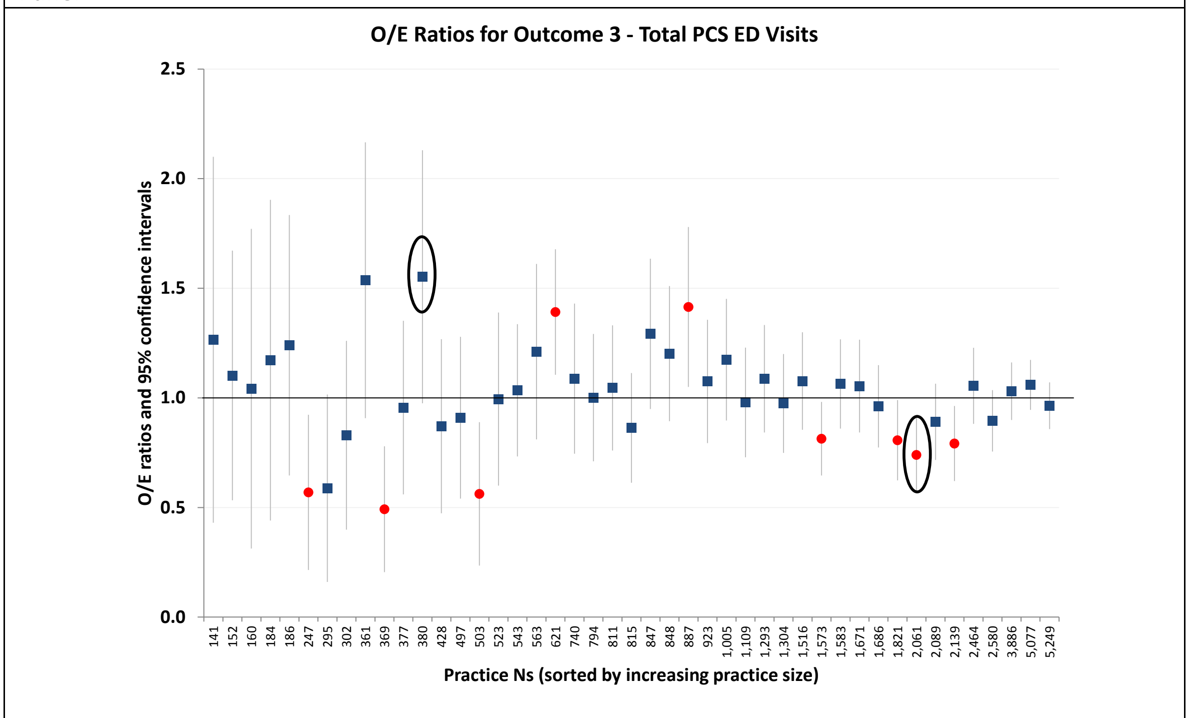
Source: Development data (n=53,112). Includes 45 practices with ≥ 100 enrollees.Note: Red circles indicate practices that differ significantly from 1. Hollow ovals highlight the two family medicine practices discussed in the text.

Figure 2 - 4. Outcome 3: mean observed and expected number of PCS emergency department visits (top) and O/E ratios (bottom) by practice

Panel A



Panel B



Source: Development data (n=53,112). Includes 45 practices with ≥ 100 enrollees. Note: Red circles indicate practices that differ significantly from 1. Hollow ovals highlight the two family medicine practices discussed in the text.

For the second outcome, number of ED visits, 11 (24%) practices had significant O/E ratios, including both Practice X (O: 0.222, E: 0.162, O/E: 1.415 [1.031 to 1.799]) and Practice Y (O: 0.143, E: 0.160, O/E: 0.871 [0.745 to 0.996].

For the third outcome, number of PCS ED visits, 9 (20%) practices had significant O/E ratios. Practice X again had a higher than expected number of visits (O: 0.096, E: 0.065) but the O/E ratio (1.552 [0.975 to 2.129]) was not significantly different from 1. Practice Y's observed utilization was again significantly lower than expected (O: 0.046, E: 0.060, O/E: 0.740 [0.573 to 0.907].

Four of the practices with significant O/E ratios on any ED visit were also significantly different on both total ED visits and total PCS ED visits. Four practices were significantly different on 2 measures, and 5 were significantly different on only 1 measure.

Discussion

In this study, we have described several measures of ED use, including a measure of primary-care sensitive ED use that sums the probability of a visit being PCS, as assigned by the NYU ED algorithm, across all visits in a year. This outcome is contrasted to earlier methods using thresholds, which discard information, inflate the number of zero outcomes, and depend upon arbitrary threshold choices that can strongly affect the outcome distribution. Applying dichotomies loses information and is not necessary.

Whether individuals had any ED utilization during a year and their total number of ED visits per year are common measures of a population's use of the ED, but both have drawbacks. Measuring only whether individuals had any ED use does not allow us to examine magnitudes or variations in intensity, and neither method takes into account whether a visit was potentially avoidable.

The sum of probabilities measure described in this paper offers several advantages over other measures: it provides an indication of magnitude, loses little information compared to counting all ED visits equally, and addresses concerns about penalties based on events not under PCP control (such as injuries, poisonings, and other emergency conditions). In terms of disadvantages, it is a continuous, non-integer value, and Poisson and negative binomial regression models are designed for integer-valued data. From a practical perspective, modeling the total PCS ED visits outcome requires more statistical expertise than modeling a simple binary or count outcome, which could limit the real-world diffusion of the measure. However, a relatively straightforward two-part model, using a logistic in the first part and either OLS or GLM in the second, accommodates the skewness.

Importantly, much of the practice-level variation in O (but not O/E) is explained by E, thus making the case for the importance of setting risk-based targets for performance assessment. While differences between Os and Es tended to decrease with increasing practice size, even among practices with more than 1,000 patients, nearly half (7 of 18) had Os that differed from their Es by 10% or more. In addition, while the smallest practices were more variable on both O and E, some larger practices also had Es that were quite different from average.

Little prior research has been published on using observed-to-expected ratios to measure physician performance on utilization. Ash and Ellis described one application of the O/E method in their 2012 article on risk-adjusted payment and performance adjustment for PCPs.[96] It is also discussed in the 4th edition of *Risk Adjustment for Measuring Healthcare Outcomes.*[103] The O/E ratio has also been used in national surgical quality improvement programs and to monitor surgical mortality.[104,105]

Many prior articles have been published on measuring emergency department utilization, but such articles have typically focused on measuring and predicting ED crowding and future demand,[106] defining the "appropriateness" of particular ED visits,[79,107,108] or counting numbers of ED visits to define "frequent" visitors.[30,39,41,109] We are not aware of any prior publications that use predictive modeling of ED use for performance measurement and case management.

Limitations

This study included only persons insured by one of 4 commercial insurers in Massachusetts who received care from one managed care network. The population was mostly white and relatively healthy, with lower ED use than the state average. Massachusetts is also a relatively wealthy state, with 98% of the population covered by health insurance. Thus, our findings may not generalize broadly.

Another important limitation is that the version of the NYU ED algorithm we used was last updated in 2009.There may have been changes in the ICD-9-CM diagnosis codes used in medical billing since 2009 that the algorithm could not capture. Roughly 10% of ED visits in our sample were unclassifiable, and thus excluded from our analysis.

Moreover, using diagnoses to classify individual visits has inherent limitations because coding practices vary among providers.

One feature of our analysis was that we treated people with partial-year observations the same as those with full-year observations. We know, based on the MCN's internal records, that most enrollees were present for all 12 months of both the base and prediction periods, and the vast majority had at least 6 months of observation in each period. However, problems in file construction meant that we were not able to incorporate the number of months present for each enrollee in our analysis.

Partial-year observations have several implications for practical implementation of these methods. For retrospective analyses (to set benchmarks for providers based on two years of recent data for the purposes of performance measurement and quality improvement efforts), treating people with partial-year observations during the target year the same as those with full-year observations provides an unfair advantage for providers whose patients' ED use is observed for fewer months per person. Specifically, the more fully observed the patient base, the higher the O/E ratio. To illustrate this, **Table 2-5** considers a hypothetical pair of providers. Provider A has 200 identical patients in her panel; 50% of them have data for 6 months, and 50% have data for 12 months. Provider B has 200 identical patients in his panel, and 100% of them have 12-month data. If each group of patients had exactly average ED utilization as the MCN development sample, thus having identical observed and expected use rates, Provider A's O/E ratio would be substantially lower than Provider B's simply because she has fewer patients with complete data. Because we observe only half the utilization in half of the panel, this unexceptional provider gets inappropriate credit for lower-than-expected utilization.

Table 2 - 5. Comparison of two hypothetical practices with different distributions of eligibility months in the target year

	N	Observed use rate	Expected use rate	O/E ratio
Provider A				
Patients with 6 months of data	100	0.073	0.147	0.5
Patients with 12 months of data	100	0.147	0.147	1
Total	200	0.110	0.147	0.75
Provider B				
Patients with 12 months of data	200	0.147	0.147	1

Future research on implementing a PCS ED use performance measure would benefit from a partial-eligibility data analysis. One option for handling unobserved utilization during part of the target year is to calculate annualized rates using the number of months of eligibility divided by the number of months in the prediction year as a weight (i.e., eligibility fractions).[110] This posits a constant rate of ED use throughout the year, which may be reasonable absent a strong seasonality effect (a question that could be explored in future research). Seasonality could be addressed using different weights for different missing months. Using eligibility fractions to weight observed utilization can also result in large outliers. For example, an individual with only 1 month of enrollment who went to the ED twice in that month is treated as 1/12 of a full-year observation, accumulating ED visits at the rate of 24 per year.

Predictive models meant to be used for case management would be built in a similar manner to our models, since analysts in real-time situations do not know, in advance, for how many months in the future an enrollee will remain enrolled. However, having missing months of eligibility in the base year presents a problem when calculating risk scores. If we are missing months of observation, this may bias the risk scores downward (underestimating morbidity, and thus risk of future expenditures and/or utilization). On the other hand, different reasons for a person being present for only part of the base year make it difficult to find simple relationships between risk scores calculated from 12 months of data versus those calculated on fewer months. We know of no published research that provides a solution for dealing with partial-year eligibility in the base year in any better way than we have here. This is another area deserving of future research.

We chose to focus only on outpatient ED visits in this analysis (i.e., visits by those who were not admitted to the hospital after their ED visit), since ED visits on the path to an inpatient stay are generally thought to be unavoidable.[52,53] Strictly speaking, however, it is not necessary to exclude visits on the path to an admission if the NYU ED algorithm works as designed. Hospital stays that start with a PCS ED visit would typically be preventable (ambulatory care sensitive [ACS]) hospitalizations, and thus could be legitimately classified as potentially preventable (PCS). However, most non-preventable hospitalizations would be likely to start with an ED visit that should be categorized as emergent, not preventable/avoidable (i.e., not PCS).

When we exclude ED visits that result in an inpatient stay, we are likely to see more nonemergent and primary care treatable ED use, and less preventable/avoidable (ACS). The ACS definition, which the NYU ED Algorithm used to define the preventable/avoidable category, includes inpatient stays that are for angina, asthma, chronic obstructive pulmonary disease [COPD], diabetes, grand mal status and other epileptic convulsions, heart failure and pulmonary edema, and hypertension.[111] The specifications also include visits for referral-sensitive conditions that PCPs may be less able to influence, such as hip replacements and pacemaker insertions.[112] Therefore, excluding ED visits that result in an inpatient stay has implications that deserve further study.

Models designed to measure or predict ED use could we applied in various ways. First, such models could be used to set expected-use targets for providers (benchmarks based on providers' own patients). Our practice-level analysis shows that factors largely outside the control of the provider contribute meaningfully to the practice-level variation in ED use. This illustrates the importance of risk adjustment in developing such targets for providers.

Second, predictive models of ED use can be used to identify high-risk enrollees for case management, clinical, or educational interventions.[42,113-115] Near real-time data availability is needed for case management. Of course, individual outcomes are highly variable and difficult to predict, and individual predictions are, at best, only correct on average. One approach would be to use the model to target only enrollees in the top 99.5th or 99.9th percentile of model-predicted risk. In that small subgroup of enrollees at highest risk, case management efforts would be most likely to be concentrated on a group that, absent intervention, would be likely to incur disproportionate numbers of ED visits.

Finally, predictive models of ED utilization could be used to examine the impact of changes in practice delivery, such as transitioning to patient-centered medical homes (PCMHs). To our knowledge, the PCMH demonstrations that have evaluated ED use to date have used only the simplest measure: percent with any ED use.[59,89,104] Measuring PCS ED use instead (or in addition) may be more likely to identify changes in ED utilization that could be attributed to changes in primary care access and quality.

The choice of which variables to include in a predictive model of ED utilization will vary based on the model's intended use. For example, physician profiling and performance assessment models incorporating prior use of the ED and prior costs could have the unintended effect of setting a low bar for providers who have historically provided sub-standard care, since their targets will be based on their previous efforts. Instead, basing such models only on enrollee demographics and morbidity will address variation within panels without biasing the targets. In contrast, analysts want models predicting ED use for case management purposes to be maximally predictive, explanatory models, using every variable that helps improve prediction.

In conclusion, current measures of ED utilization, such as simple binary indicators of any ED visit or counting the number of ED visits in a year, provide less relevant information to policymakers and administrators than our proposed measure, which examines the subset of ED visits that are potentially sensitive to primary care. Others have proposed using thresholds based on probabilities assigned by the NYU ED algorithm to create binary indicators of whether certain visits were PCS or not. This method is problematic because few visits are assigned a 100% probability of being PCS; the choice of which threshold to use is arbitrary; and when visits below the threshold are discarded, the analyst loses important information.

Therefore, we propose a method that sums the probabilities generated by the algorithm across all ED visits by each person, and then calls that sum the estimated number of PCS ED visits. In our data, the assessment method based on our PCS measure flagged fewer outliers than the one that counted all ED visits.

CHAPTER III.
RISK-ADJUSTED PREDICTIVE MODELS OF EMERGENCY DEPARTMENT UTILIZATION BASED ON ADMINISTRATIVE DATA

Preface

We gratefully acknowledge the contributions of Zoe (Mengyao) Zhao, MS, a research analyst at Verisk Health, who assisted with SAS programming.

Abstract

BACKGROUND: Many emergency department (ED) visits could be avoided with high quality primary care. A model to predict ED use could be useful for identifying patients at high risk for ED visits and rewarding (penalizing) primary care providers whose patients use the ED less (more) than expected.

OBJECTIVE: To construct models predicting ED utilization using administrative data, identify individuals at high risk of any ED visit, and estimate the number of PCS ED visits that groups of enrollees are expected to incur.

METHODS: Retrospective, observational study using MarketScan claims data from 2007 (n=15,136,261) as baseline information to predict: 1) the likelihood of any ED use in the first 6 months of 2008; and 2) the number of primary-care sensitive (PCS) ED visits. We used the New York University ED algorithm to quantify PCS ED utilization and multivariable logistic and ordinary least-squares regression modeling with splining to estimate the probability of ED utilization, controlling for age, sex, morbidity, prior-year ED visits by quarter, and costs. We also generated top risk groups (those predicted to be at highest risk) according to the model-predicted likelihood of ED use for each measure, using cutpoints of 99.5%, 99%, 97.5%, 95%, and 90%. Within each top risk group, for each measure, we evaluated sensitivity, specificity, and positive predictive value (PPV).

RESULTS: In the first 6 months of 2008, 10.6% of enrollees had at least one ED visit, with about half of utilization scored as PCS. Among the 0.5% of the population predicted to be at greatest risk of any ED use (the top risk group), 49.7% had at least one ED visit in the prediction period, and 40.5% had any PCS ED use. For the top risk group, the model's sensitivity was 3.1% and specificity was 99.7%. A splined OLS model predicting PCS visits yielded sensitivity of 3.8% and specificity of 99.7% for the top risk group.

CONCLUSIONS. Prediction models using administrative data may be used to help identify enrollees at high risk of ED visits and estimate the number of PCS ED visits that a group of enrollees will incur.

KEYWORDS: emergency department, claims analysis, risk drivers, utilization

Introduction

The Institute of Medicine (IOM) described emergency medicine as "at the breaking point" in 2007.[1] Emergency department (ED) utilization is growing as the number of EDs nationwide is shrinking. From 1999 to 2009, ED visits increased by 32%, while the number of EDs decreased by 2%.[2] Much of this ED care could be provided in far less resource-intensive settings.

Care in the ED is often associated with lack of coordination between providers, potentially resulting in unnecessary procedures and worse care.[16] With less than half of EDs completely transitioned to using electronic medical records, and with understaffing common, communication and follow-up are often challenging.[1,11]

A substantial proportion of ED visits are potentially avoidable and/or primary care sensitive (PCS), since many patients seek ED care for conditions that could have been treated in primary care. In surveys, as many as half of patients visiting the ED for nonurgent reasons say that being unable to get a timely appointment with their healthcare provider was a reason for their visit.[17-21] In one national study of 56 EDs, admitting triage nurses classified 37% of all ED visits as nonurgent.[7] Another study, using an algorithm developed by New York University to categorize PCS ED use, found that nearly 75% of patients who walked in to hospital EDs in New York City in 1998 were either non-emergencies or were treatable in a primary care setting.[8]

An ED utilization prediction model (especially one that predicts PCS use) could be useful for reducing unnecessary utilization and for profiling providers. ED utilization has been targeted for reduction under patient-centered medical home (PCMH) models and other practice and payment reform systems. However, there have been few previous attempts to create ED risk models.[3,44,96] ED risk models are important for accurately targeting high-risk enrollees with educational and care management programs, with the goal of preventing future ED visits. An ED risk model could also be used to set risk-adjusted "expected use" targets for panels of enrollees, against which actual use can be judged.

Care in the ED is more expensive than in other settings. Many studies have found that costs to Medicare, Medicaid, and other third-party payors, as well as individual out-of-pocket costs, are considerably higher (320%-728% in one study[22]) for the same services provided in the ED versus less acute settings.[22-25] Reducing ED overuse could save as much as $38 billion per year.[26]

There is no consensus on how to use administrative data to infer whether a particular visit could have been avoided.[54] However, the NYU ED Algorithm has been validated in its ability to distinguish cases with a higher risk of mortality or subsequent hospital admission from less acute cases;[68] it is the only validated tool for classifying ED visits using administrative data.[72]

This study's objective was to develop and evaluate claims-based models to predict ED use in a nationwide sample of the privately insured US population.

Methods

Conceptual Framework

The conceptual framework for this study was adapted from Andersen's behavioral model of health services utilization.[60] This model posits that utilization is influenced by need (e.g., number of medical conditions), predisposing (e.g., sociodemographic), and enabling (e.g., insurance and income) factors – both at the individual and societal level – as well as health behaviors. Further, primary care utilization has been shown to influence ED use. We used this conceptual framework to help guide our selection of covariates in our ED risk prediction models and to aid in interpreting our findings.

Data and Sample Selection

We used the Truven Health Analytics MarketScan Commercial Claims and Encounters database (Truven Health Analytics, Ann Arbor, MI), a proprietary database that we accessed through a partnership with Verisk Health, Inc. MarketScan includes data from approximately 45 large employers who self-insure employees and their dependents.[87] This nationwide claims database is widely used by researchers to examine health services utilization and costs, with over 550 peer-reviewed articles published since 1990.[88] The MarketScan data are validated to ensure that they are complete, accurate, reliable, and HIPAA compliant. Our study was approved by the University of Massachusetts Medical School institutional review board.

The MarketScan Commercial Claims and Encounters database includes enrollees of all ages in participating private health plans and large self-insured employer plans, including comprehensive, HMO, POS, and PPO plans. The database contains residents of the 50 US states, the District of Columbia, and the US Virgin Islands; in 2007, it contained 35 million unique enrollees, and in 2008, 49 million. To be included in our study, individuals were required to have drug benefits and to have at least 1 month of enrollment in 2007 and at least 1 month of enrollment in the subsequent 6-month period (January-June, 2008). We chose a prediction period of 6 months because a shorter period enables case managers to focus on individuals who are at high risk in the near future. Data from 2007 were used to predict the likelihood of ED use between January and June, 2008. Our final dataset included 15,136,261 unique individuals.

DxCG Clinical Classification Systems

The DxCG risk adjustment clinical classification system (DxCG Intelligence v.4.1, Verisk Analytics, Jersey City, NJ) was used to create individual-level summaries of illness burden and two sets of morbidity scores—prospective, using the base year's data to predict the next year's costs, and concurrent, predicting current-year costs—one set based on diagnoses, and one based on prescription drug fills. The DxCG illness burden summaries and morbidity scores are based on all diagnoses (excluding laboratory and radiology claims and other services without a face-to-face clinician encounter) and pharmacy codes recorded on claims.[97,116] ICD-9-CM diagnosis codes are regrouped into 1,010 homogeneous clinical groups, called DxGroups,[117] which are mapped into 394 condition categories (CCs) based on the clinical body system and relative resource use. Empirical studies have shown that using CCs to predict outcomes, such as next year's disease progression and costs, is robust to many variations in practices for coding diagnoses in administrative claims data.[97,116]

The National Drug Codes (NDC) for prescription medications were also classified into one of 251 mutually exclusive categories, called RxGroups, and used to create the second set of morbidity scores based on enrollees' prescription drug fills. This second morbidity score is useful because it captures additional morbidity not captured by the claims for face-to-face encounters. For example, many individuals who take a daily antidepressant medication might not have a clinic visit coded with a depression diagnosis during a given observation period. In addition, researchers have found that including both encounter-based and prescription-based morbidity scores improves a model's predictive ability.[118]

Predictor Variables and Outcome Measures

Predictor Variables

Our predictor variables included 22 age/sex groups (11 for males and 11 for females, as follows: 0 – 1 years, 2 – 5 years, 6 – 12 years, 13 – 17 years, 18 – 24 years, 25 – 34 years, 35 – 44 years, 45 – 54 years, 55 – 59 years, 60 – 64 years, and 65+ years), the prospective medical morbidity score, the prospective pharmacy morbidity score, the number of ED visits in each of the preceding 4 quarters, the number of chronic conditions, and 3 annualized expenditure variables (1 each for inpatient, outpatient, and pharmacy).

Outcome Measures

We examined 2 outcomes related to ED visits in the first 6 months of 2008, as predicted from 2007 data. The first outcome measure was any ED use during the prediction period. Only outpatient ED visits were considered (that is, visits by those who were not admitted to the hospital from the ED). This approach is often used when examining potentially avoidable ED use, under the assumption that ED visits on the path to a hospital admission are usually unavoidable.[52,53]

Our second outcome measure, total PCS ED visits, was defined by first applying the NYU ED algorithm to the principal diagnosis codes associated with each ED visit, then creating a summary measure across the 6-month prediction period (described in further detail below). The algorithm, developed by Billings et al.,[93] uses the principal diagnosis of an ED visit to assign probabilities that it belongs in one of 9 categories: 1) nonemergent; 2) emergent, primary-care treatable; 3) emergent, preventable/avoidable (consisting of the same diagnoses as for ACS hospitalizations); 4) emergent, not preventable/avoidable; 5) mental health; 6) substance use; 7) alcohol; 8) injury; 9) unclassifiable. The sum of each visit's probabilities is 1. Categories 1-4 may be 0, 1, or any value in between; categories 5-9 are either 0 or 1.

According to Billings et al., “nonemergent” refers to a condition for which treatment was not required within 12 hours.[119] The “primary-care treatable” category consists of conditions in the original chart review that did not require imaging or other resources not typically available in primary care settings. The “preventable/avoidable” category includes the same conditions as those that define ambulatory care sensitive (ACS) hospitalizations, another commonly used quality indicator.[120]

The probabilities assigned to categories 1-4 were estimated based on detailed chart review of approximately 6,000 ED visits at New York hospitals in the 1990s.[93] Consequently, most rare conditions are categorized as 9) unclassifiable, as are any ICD-

9-CM codes introduced since 2003, the last year that Billings et al. updated the algorithm (please see APPENDIX A for further details).

Prior users of the algorithm have measured PCS by first categorizing each visit as PCS or not, depending on whether the sum of the probabilities in a designated subset of the categories (e.g., categories 1-3 or categories 1 and 2) exceeds a threshold, such as 0.50.[68] Then, for each person, they either count them as having "any PCS ED visit" (vs. not), or count the "number of PCS ED visits." Using a threshold to dichotomize the outcome loses information, and requires specifying an arbitrary number as the threshold. For example, with a 0.50 threshold, 10 ED visits, each of which has probability summing to 0.60, counts for 10, while with a 0.75 threshold, they count for nothing. To avoid these problems, we developed an alternative approach.

Our PCS ED measure first uses the algorithm to assign probabilities to each ED visit and then sums the probabilities from categories 1-3 for all of a person's ED visits to calculate total PCS ED visits. For example, suppose a person had 1 ED visit for asthma, unspecified (ICD-9-CM 493.9) and 1 ED visit for headache (784.00). Asthma is assigned a prob_PCS of 1, while headache is 0.87, resulting in a total of 1.87 PCS ED visits for this enrollee.

Statistical Analyses

We first described the sample in terms of age, sex, medical and pharmacy morbidity scores, number of chronic conditions, number of ED visits in each quarter of the base year, and base-year inpatient, outpatient, and pharmacy expenditures. We calculated rates per 100 persons for each outcome in both the base and prediction periods (with the unit of analysis being observed utilization during the specified period per person). We then estimated separate regression models for each outcome measure.

Model 1 was a multivariable logistic regression model predicting the likelihood of any ED visit in the first 6 months of 2008, based on 2007 data. The model included all the predictor variables described in the previous section. Overall model performance was evaluated in several ways, including: R^2, sensitivity (true positive rate), specificity (1-false positive rate), and positive predictive value (PPV). **Table 3-1** provides the formulas used to calculate each metric.

Table 3 - 1. Formulas for model performance metrics

$$R^2 = 1 - \sum(\hat{y}_i - y_i)^2 / \sum(\bar{y}_i - y_i)^2$$

Sensitivity (True Positive Rate) = TP/(TP+FN)

Specificity (1-False Positive Rate) = TN/(FP+TN)

Precision (Positive Predictive Value, or PPV) = TP/(TP+FP)

TP = # of true positives; TN: # of true negatives; FP: # of false positives; FN: # of false negatives.

R^2 was calculated as the squared correlation between outcomes, or Ys, and predicted outcomes, or Y-hats. When modeling with OLS, this coincides with the R^2 value reported by standard statistical packages. Computing R^2 in this way allows for comparing how well different models' predictions fit the actual outcomes that they are intended to predict—something that is otherwise hard to do when the models have very different structural forms.[121]

Sensitivity is the proportion of actual positives that is correctly identified as having ED use. Specificity (1-false positive rate) is the proportion of actual negatives that is correctly predicted not to have ED use. PPV is the proportion of members identified in the model who actually used ED services in the first 6 months of 2008 (for example, the percentage of enrollees in the 99th percentile with any ED visit).

Model 2, predicting total PCS ED visits, included the same covariates as in Model 1. Because this outcome measure was neither an integer nor normally distributed, we explored several different estimation methods, including: a generalized linear model (GLM, using PROC GENMOD in SAS) with several different distribution and link specifications; an ordinary least-squares (OLS) model, with and without splining; and a two-part (logistic plus OLS) model to estimate PCS visits among those with an ED visit. Splining enables linear models to capture non-linear relationships through the use of multiple points of inflexion (knots).[122,123] Our splining method used three steps: 1) specify OLS regression models to predict ED use; 2) based on that model, create variables representing the splined inflexion points for each age/sex group's predicted risk; 3) re-specify the OLS regression model including the spline variables created in Step 2.

For our second outcome measure, total PCS ED visits, model performance was evaluated using R^2, sensitivity, specificity, and PPV, as defined above. However, rather than evaluating the model's ability to predict any ED use within each of the top groups, we evaluated the model's ability to predict any PCS ED use for this outcome.

Sensitivity analyses were conducted using different predictive scenarios, varying the target percentages defining ED risk in the top 0.5%, 1%, 2.5 %, 5%, 10%, etc. These analyses of the top risk groups allow us to assess the predictive ability of the models within quantiles of predicted likelihood of ED use. All statistical analyses were performed using SAS v. 9.3 (SAS Institute, Research Triangle Park, NC) and StataIC v. 11.3 (StataCorp LP, College Station, TX).

Results

Table 3-2 summarizes population demographics and ED utilization during the base year. The average age was 34.2 years, and 52% were female.

Table 3 - 2. Descriptive characteristics of the MarketScan sample, 2007 (base year)

	% of Sample	Percent with Any ED Use in Base Year
All (N= 15,136,261)	100%	12.9
Female	51.8%	13.5
Male	48.2%	12.3
Age (mean = 34.2)		
0-17	25.2%	13.9
18-44	38.4%	13.0
45-64	36.1%	12.1
65+	0.2%	14.6
Number of ED visits in base year		
0	87.1%	
1	9.5%	
2	2.3%	
3+	1.1%	
	Mean	**SD**
Prospective medical morbidity score	0.05	0.04
Prospective pharmacy morbidity score	0.05	0.03
Number of chronic conditions	0.92	1.40
Base year expenditures (2007 US$)		
Inpatient	$417	$2,230
Outpatient	$1,478	$2,837
Pharmacy	$482	$769

Source: Authors' analysis of MarketScan Commercial Claims and Encounters database, 2007

Outcome Measure 1: Any ED Use

In the 6-month prediction period, the rate of overall ED use was 10.6 per 100 persons (1,598,150 visits among 15,136,261 persons). The R^2, calculated as the squared correlation between outcomes (Ys) and predicted outcomes (Y-hats), was 3.83%. **Table 3-3** presents the metrics for Model 1, which predicted any ED use, within each risk group;

this table is intended to show how our model could be used to identify individuals at highest risk of ED use.

Among the 0.5% of the population with the highest predicted probability of ED use, 49.7% had at least one ED visit in the prediction period. The false positive rate and the true positive rate were 0.3% and 3.1%, respectively. That is, the model's specificity was 99.7%, and 3.1% of members who visited the ED were included in the top 0.5 percentage cohort. When the screening threshold was set at a higher level—for example, the top 10%—the resulting top group contains 20.3% of members who will have an ED visit in the following 6 months; sensitivity increased to 25.3%, while specificity remained high at 91.3%. As would be expected, the mean number of PCS ED visits was highest in those that the model predicted to be at highest risk of any ED visit, and decreased monotonically with decreasing predicted risk.

Table 3 - 3. Model 1 (any ED use) performance metrics, MarketScan data, 2008 prediction period

Predicted Risk of ED Visit	Sensitivity (True Positive Rate)	Specificity (1-False Positive Rate)	Positive Predictive Value for Any ED Use	Mean Number of PCS Visits
Top 0.5%	3.1%	99.7%	49.7%	0.73
Top 1%	5.2%	99.4%	41.6%	0.53
Top 2.5%	9.9%	98.2%	31.9%	0.34
Top 5%	15.9%	96.0%	25.6%	0.25
Top 10%	25.3%	91.3%	20.3%	0.18
100%	100.0%	0%	8.1%	0.05

Source: Authors' analysis of MarketScan 2007-08 Commercial Claims and Encounters database (n=15,136,261). PCS: primary-care sensitive.

Outcome Measure 2: PCS ED Use

The population-wide mean number of PCS visits was 5.4 visits per 100, with 795,610 (5.3%) of the total sample (that is, 49.8% of those with any ED users) having any PCS use. The highest R^2 was achieved using the OLS with splining method, as shown in **Table 3-4**. As discussed earlier, splining improves the model fit by changing the regression line for specific age-sex groups. The remainder of these results report characteristics of the final version of the model (OLS with splining).

Table 3 - 4. R^2 results for different estimation methods of Model 2 - PCS ED Use, MarketScan data, 2008

Model type	R^2
GLM (PROC GENMOD)	
Gamma distribution, log link	0.52%
Poisson distribution, log link	3.52%
Normal distribution, log link	6.54%
Two-part model (logistic + OLS)	7.08%
OLS	5.48%
OLS with splining	7.28%

R^2 was calculated as the squared correlation between outcomes (Ys) and predicted outcomes (Y-hats).
GLM: Generalized linear model; OLS: ordinary least-squares

Table 3-5 presents sensitivity, specificity, PPV, and the mean number of PCS ED visits within each top risk group for Model 2 (PCS ED use). The table demonstrates the model's ability to predict those with any PCS ED use in each of the risk groups. Model 2 showed slightly higher sensitivity and specificity in each risk group than did Model 1, predicting any ED visit. However, the PPVs are slightly lower. For example, in the top risk group (those in the 99.5th percentile of predicted risk), 40.5% had any PCS ED use, with the mean number of PCS ED visits in that group being similar.

Table 3 - 5. Model 2 (PCS ED use) performance metrics, MarketScan data, 2008 prediction period

Predicted Risk of ED Visit	Sensitivity (True Positive Rate)	Specificity (1-False Positive Rate)	Positive Predictive Value for Any PCS ED Use	Mean Number of PCS Visits
Top 0.5%	3.8%	99.7%	40.5%	0.74
Top 1%	6.2%	99.3%	32.4%	0.53
Top 2.5%	11.2%	98.0%	23.6%	0.34
Top 5%	17.3%	95.7%	18.2%	0.24
Top 10%	26.7%	90.9%	14.0%	0.17
100%	100.0%	0%	5.3%	0.05

Source: Authors' analysis of MarketScan 2007-08 Commercial Claims and Encounters database (n=15,136,261). PCS: primary-care sensitive.

Factors Related to Overall and PCS ED Use

The strongest predictors of ED utilization (overall and PCS) were the pharmacy risk score, medical risk score, number of ED visits in the base year, and age (**Table 3-6**). Infants (of either sex) had the highest risk. Higher inpatient, outpatient, and pharmacy spending in the base year were significantly associated with a *reduced* risk of having an ED visit, while larger numbers of chronic conditions was significantly associated with *increased* risk, but the magnitudes of these effects were small for both factors.

Table 3 - 6. Coefficients, standard errors, z-scores, and *P*-values from models of both outcome measures, MarketScan data, 2008

	Any ED Visit				Total PCS ED Visits			
	Coef.	**SE**	**z**	***P***	**Coef.**	**SE**	**z**	***P***
Female 0 - 1	0.06	0.00	32.26	**<.001**	0.09	0.00	28.04	**<.001**
Female 2 - 5	0.02	0.00	8.47	**<.001**	0.02	0.00	6.62	**<.001**
Female 6 - 12	0.00	0.00	0.07	.948	0.00	0.00	0.39	.696
Female 13 - 17	0.01	0.00	4.18	**<.001**	0.02	0.00	7.11	**<.001**
Female 18 - 24	0.03	0.00	15.89	**<.001**	0.04	0.00	14.67	**<.001**
Female 25 - 34	0.03	0.00	14.21	**<.001**	0.03	0.00	11.81	**<.001**
Female 35 - 44	0.01	0.00	6.77	**<.001**	0.01	0.00	4.29	**<.001**
Female 45 - 54	0.00	0.00	1.32	.188	0.00	0.00	-0.44	.660
Female 55 - 59	0.00	0.00	-1.08	.279	-0.01	0.00	-2.48	**.013**
Female 60 - 64	-0.01	0.00	-3.07	**.002**	-0.01	0.00	-4.32	**<.001**
Female 65+	0.00	0.00	0.61	.543	0.01	0.00	1.30	.192
Male 0 - 1	0.08	0.00	38.63	**<.001**	0.11	0.00	35.91	**<.001**
Male 2 - 5	0.02	0.00	8.86	**<.001**	0.03	0.00	10.93	**<.001**
Male 6 - 12	0.00	0.00	-2.06	**.040**	0.01	0.00	1.94	.053
Male 13 - 17	-0.01	0.00	-5.75	**<.001**	0.01	0.00	3.82	**<.001**
Male 18 - 24	0.00	0.00	0.36	.718	0.01	0.00	4.79	**<.001**
Male 25 - 34	0.01	0.00	3.64	**<.001**	0.01	0.00	4.36	**<.001**
Male 35 - 44	0.00	0.00	-0.03	.976	0.00	0.00	0.44	.657
Male 45 - 54	0.00	0.00	-2.66	**.008**	-0.01	0.00	-2.67	**.008**
Male 55 - 59	-0.01	0.00	-3.61	**<.001**	-0.01	0.00	-3.41	**.001**
Male 60 - 64	-0.01	0.00	-5.13	**<.001**	-0.01	0.00	-5.06	**<.001**
Male 65+	*Reference*				*Reference*			
Prospective medical morbidity score	0.22	0.00	90.14	**<.001**	0.23	0.00	101.45	**<.001**
Prospective pharmacy morbidity score	0.40	0.00	146.65	**<.001**	0.42	0.00	175.37	**<.001**
Number of ED visits, Q1 2007	0.08	0.00	272.12	**<.001**	0.13	0.00	281.87	**<.001**
Number of ED visits, Q2 2007	0.08	0.00	289.73	**<.001**	0.14	0.00	307.81	**<.001**
Number of ED visits, Q3 2007	0.09	0.00	338.46	**<.001**	0.16	0.00	363.30	**<.001**
Number of ED visits, Q4 2007	0.12	0.00	384.65	**<.001**	0.21	0.00	411.36	**<.001**
Number of chronic conditions	0.00	0.00	57.14	**<.001**	0.01	0.00	70.00	**<.001**
Expenditures, 2007 (each additional $10k)								
Inpatient	-0.01	0.00	-38.04	**<.001**	-0.02	0.00	-33.70	**<.001**
Outpatient	-0.02	0.00	-73.02	**<.001**	-0.03	0.00	-58.03	**<.001**
Pharmacy	-0.02	0.00	-20.04	**<.001**	-0.04	0.00	-20.47	**<.001**

Coef: coefficient; SE: standard error; z: z-score (standardized coefficient). **Bold** indicates statistically significant *P*-values.

Discussion

In this study, we developed predictive models of any ED visit and the estimated number of primary care sensitive (PCS) ED visits during a 6-month period, using administrative data from a large, nationwide sample of commercially insured individuals. We found that models predicting the number of PCS ED visits had higher R^2s than models predicting any ED visit, and that approximately 40-50% of those predicted, based on their prior-year characteristics, to be at highest risk of ED utilization actually did use the ED in the subsequent 6-month period.

Our method for measuring PCS visits is innovative. Prior users of the NYU algorithm have generally categorized visits as PCS or not, based on whether certain probabilities or the sum of probabilities met or exceeded a threshold, such as .5, .75, or even 1.[44,65,68] This method has several problems: it discards information, unnecessarily inflates the number of people with zero outcomes, and depends upon an arbitrary choice of threshold that strongly influences the outcome distribution. In this study, we show that applying dichotomies is not necessary. We expect that this approach will facilitate efforts by planners and researchers to design intervention models to prevent PCS ED use.

In choosing a model specification for the PCS outcome measure, those wishing to develop models will be constrained by the size of their sample population. Our sample of more than 15 million enrollees allowed us to develop a robust OLS model with splining, which generated the highest R^2. Those who build such models recommend using at least 100,000, and typically 500,000 or more, observations in development samples if splining is desired. However, among the models we considered, a hurdle model (logistic plus OLS) yielded an R^2 that was nearly as high.

Although their sensitivity is not high, our models can identify small cohorts of high-risk enrollees. For example, if PCPs for enrollees in this sample had been given a list of their enrollees who were in the top 5% of predicted risk for PCS ED use, about one fifth of them would—absent intervention—go on to have some PCS ED use in the next 6 months, as opposed to the one in twenty that would occur in a random selection. It should be noted that even for a perfect predictive model—one that accurately identified 100% of those in the top risk group who would have ED use—the maximum sensitivity we could achieve would be about 6%. To illustrate this point, imagine a cohort of 10,000 people with an 8% ED use rate (i.e., 800 ED users). A perfect model that predicted any ED use in a top risk group containing ½ of 1% of the population (50 people) would have a specificity and PPV of 100%, but its sensitivity would be 6.3% (50/800).

To our knowledge, only two prior peer-reviewed papers have described predictive models of ED utilization with information about model accuracy and precision. In a 2013 study, Billings and Raven developed models to predict frequent ED use in a cohort of 212,259 Medicaid-insured ED users in New York.[44] The strongest results reported were obtained when predicting patients with 3 or more visits during the index year, with a PPV of 66.3%, sensitivity of 22.9%, and specificity of 95.2%. Note that most (87%) of our sample consisted of non-users of the ED, whereas all of their population had used the ED at least once. Thus, it is not clear that there is any useful comparison to be made with our models' performance on very different outcomes.

In a 2012 study, Ash and Ellis developed models to predict 9 different outcome measures, including any ED use, in a cohort of 456,781 insured primary-care patients in upstate New York.[96] They report achieving an individual-level R^2 of 3% using a generic risk score and 25% using a tailored risk score. Our individual-level R^2 was 3.8%.

Small R^2s indicate that predicted outcomes typically differ from observed outcomes by almost as much as they would if predictions equaled the population mean. Predictive models of a random event such as emergency department use, especially in a larger sample, will typically have low R^2s. However, even predictive models with small R^2s can identify important systematic differences in a population and facilitate finding individuals with expected high costs who might be good candidates for case management. For example, the Centers for Medicare and Medicaid Services (CMS) currently adjusts payments for roughly 40 million Medicare Advantage (HMO) enrollees based on a risk score calculated using hierarchical condition categories (HCCs) derived from claims, achieving a model R^2 of 14.3%.[124]

Our study had several limitations, including our focus on a commercially insured population, which limits generalizability. In addition, our 6-month prediction period included the months of January to June; a different period, such as July to December, could have had different results because of seasonal trends in ED usage. The NYU ED algorithm has not been updated since 2003, which means that a relatively large fraction of ED visits could not be categorized (about 10% in our data). It will also need to be updated if it is to be used after the US healthcare system transitions to ICD-10-CM coding, expected to occur in October 2015. Future research is needed to update the algorithm and, perhaps, incorporate additional data, such as Current Procedural

Terminology (CPT) codes representing intensity of care or whether patients received imaging or surgical procedures. In addition, we chose to focus only on outpatient ED visits in this analysis (i.e., visits by those who were not admitted to the hospital after their ED visit), a commonly used strategy, since ED visits on the path to an inpatient stay are generally thought to be unavoidable.[52,53]

Our inclusion criteria required only 1 month of eligibility in the base period and 1 month in the prediction period, and our analysis treats people with partial-year observations the same as those with full observations during the prediction period. This approach has several advantages: it includes those who were born or died during either period, it uses all people in the data with enough information to contribute to our understanding of how year-1 variables relate to year-2 ED utilization, and it is consistent with an implementation-oriented approach using real-world data. For example, if we had required individuals to be eligible for the entire base and prediction periods, we would not have learned important information about the heightened risks associated with infants under 1 year old. In prior research, others have found that risk adjustment models have improved predictive ability—meaning, they are better at adjusting for risk and accurately predicting future outcomes—when enrollees with partial data are included, compared to requiring full-year eligibility.[89] Moreover, expected users of these models and methods will confront the same type of partially observed data as we used in this study. We demonstrated only modestly predictive results in these data, which likely has more to do with the importance of non-medical factors as drivers of ED use than from the presence of people with partial-year eligibility.

Our approach has several implications for practical implementation of these methods. For retrospective analyses (to set benchmarks for providers based on two years of recent data for the purposes of performance measurement and quality improvement efforts), treating people with partial-year observations the same as those with full-year observations could bias the model's benchmarks for providers with more individuals with partial-month data.

Predictive models meant to be used for case management purposes would be built in a very similar manner to the models we used in this analysis, since analysts in concurrent predictive scenarios do not know, in advance, for how many months in the future an enrollee will remain enrolled. However, when using predictive modeling to support case management efforts, having missing months of eligibility in the base year presents a problem when calculating risk scores. If we are missing months of observation, we may be biasing the risk scores downward (underestimating morbidity, and thus risk of future expenditures and/or utilization). We are not aware of any published research that provides a solution for appropriately dealing with partial-year eligibility in the base year.

We recognize the potential for problems resulting from discouraging ED use through higher copayments for enrollees; such changes could lead patients to avoid needed care for true emergencies. Also, if providers are incentivized to prevent (outpatient) PCS ED visits, this could potentially lead to an increase in inpatient admissions, since ED visits that result in a hospitalization are usually excluded when analyzing PCS ED utilization. Implementation of this type of policy would be best done in concert with other quality measures, such as those for ambulatory-care sensitive hospitalizations, to limit the risk of such an outcome.

Prediction models using administrative data can identify enrollees at high risk of having any ED usage and estimate the number of PCS ED visits that a group of enrollees will incur. Our approach to quantifying PCS ED use continuously—rather than applying arbitrary dichotomies—may represent a better way to measure ED use.

CHAPTER IV.
USING ENHANCED ADMINISTRATIVE DATA TO PREDICT EMERGENCY DEPARTMENT UTILIZATION: THE ROLE OF NEIGHBORHOOD POVERTY

ABSTRACT

BACKGROUND: In a pay-for-performance (P4P) environment, primary care providers (PCPs) may be held accountable for their patients' use of the emergency department (ED). Whether because of reduced access to primary care or other complex social, behavioral health, or physical health reasons, lower-income individuals are at higher risk of ED utilization.

OBJECTIVE: To explore the strength of the association between ED use and neighborhood poverty after adjusting for morbidity and other factors influencing ED use from Andersen's behavioral model of health services utilization.

METHODS: This retrospective, observational study included 64,623 unique enrollees (107,449 observations) in a Massachusetts managed-care network in 2009-11, with data on age, sex, race, morbidity (including claims-based morbidity scores and 10 conditions identified using electronic medical record [EMR] problem lists), prior use of the ED, payor (4 commercial insurers), and PCP type and quality. Census tract data on socioeconomic characteristics were linked to each enrollee. Multivariable regression models predicted 3 year-2 outcomes from year-1 data: 1) any ED visit, 2) total ED visits, and 3) total primary-care sensitive (PCS) ED visits, as defined using the NYU ED algorithm, a validated tool for categorizing ED visits using diagnosis codes.

RESULTS: About 14.6% (±0.1%) of the sample had 1 or more ED visits during the prediction period, with an overall mean ED visit rate of 18.8 (±0.2) visits per 100 persons and 7.6 (±0.1) PCS ED visits per 100 persons. Models predicting ED utilization using age, sex, race, morbidity, and prior use only (claims-based models) had lower R^2 (ranging from 2.9% to 3.7%) and poorer predictive ability than the enhanced models that also included payor, PCP type and quality, problem list conditions, and covariates from the EMR, Census tract, and MCN provider data (enhanced model R^2s ranged from 4.2% to 5.1%). In adjusted analyses, age, claims-based morbidity score, any ED visit in the base year, asthma, congestive heart failure, depression, and tobacco use, and neighborhood poverty were strongly associated with increased risk for all 3 measures (all $P<.001$).

CONCLUSIONS. Models predicting ED utilization should incorporate publicly available neighborhood-level variables, such as income, along with common risk adjustors such as age, sex, and morbidity. Otherwise, targets for ED use—even if adjusted for traditional “case mix” variables—may be unfair. ED utilization is driven by medical need, but other factors not under PCP control are also important.

KEYWORDS: emergency department, claims analysis, electronic medical records, small-area analysis, predictive modeling, utilization

Introduction

Use of the emergency department (ED) in the US is growing[11] and expensive.[2] According to many different estimates, about half of all outpatient (nonadmitted) ED visits are potentially avoidable;[48,50,93,125,126] in a perfect world, those visits would not have occurred. Potentially avoidable ED visits include visits for low-acuity and nonemergent conditions, such as a hangnail; conditions that could be treated in a primary-care setting, such as a urinary-tract infection; and conditions that could potentially be prevented or avoided with high-quality primary care, such as an asthma exacerbation. These types of ED visits are sometimes referred to as primary-care sensitive (PCS),[48,49,127] a term that highlights the association with primary care without saying that every instance of such utilization is "inappropriate".

Whether because of reduced access to primary care or other complex social, behavioral health, or physical health reasons, lower-income individuals are at higher risk of ED utilization. The association between neighborhood poverty and increased ED use has been well known since at least the 1980s.[128] The association persists after adjustment for numerous other risk factors.[52,53,128] However, the question of whether to risk-adjust quality measures for socioeconomic status is controversial, as demonstrated by the recent lively debate on a draft report by the National Quality Forum advocating risk-adjustment.[129]

According to Andersen's behavioral model of healthcare utilization, factors influencing ED utilization can be grouped into the following categories: contextual (or social/environmental) and individual *need*, contextual and individual *predisposing*, contextual and individual *enabling*, and *health behavior* factors.[60-62] Seen through this lens, administrative data, such as diagnosis codes from encounters and beneficiary characteristics like age and sex, may have limited power to predict ED use because they capture partial data in only 2 categories: need and predisposing factors. In this study, we aimed to develop more comprehensive predictive models of ED utilization—both overall and primary-care sensitive—by incorporating data from multiple sources, including expanded clinical data from the electronic medical record (EMR), data about the healthcare system (from administrative files), and information about the neighborhoods in which enrollees lived (from the US Census).

Methods

Data Sources

Our data on individuals and their providers was provided by a managed-care network (MCN) in Massachusetts, which requested help in developing a performance measure for PCPs in the network based on practice-level ED use. The MCN includes several hospitals (including one affiliated with a medical school) and over a thousand providers. In addition, we incorporated data from the 2011 American Community Survey, an annual survey conducted by the US Census Bureau.[130] The study was approved by the University of Massachusetts Medical School institutional review board.

Study Sample

The study population included residents of Massachusetts: 1) who were enrolled for at least 1 month in each of 2 consecutive years in 1 of 4 large commercial insurance plans (referred to as Plans 1-4); and 2) whose primary care provider (PCP) was affiliated with the MCN and had transitioned from paper medical records to the Allscripts electronic medical record (EMR) system prior to 2009.

Complicated models with many risk factors that have been (over)fit to one data set may not predict outcomes as well when applied to new data.[131] Thus, we split our population into two samples: the development sample included individuals observed in 2010 (base year) and 2011 (prediction year), while the validation sample included individuals observed in 2009 (base year) and 2010 (prediction year).

Measures

Using data from the base year, we measured and predicted 3 outcomes during the prediction year: 1) any outpatient ED visit; 2) total number of outpatient ED visits during the prediction period; and 3) total number of PCS ED visits. PCS visits included outpatient nonemergent, primary-care treatable, and preventable/avoidable visits, as categorized using the NYU ED Algorithm.[93]

Although diagnosis codes change from year to year, the NYU ED algorithm has not been updated by its original developers since 2003. To reduce the number of diagnosis codes that could not be classified, we used a version developed in 2009 by the Massachusetts Center for Health Information and Analysis (CHIA) (personal communication, April 30, 2013). CHIA's version of the algorithm built on the original to incorporate new codes with input from the original developer and an emergency medicine physician, but did not involve new data abstractions.[50] When we applied the CHIA-updated algorithm to our 2010 sample, it reduced the percentage of unclassifiable claims from 14.8% to 10.1% (please see APPENDIX A for further details). The updated algorithm is available from the authors.

We excluded visits to the ED that resulted in an inpatient admission (approximately 15%), a standard approach when examining ED visit data, since visits that terminate with an inpatient stay are considered unlikely to be PCS.[52,53] Visits related to injuries, alcohol/drug use, and mental health complaints, and the approximately 10% of visits that the algorithm could not classify were not counted as PCS.

Practice and Provider Characteristics

Data on practice and provider characteristics were obtained from MCN administrative records. Practice characteristics included location, practice specialty, and a provider quality score. The specialties were family practice, internal medicine, maternal/pediatrics, and multi-specialty/other. The provider quality scores were developed by the MCN as part of their internal rewards program, and were based on 21 quality and efficiency measures (primarily HEDIS measures, such as well-child visits) described in detail inAPPENDIX B.[86]

Problem List Variables

Data from the Allscripts EMR included, for each individual in the prediction year of either sample, problem list entries consisting of diagnosis codes and verbal descriptions (802,420 entries after removing "normal/routine" entries). We used these problem lists to identify 10 conditions (arthritis, asthma, cancer, congestive heart failure [CHF], chronic obstructive pulmonary disease [COPD], depression, diabetes, hypertension, overweight, and tobacco use). The conditions were selected from the list of "priority conditions" developed by the Agency for Healthcare Research and Quality.[132] The algorithms were developed by a PhD candidate with input from an MD, an MD/PhD student, and a PhD researcher. The problem lists obtained from Allscripts included both a diagnosis field (containing an ICD-9-CM code) and a description field. Since many records were missing either the diagnosis code or description, we used a two-stage algorithm to identify cases from either field.

We identified the initial diagnosis codes to be matched using the ICD-9-CM code manual at http://www.icd9data.com/. We developed the description search terms using a software programming technique known as "regular expression matching".[133] Regular expressions are strings of letters and special characters known as operators, which can be used to match substrings and portions of text in a text-based variable (in this case, the description field in a problem list). For example, to identify individuals with arthritis, we first flagged every record that contained any ICD-9-CM code within the range 714.00 to 716.99. We then searched the description fields for the word "Arthritis", which could be either upper- or lower-case. We then flagged and removed any records that contained the terms allergic, bacterial, bowel, infect* (using a wildcard character, *, to match any word that started with "infect"; such as infectious), reactive, and septic, to exclude acute forms of arthritis. We then scanned the remaining records to ensure that the preliminary set of arthritis cases was accurate and that no other description field terms should be included or excluded.

In addition, we generated a list of enrollees with a claim for one of the 10 conditions who had not been flagged by the problem list algorithm for that condition. We randomly selected 25 cases per condition for detailed review. For each set of cases, two researchers independently reviewed all problem list entries for each enrollee to determine whether any diagnosis codes or description field terms should be added to the algorithm. Further details, including a description of our analysis of the concordance between the problem lists and claims, may be found in APPENDIX D.

Morbidity Measurement

DxCG Intelligence version 4.1 (Verisk Analytics, Jersey City, NJ) was used to classify diagnoses from the base-year utilization files into hierarchical condition categories and to generate 2 morbidity scores: a concurrent score and a prospective score. The difference between these two scores is that the concurrent score uses one year's data to predict illness burden in that same year, while the prospective score uses one year's data to predict illness burden in the next year. Morbidity scores were normalized to the study sample and top-coded at the 99.5th percentile (normalized prospective risk score: mean=1, SD=1.34; normalized concurrent risk score: mean=1, SD=1.93). See APPENDIX C for further details on our top-coding process and resulting values.

Small-area Analysis

Area-based measures, such as median income and percent of residents living in poverty, are publicly available and have been used in many analyses to provide insight into socioeconomic status.[134-137] These measures are not only a reasonable proxy of individual situations, but also measure contextual predisposing and enabling factors in the residential environment that are relevant in Andersen's model.

To better understand the effects of geographic residence on ED utilization, we conducted small-area analyses using enrollee addresses and data from the 2011 American Community Survey, an annual survey conducted by the US Census Bureau. We used addresses in our data to assign each enrollee to a Census tract. A census tract (CT) is a small area, designed to be relatively homogenous with respect to population characteristics, economic status, and living conditions (compared to ZIP codes). This makes CTs a better geographic area for analysis than ZIP codes.[134,136] In the Massachusetts data from 2011, there were 1,478 CTs containing an average of 4,406 persons per CT (in contrast, there are approximately 700 ZIP codes in Massachusetts). We used 5-year estimates from the 2011 survey, which are the most reliable and precise estimates of neighborhood-level characteristics available.[138]

The variables chosen for analysis were those believed, *a priori*, to potentially delineate differences between neighborhoods that could affect ED utilization, based on our theoretical framework and review of the literature. The factors we analyzed included: median age, percent of high school graduates, mean travel time to work, median household income, percent owner-occupied housing, median house value, percent black, percent Hispanic, percent unemployed, percent foreign-born, percent who speak English less than very well, percent of housing units that are vacant, and percent in poverty. Some of these risk factors for ED use are well documented in the literature, such as those related to education and poverty.[3,5,6,63,67,128,137,139-144] Others, such as mean travel time to work, were included because they were thought to be potentially associated with ED use despite a lack of prior research substantiating the association. For example, longer travel time to work could indicate communities that are more rural or remote, and more time spent commuting is a risk factor for stress and fatigue, which could be associated with increased ED use. APPENDIX E provides further details on these variables, their values in our dataset and in the US as a whole, and the Census files from which the data were taken, as well as the results of an exploratory factor analysis. After obtaining the Census data and merging it with the enrollee data, we used multivariable regression modeling on the merged data to explore relationships between person-level and CT-based variables and ED visits.

We investigated the role of distance from enrollees' homes to the nearest ED and to their PCP as possible explanatory variables. We operationalized "distance" based on the addresses for enrollee's primary care practices from the MCN and the addresses of all hospitals with EDs in Massachusetts, obtained from the Massachusetts College of Emergency Physicians (personal communication, November 1, 2013). We then geocoded all addresses and computed driving distances from each enrollee's home to the nearest ED and their PCP (and the differences between those distances) using ArcGIS version 10.2 (Esri, Inc., Redlands, CA).

Poverty Analysis

We conducted initial analyses of the effect of neighborhood income on ED utilization using the CT-level median household income data. We used these data and the 2009-2010 Federal poverty threshold for a family of 4 (FPT: $22,050) to create 3 commonly used categories of neighborhood income: < 200% FPT, 200-399% FPT, and 400% FPT or more.[143] We also investigated the effect of different categorizations (<200%, 200-399%, 400-599%, and 600%+ and <200; 200-350, 350-500, 500+) and—

since results were essentially the same—chose to use the 3 standard categories just described to facilitate comparisons of our results with those of others.

We also created a binary indicator of living in an impoverished area (that is, an area with more than 20% of the population living below the FPT) and a continuous measure of the percentage of individuals living under the FPT. The final models included only the continuous measure, both because it is conceptually more appealing (because it discards less information), and empirically more predictive (models using it had higher R^2 values).

Statistical Analyses

We described the population in terms of sociodemographic and clinical characteristics, and examined bivariate associations between each predictor and likelihood of ED use. We tested for collinearity between predictor variables, and calculated the prevalence of ED use in the population.

We then used multivariable logistic regression models to predict the likelihood of any ED visit during the prediction year. We used zero-inflated negative binomial (ZINB) regression models to predict the number of ED visits during the prediction year. To predict the number of PCS ED visits, we explored several estimation strategies: two-part (hurdle) models, using a logistic regression to predict any ED visit in the first part and either an ordinary least-squares (OLS) model or a generalized linear model (GLM) in the second part to predict the number of PCS ED visits among those with any ED use; and GLMs with a log link and either a Gaussian or gamma family (distribution). For each model, we calculated the squared correlations between actual and predicted outcomes (R^2), using this metric to choose our final model.

The base model was built using factors from administrative data (age, sex, race, morbidity scores, and prior ED use (all from administrative data) as covariates. Three enhanced models adjusted for additional covariates drawn from payor and practice characteristics (from the MCN's administrative records), neighborhood characteristics (from the Census), and the EMR data (from Allscripts). We evaluated improvements in model fit and performance in base models versus enhanced models by comparing R^2s and inspecting graphs of predicted versus actual results within quantiles of predicted risk in both the development and validation samples.

Several model development strategies were used, including forced entry of all potential risk factors, as well as forward and backward stepwise selection with Bonferroni corrections to account for the number of simultaneous significance tests.[145] All risk factors statistically significant in one or more of those approaches were retained in the final models.

The validation procedure consisted of the following steps: 1) creating 3 Ys (outcome variables), one for each outcome, using the development sample; specifying each predictive model for each Y; estimating model-specific predicted outcomes (Y-hats) using both the development and the validation sample; and comparing Y-hats within quantiles of predicted risk to determine how well the models fit the validation sample.

To examine practice-level variations in outcomes, we excluded practices with fewer than 100 enrollees in their panels (which reduced the number of practices from 54 to 45). We examined differences among practices in their panel's average observed outcomes, in their expected outcomes based on multivariable regression models, and in their observed-to-expected (O/E) ratios. We created a pooled estimate of the standard deviation (SD) of an outcome from its prediction by first taking the square root of the sum of the variances for each practice using weights based on practice size (number of enrollees in the development sample). We used this to generate standard errors for each practice's expected outcomes. Analyses were conducted in Stata/IC version 11.2 (Stata Corp., College Station, TX) and SAS version 9.2 (SAS Institute, Research Triangle Park, NC).

Results

Baseline Characteristics

Enrollees

The development sample, with data from 2010-11, had 53,112 observations; the validation sample (2009-10) had 54,337 observations; combined, the dataset included 107,449 observations on 64,623 unique individuals. **Table 4-1** provides sociodemographic and clinical characteristics by sample. During the prediction periods, the two samples had similar rates (± SEs) of any ED use (development: 0.147 ± 0.001; validation: 0.142 ± 0.001), mean number of ED visits (development: 0.189 ± 0.002; validation: 0.181 ± 0.002), and PCS ED visits (development: 0.076 ± 0.001; validation: 0.073 ± 0.001).

Table 4 - 1. Demographic and clinical characteristics in the base year, by sample, MCN data, 2009-10

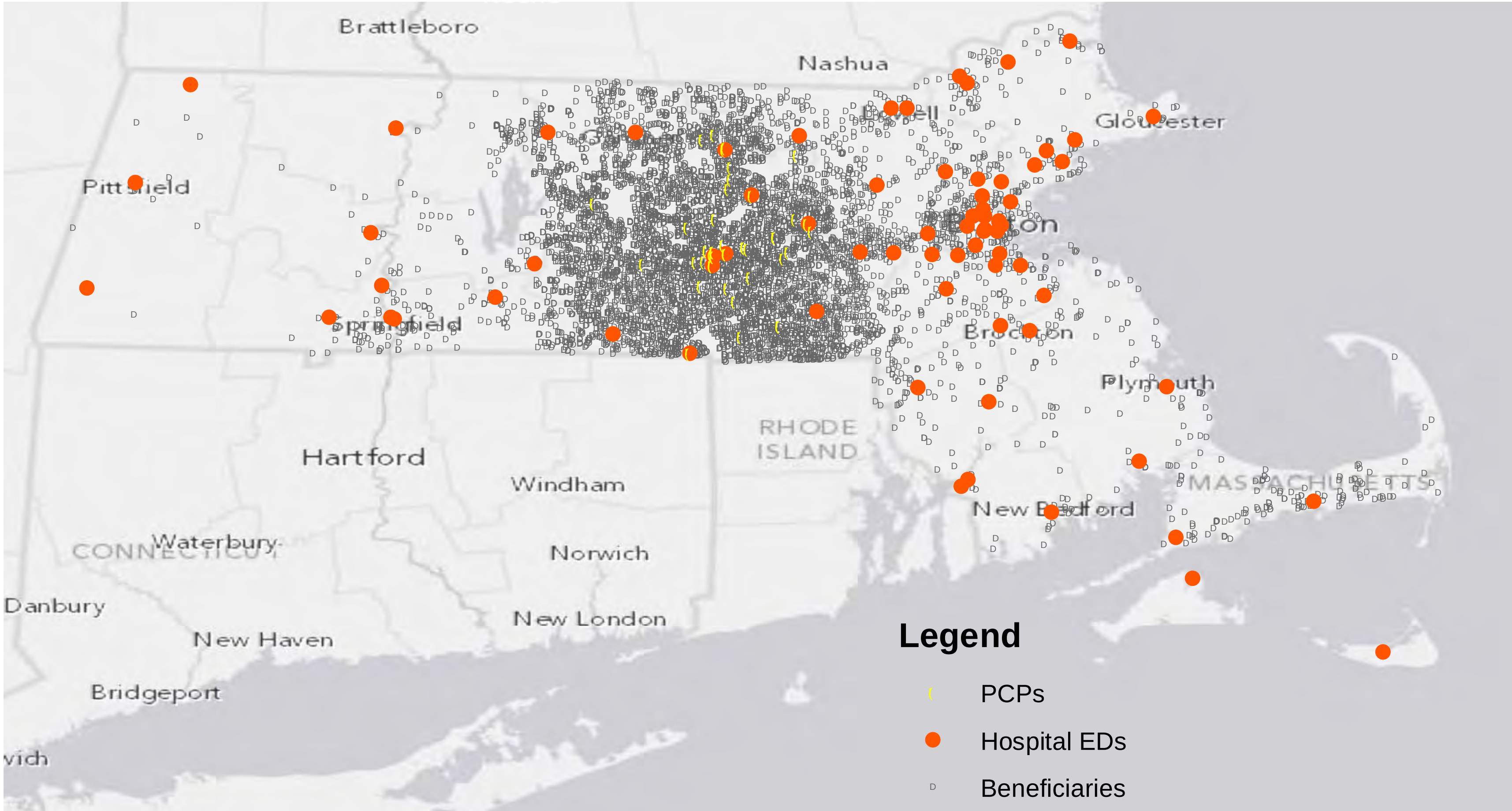
	Development (2010)		Validation (2009)		*P*-value
	N enrollees	%	N enrollees	%	
N	53,112		54,337		
Age (as of Jan. 1 of base year)					<.001
< 1	920	1.7%	749	1.4%	
1-10	5,889	11.1%	5,178	9.5%	
11-17	4,555	8.6%	4,187	7.7%	
18-24	5,075	9.6%	4,978	9.2%	
25-39	11,516	21.7%	10,641	19.6%	
40-64	25,237	47.5%	24,770	45.6%	
65+	1,145	2.2%	2,609	4.8%	
Female	27,983	52.7%	27,199	50.1%	.344
Race/ethnicity					.006
White	38,075	71.7%	36,947	68.0%	
Black	1,362	2.6%	1,323	2.4%	
Other	2,771	5.2%	2,563	4.7%	
Unknown	12,129	22.8%	12,279	22.6%	
Neighborhood income category					.089
Low (<200% FPT)	3,528	6.7%	3,585	6.6%	
Middle (200-399% FPT)	31,947	60.2%	32,365	59.6%	
High (400+ FPT)	17,637	33.2%	18,387	33.8%	
PCP type					<.001
Internal medicine	25,700	48.4%	25,657	47.2%	
Family medicine	14,278	26.9%	15,229	28.0%	
Maternal/pediatrics	13,056	24.6%	13,340	24.6%	
Other	78	0.2%	111	0.2	
Insurance plan					<.001
Plan 1	29,935	55.1%	26,791	50.4%	
Plan 2	8,578	15.8%	9,585	18.0%	
Plan 3	10,265	18.9%	10,819	20.4%	
Plan 4	5,559	10.2%	5,917	11.1%	
Selected conditions from problem lists					
Arthritis	3,824	7.2%	3,628	6.7%	.001
Asthma	6,481	12.2%	6,566	12.1%	.552
Cancer	11,225	21.1%	10,943	20.1%	<.001
Chronic obstructive pulmonary disease	215	0.4%	180	0.3%	.046
Congestive heart failure	253	0.5%	142	0.3%	<.001
Depression	7,431	13.4%	7,628	13.4%	.905
Diabetes	8,541	16.1%	8,322	15.3%	.001
Hypertension	12,655	23.8%	11,960	22.0%	<.001
Overweight	5,812	10.9%	5,862	10.8%	.415
Tobacco use	8,827	16.6%	9,021	16.6%	.938
	Mean	**SD**	**Mean**	**SD**	
Concurrent morbidity score (top-coded)	1.52	2.97	1.36	2.61	<.001
Prospective morbidity score (top-coded)	1.40	1.87	1.28	1.72	<.001
Provider quality score	1.08	0.85	1.10	0.87	.006

FPT: Federal poverty threshold; PCP: primary care provider; MCN: managed care network

Small-area Analysis

About 92% of enrollees lived in Worcester county, another 5% in Middlesex county, 1% in Norfolk county; the remaining 2% were spread over 9 other Massachusetts counties. As shown on the below map (**Figure 4-1**), the PCPs associated with enrollees in this sample were all clustered in central Massachusetts, whereas both enrollees and hospital EDs were spread across the state. Because of this, we found that the median (interquartile range [IQR]) driving distance to an enrollee's PCP was 5.8 (3.2-10.5) miles, and the median (IQR) driving distance to the nearest ED was 5.1 ± 2.8 miles. That is, on average, individuals had to drive 0.7 miles further to reach their PCP than to reach the nearest ED, which we refer to as the "extra distance to their PCP". Again, these differences were driven by the fact that some individuals lived quite far from central Massachusetts, where all the study's PCPs were located. Only 10% of enrollees lived more than 10 miles from a hospital with an ED, but 25% lived more than 10 miles from their PCP.

Figure 4 - 1. Map of primary care practices, hospital emergency departments, and enrollees

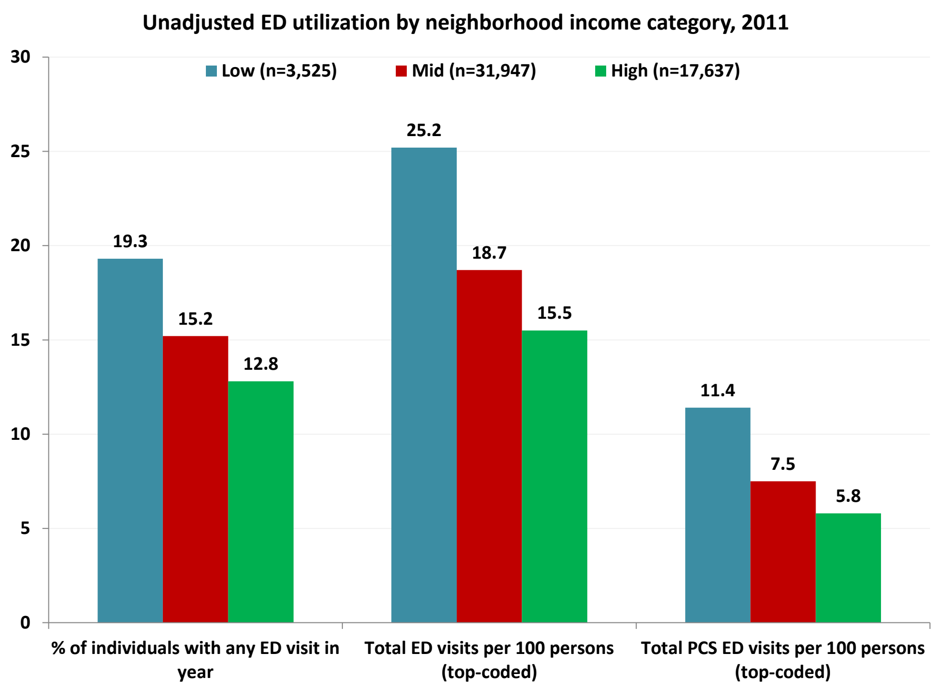


This map illustrates that both hospitals (red circles) and enrollees (gray hollow circles) were located across the state, but the PCPs (yellow bars) were clustered in central Massachusetts. This helps explains the fact that enrollees, on average, were located further from their PCP offices than from the nearest hospital ED.

Unadjusted Effects of Poverty and Morbidity on ED Utilization

Figure 4-2 shows the unadjusted relationship between neighborhood income category and ED utilization. Individuals living in neighborhoods with median family incomes of less than 200% of the FPT were most likely to have any ED visit (19% vs. 15% for middle income [*P*=.041] and 13% for high income [*P*=.017]) and had the highest number of total ED visits and total PCS ED visits (all *P*<.05). Those in middle-income neighborhoods (200-399% of FPT) also had higher ED utilization than those in the highest-income neighborhoods (400% or more of the FPT).

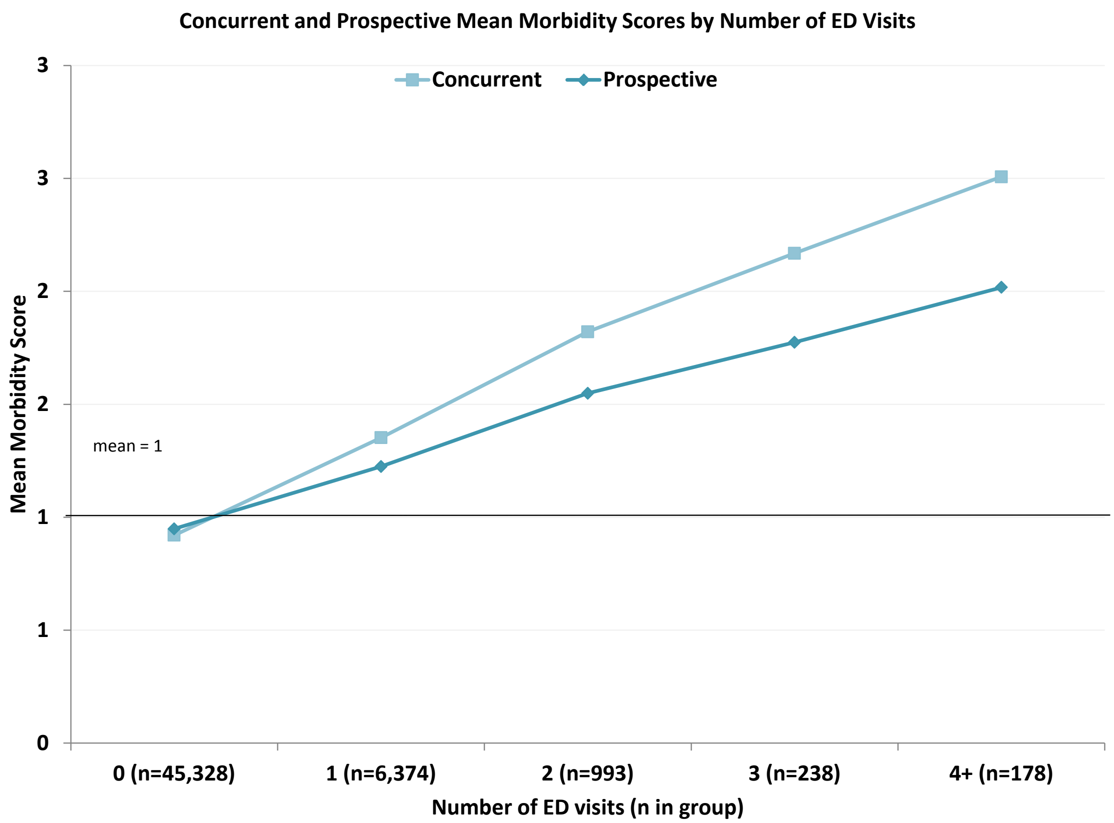
Figure 4 - 2. Emergency department use varies by neighborhood income category, MCN development data



Source: MCN development data (n=53,112); MCN: managed care network. Low: median household income in Census tract was <200% of Federal poverty threshold (FPT); mid: 200-399% of FPT; high: >400% of FPT.

In **Figure 4-3**, we illustrate the unadjusted relationship between the concurrent and prospective morbidity scores (each calculated using the base-year claims) and the number of ED visits in the following year. The purpose of this figure is to show the unadjusted relationship between the concurrent and prospective morbidity scores and the number of ED visits, demonstrating the importance of morbidity in predicting ED volume.

Figure 4 - 3. Mean morbidity score using base-year claims by number of ED visits in the subsequent year



Source: MCN development data (n=53,112); ED: emergency department; MCN: managed care network

Prediction Models and Predictors

Models predicting ED utilization using age, sex, race, morbidity score, and prior use (claims-based models) had lower R^2 and poorer predictive ability than models that also included payor, PCP type and quality, problem list conditions, and neighborhood poverty (enhanced models). **Table 4-2** shows the improvements in R^2 for each outcome measure when additional variables were included in the regression models within the following sets: Model 1 – administrative data only (age, sex, race, claims-based morbidity score, and any base-year utilization (office, inpatient, and ED); Model 2 – added payor (1 of 4 plans), practice specialty, and provider quality score; Model 3 – added extra distance to PCP and percent living below poverty in Census tract; and Model 4 – added indicators for each of the 10 EMR-derived conditions. The biggest improvements in R^2 were seen when adding variables derived from the enrollees' census tract—moving from Model 2 to Model 3—which added approximately 1 percentage point to the R^2 for each measure.

Table 4 - 2. Improvements in R^2 values for each outcome measure when additional sets of predictors are included

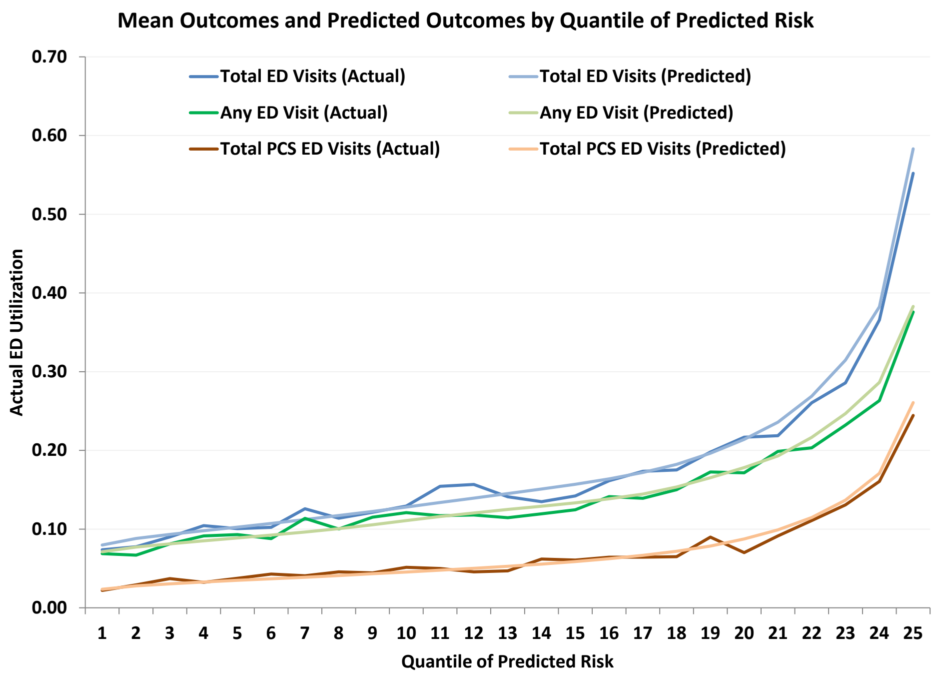
	Outcome 1--Any ED Visit	Outcome 2--Total ED Visits	Outcome 3--Total PCS ED Visits
Model 1 – Administrative data only	3.3%	3.7%	2.9%
Model 2 – Add provider/payor data	3.4%	3.9%	3.0%
Model 3 – Add neighborhood data	4.2%	5.0%	4.0%
Model 4 – Add EMR data (most predictive enhanced models)	4.3%	5.1%	4.2%

ED: Emergency department; EMR: electronic medical record; PCS: primary-care sensitive. Model covariates are listed in Table 4-3 (age, sex, race, morbidity, PCP/ED/inpatient use in base year, payor, PCP type, PCP quality score, extra distance to PCP, neighborhood % in poverty, and 10 conditions from problem lists).

For outcome 3—total PCS ED visits, the highest R^2 was obtained with a GLM model with a log link and Gaussian distribution (4.17%); the alternative specifications of the most predictive models (Model 4) produced R^2s of 3.58% to 4.10% (detail not shown). From this point forward, results for outcome 3 presented in this paper pertain to the enhanced GLM model only.

Outcomes in the validation data were consistent with predictions based on the models built with the development data. This is illustrated in **Figure 4-4**, which plots the actual ED utilization in the validation sample and the predicted utilization by quantile of predicted risk for each of the 3 outcome measures.

Figure 4 - 4. Model predictions vs. actual utilization in the validation sample by predicted quantile of risk



Source: MCN development and validation samples (n_{obs}=107,449 in 64,623 unique individuals); MCN: managed care network. The darker blue, green, and red lines correspond to actual ED utilization within each quantile of predicted risk in the validation sample; the lighter blue, green, and orange lines represent predicted ED utilization in the validation sample, based on the characteristics of the development sample. Models adjusted for all covariates listed in Table 4-3 (age, sex, race, morbidity, PCP/ED/inpatient use in base year, payor, PCP type, PCP quality score, extra distance to PCP, neighborhood % in poverty, and 10 conditions from problem lists).

Table 4-3 provides the coefficients from the most predictive enhanced models for each outcome measure. After adjusting for other factors, infants were at highest risk of having any ED visit; each age group was at significantly elevated risk of ED utilization relative to the reference group (age 40-64). Being female was associated with decreases in outcome 1—any ED use (*P*=.014) and outcome 2—total ED visits (*P*=.074), but with increases in outcome 3—PCS ED visits (*P*<.001). Black race was significantly associated with increased risk on all 3 measures, as were the prospective morbidity score, having any ED visit in the base year, and having any PCP visit in the base year. Having any inpatient stay in the base year was significantly associated with outcome 2—total ED visits. One of the 4 payors, the second-largest (Plan 3), was significantly associated with fewer ED and PCS ED visits, relative to the smallest payor (Plan 4). In our final enhanced model, the top 10 predictors of outcome 3—PCS ED use, as ranked by the standardized coefficients (z-scores), were prior ED use, asthma, age 18-24, morbidity score, depression, neighborhood poverty, age 1-10, age <1, tobacco use, and being female (all *P*<.001).

Of the 10 selected conditions noted in EMR problem lists, we found that 9 were significantly associated with ED risk on at least 1 outcome measure (**Table 4-3**). Asthma, congestive heart failure, depression, and tobacco use were associated with increased risk on all 3 outcome measures. Arthritis and hypertension were associated with higher risk on outcome 1— any ED visit and an increase in outcome 2—total ED visits, but not with outcome 3—total PCS ED visits. Cancer was significantly associated with fewer ED and PCS ED visits. Individuals with overweight in their problem list had a significantly higher number of ED visits. A higher likelihood of any ED visit was found for persons with diabetes. COPD was not significantly associated with increased risk of ED utilization (although it was very rare in this population). **Table 4-1** provides the number and percent of individuals with each of these conditions in each sample.

Table 4 - 3. Coefficients, standard errors, z-scores, and *P*-values from models of all 3 outcome measures, MCN data, 2010-11

	Outcome 1—Any ED Visit				Outcome 2—Total ED Visits				Outcome 3—Total PCS ED Visits			
	Coef.	**SE**	**z**	***P***	**Coef.**	**SE**	**z**	***P***	**Coef.**	**SE**	**z**	***P***
Age <1	0.98	0.09	10.37	**<.001**	0.87	0.08	10.36	**<.001**	0.93	0.16	5.89	**<.001**
Age 1-10	0.58	0.05	11.28	**<.001**	0.49	0.05	10.33	**<.001**	0.56	0.09	6.02	**<.001**
Age 11-17	0.65	0.05	12.38	**<.001**	0.58	0.05	12.38	**<.001**	0.39	0.10	3.74	**<.001**
Age 18-24	0.53	0.05	11.53	**<.001**	0.52	0.04	12.81	**<.001**	0.67	0.08	8.16	**<.001**
Age 25-39	0.19	0.04	5.13	**<.001**	0.18	0.03	5.40	**<.001**	0.33	0.07	4.69	**<.001**
Age 40-64						*Reference*						
Age 65+	0.15	0.06	2.39	**.017**	0.17	0.06	3.06	**.002**	0.13	0.11	1.18	.239
Female	-0.06	0.03	-2.45	**.014**	-0.04	0.02	-1.78	.074	0.23	0.05	4.93	**<.001**
Black	0.18	0.08	2.44	**.015**	0.18	0.07	2.73	**.006**	0.26	0.12	2.20	**.028**
Prospective morbidity score	0.08	0.01	11.70	**<.001**	0.07	0.01	11.91	**<.001**	0.07	0.01	6.66	**<.001**
Any PCP visit in base yr.	0.11	0.04	2.98	**.003**	0.11	0.03	3.41	**.001**	0.18	0.06	2.86	**.004**
Any ED visit in base yr.	0.79	0.03	25.94	**<.001**	0.74	0.03	27.66	**<.001**	1.02	0.05	22.17	**<.001**
Any inpatient stay in base yr.	-0.11	0.06	-1.75	.080	-0.12	0.06	-2.11	**.035**	-0.15	0.12	-1.33	.183
Plan 1	-0.06	0.04	-1.35	.178	-0.05	0.04	-1.20	.231	-0.10	0.08	-1.23	.219
Plan 2	-0.05	0.05	-0.93	.354	-0.01	0.04	-0.13	.894	-0.03	0.09	-0.32	.747
Plan 3	0.02	0.05	0.35	.729	-0.18	0.04	-4.22	**<.001**	-0.33	0.08	-3.92	**<.001**
Plan 4						*Reference*						
PCP type: Maternal/pediatric	0.08	0.04	2.07	**.039**	0.06	0.03	1.88	.060	0.14	0.07	1.88	.059
PCP type: Family med						*Reference*						
PCP type: Internal med	0.04	0.03	1.16	.245	0.03	0.03	0.88	.379	0.05	0.06	0.87	.387
PCP type: Other	-0.27	0.40	-0.69	.492	-0.16	0.34	-0.46	.648	0.11	0.84	0.14	.891
Mean PCP quality score	-0.06	0.02	-3.86	**<.001**	-0.06	0.01	-4.34	**<.001**	-0.05	0.03	-1.96	.050
Extra distance to PCP	0.00	0.00	1.17	.243	0.00	0.00	2.46	**.014**	0.01	0.00	2.13	**.033**
CT: % living below poverty	0.01	0.00	6.18	**<.001**	0.01	0.00	5.84	**<.001**	0.02	0.00	6.20	**<.001**
PL: Arthritis	0.12	0.05	2.45	**.014**	0.10	0.04	2.32	**.020**	0.06	0.09	0.67	**.500**
PL: Asthma	0.34	0.03	9.85	**<.001**	0.29	0.03	9.50	**<.001**	0.51	0.06	8.95	**<.001**
PL: Cancer	-0.06	0.03	-1.74	**.082**	-0.06	0.03	-2.10	**.036**	-0.17	0.07	-2.53	**.011**
PL: CHF	0.45	0.15	2.99	**.003**	0.47	0.12	3.88	**<.001**	0.98	0.21	4.73	**<.001**
PL: COPD	0.16	0.17	0.95	**.340**	0.23	0.14	1.58	**.114**	0.26	0.31	0.82	**.412**
PL: Depression	0.39	0.04	11.00	**<.001**	0.36	0.03	11.53	**<.001**	0.38	0.06	6.34	**<.001**
PL: Diabetes	0.07	0.04	2.00	**.046**	0.04	0.03	1.08	**.278**	0.05	0.06	0.78	**.437**
PL: HTN	0.07	0.03	2.10	**.036**	0.07	0.03	2.10	**.035**	0.11	0.07	1.67	**.095**
PL: Overweight	0.07	0.04	1.68	**.093**	0.09	0.04	2.54	**.011**	0.02	0.08	0.31	**.757**
PL: Smoker	0.29	0.03	8.60	**<.001**	0.28	0.03	9.37	**<.001**	0.32	0.06	5.44	**<.001**
Constant	-2.49	0.05	-46.12	**<.001**	-2.32	0.13	-17.38	**<.001**	-3.61	0.10	-35.74	**<.001**

CHF: Congestive heart failure; coef: coefficient; COPD: Chronic obstructive pulmonary disease; CT: Census tract; HTN: Hypertension; MCN: managed care network; PCP: Primary care provider; PL: problem list; SE: standard error; z: z-score (standardized coefficient). **Bold** indicates statistically significant *P*-values.

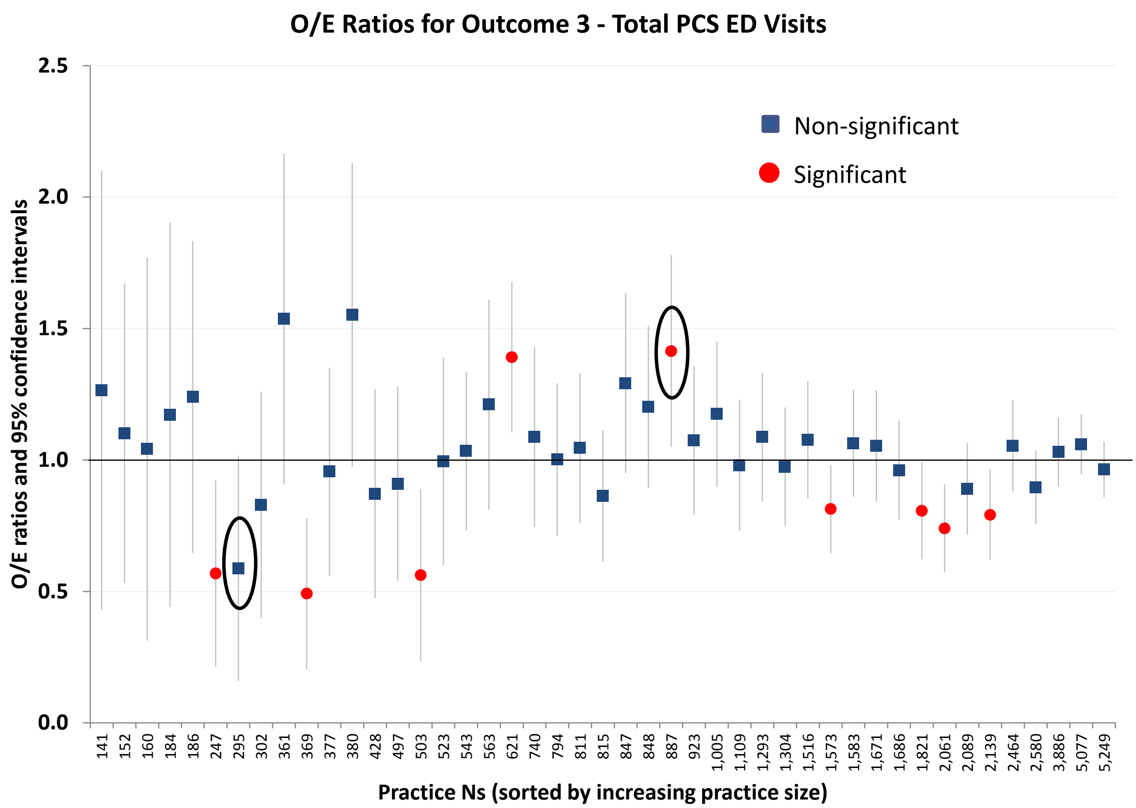
Several clinical conditions (CCs) from base-year claims were also significantly (*P*<.001) associated with increased risk of visiting the ED during the prediction year in stepwise multiple regression analyses, including: torn ligament in knee (AOR [95% CI] 1.49 [1.20 to 1.85]), back pain (1.34 [1.23 to 1.45]), ankle sprain (1.32 [1.11 to 1.58]), depression (1.51 [1.35 to 1.69]), ADD (1.50 [1.20 to 1.87]), asthma (1.46 [1.33 to 1.61]), other gastrointestinal disorders (1.23 [1.14 to 1.32]), and non-chronic ear-nose-throat disorders (1.19 [1.12 to 1.26]) (data not shown).

The relationship between ED utilization and neighborhood income persisted in models that included age, sex, race, prior use, morbidity score, EMR-based problem list conditions, payor, provider specialty, and provider quality. In our most predictive enhanced models, for every 1-percentage-point increase in the percent living in poverty in a Census tract, the AOR for outcome 1—any ED visit increased by 1.01 (95% CI: 1.01-1.01). Outcome 2—total ED visits, increased by 0.01 (±0.001), and outcome 3—total PCS ED visits, increased by 0.02 (±0.003) (**Table 4-3**).

In **Figures 4-5** and **4-6**, we show the practice level differences in O/E ratios when regression models predicting ED utilization include or exclude the neighborhood poverty variable. For this analysis, we specified two sets of regression models for each of the 3 outcome measures. The first models included the percent in poverty in the Census tract as a predictor, and the second set omitted that variable. We then calculated and plotted the two sets of O/E ratios and 95% confidence ratios.

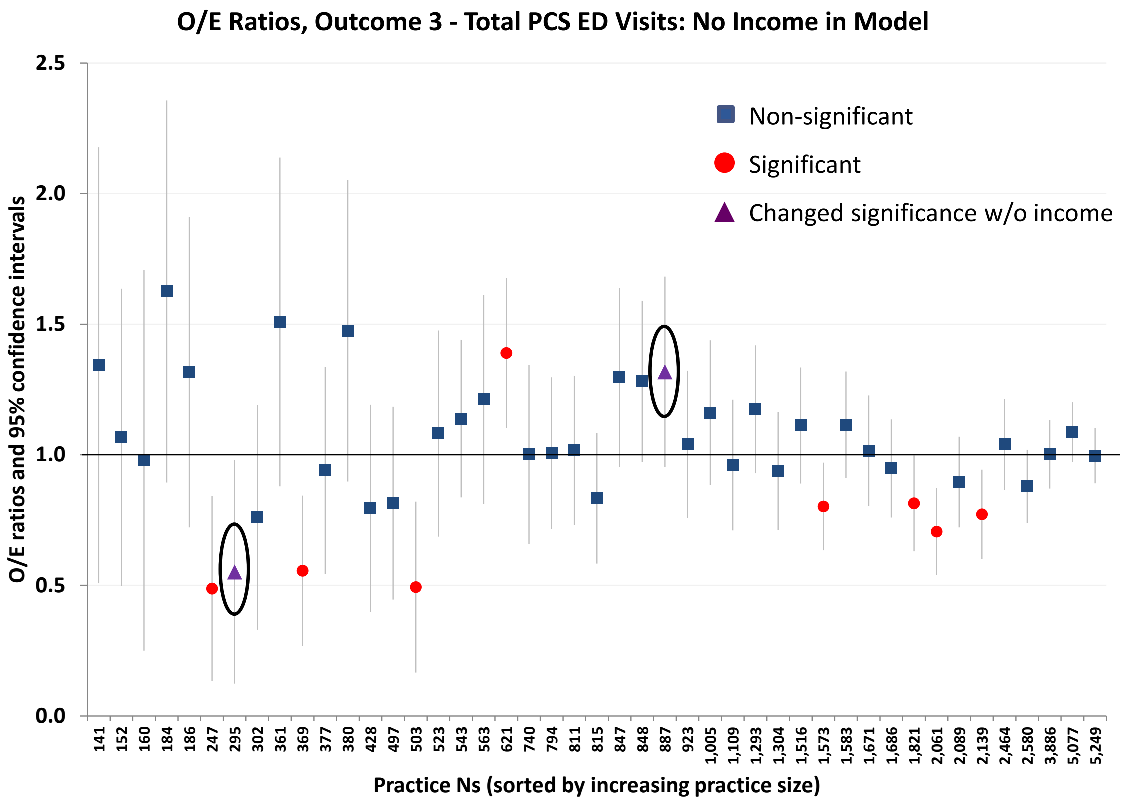
These two figure illustrate the fact that, for 2 out of 45 practices, models predicting outcome 3—total PCS ED visits were sensitive to whether neighborhood poverty was included or not. The two circled practices would have O/E ratios that were significantly different with poverty included than without. If practices were rewarded or penalized for having outlier O/E ratios, practices with a higher proportion of enrollees from low-income neighborhoods would be penalized if their expected-use targets did not take their enrollees' impoverished neighborhoods into account (and the reverse is true as well – practices with wealthier patients would benefit).

Figure 4 - 5. Practice-level observed-to-expected ratios for total PCS ED visits when poverty is included as a predictor



Source: Development data (n=53,112); MCN: managed care network. The two circled practices (discussed in text) are family practices with 295 and 887 patients, respectively. The smaller practice's O/E ratio is not significantly different with poverty included in the model, while the larger practice would be judged as having significantly higher PCS ED use than expected with poverty included. Models adjusted for all covariates listed in Table 4-3 (age, sex, race, morbidity, PCP/ED/inpatient use in base year, payor, PCP type, PCP quality score, extra distance to PCP, neighborhood % in poverty, and 10 conditions from problem lists).

Figure 4 - 6. Practice-level observed-to-expected ratios for total PCS ED visits when poverty is NOT included as a predictor



Source: Development data (n=53,112); MCN: managed care network. The two circled practices (discussed in text) are family practices with 295 and 887 patients, respectively. The smaller practice would be judged as having significantly lower PCS ED use than expected with poverty omitted, while the larger practice would no longer be judged an outlier. Models adjusted for all covariates listed in Table 4-3 except neighborhood % in poverty.

Extra distance to the PCP made no significant difference in terms of likelihood of outcome 1—any ED visit after adjusting for multiple potential confounders. Among enrollees who did visit the ED, outcome 2—total ED visits was 0.003 higher for every 1 mile of extra distance to their PCP (P=.014). Outcome 3—total PCS ED visits was 0.01 higher for every 1 mile of extra distance to their PCP (P=.033).

Discussion

In this study, models enhanced with data about enrollees' providers, payors, clinical conditions, and neighborhoods are better able to predict several aspects of ED utilization, including whether individuals will go to the ED and how many ED visits and PCS ED visits they will have, than models using claims data alone. We demonstrate that neighborhood-level income predicts ED use, even after adjusting for common risk adjustors. People in lower-income neighborhoods are more likely to go to the ED, have more ED visits, and are more likely to use the ED for primary-care sensitive conditions. In addition, if practices were measured on outcome 3—total PCS ED visits, some practices would have markedly different observed-to-expected utilization ratios if neighborhood poverty were included (or omitted) as a covariate in the model. Models predicting ED use should incorporate publicly available neighborhood-level variables, such as percent in poverty, when available. Otherwise, targets for ED use—even if adjusted for traditional "case mix" variables—may be unfair.

We also find that age is one of the strongest predictors of any ED use, with those younger than age 25 being at highest risk of increased utilization (any ED visit, total ED visits, and total PCS ED visits). Using the ED in the previous year is also one of the strongest predictors of future ED utilization. This latter association may reflect some combination of chronic morbidity, individual preferences for the ED as a place of care, and chronic problems accessing care in other settings. In terms of clinical factors, congestive heart failure, asthma, tobacco use, and depression were among the strongest predictors on all 3 measures of ED use. Higher PCP quality scores were associated with a reduced risk of any ED visit (*P*<.001) and fewer ED visits (*P*<.001), but the effect was not statistically significant for PCS ED visits (*P*=.050) in this analysis. Therefore, the evidence that PCP quality is a significant factor predicting total PCS ED use is somewhat equivocal, but the trend toward significance, combined with the results for the other outcomes, strongly suggest that there is a likely association.

Our findings are similar to those reported in prior studies in terms of the associations found between ED utilization and age, race, neighborhood income, and prior ED use.[6,52,75,128,146,147] The association between PCP quality and ED use has been known since at least the 1980s.[148] Of the 4 conditions identified as strongly and consistently associated with increased risk of ED utilization in our study, asthma and depression have been reported as risk factors in prior work,[149-151] whereas congestive heart failure and tobacco use have not been previously reported as risk factors, to our knowledge.

Limitations

This study was limited by a geographically constrained population that included only those persons insured by one of 4 commercial insurers in Massachusetts. Our sample population was mostly white and relatively healthy, with lower ED use than the state average. Massachusetts is also a relatively wealthy state, with 98% of the population covered by health insurance.[51] Thus, our findings may not generalize to other populations.

A limitation of the EMR data we used is that the information represents "present-moment" characteristics and may not accurately reflect enrollees' characteristics at the time of an ED visit (or at the end of the base year, in the case of a prospective model). For example, an enrollee's smoking status in our data may not match what it was at the time of his or her ED visit(s). We were also unable to verify the accuracy of the EMR problem lists.

Another important limitation is that the version of the NYU ED algorithm we used was last updated in 2009.There may have been changes in the ICD-9-CM diagnosis codes used in medical billing since 2009 that the algorithm could not capture. Moreover, using diagnoses to classify individual visits has inherent limitations because including the fact that coding practices vary among by providers. The algorithm itself has been criticized for insensitivity to changes in access to care.[71,73] Future research using our PCS measure is needed to determine whether the methods we propose are better able to capture such changes.

Finally, one of the features of our analysis was that we treated people with partial-year observations the same as those with full-year observations. We know, from the MCN's records, that most enrollees were present for all 12 months of both the base and prediction periods, and the vast majority had at least 6 months of observation in each period. However, because of technical issues during the file construction phase of this research, we were not able to incorporate the number of months each enrollee was present into our analyses.

Partial-year observations have several implications for practical implementation of these methods. For retrospective analyses (to set benchmarks for providers based on two years of recent data for the purposes of performance measurement and quality improvement efforts), treating people with partial-year observations the same as those with full-year observations could bias the model's benchmarks for providers with more individuals with partial-month data. Future research on implementation of a PCS ED use performance measure would require a partial-eligibility data analysis. One option for handling the potentially unobserved utilization in the target year would be to calculate annualized rates using the number of months of eligibility divided by the number of months in the prediction year as a weight (i.e., eligibility fractions).[110]

Predictive models meant to be used for case management purposes would be built in a very similar manner to the models we used in this analysis, since analysts in concurrent predictive scenarios do not know, in advance, for how many months in the future an enrollee will remain enrolled. However, when using predictive modeling to support case management efforts, having missing months of eligibility in the base year presents a problem when calculating risk scores. If we are missing months of observation, we may be biasing the risk scores downward (underestimating morbidity, and thus risk of future expenditures and/or utilization). We are not aware of any published research that provides a solution for appropriately dealing with partial-year eligibility in the base year.

Conclusions

As of April 2013, more than half of US physicians had transitioned to using electronic medical records (EMRs).[152] Wider implementation of EMRs provides researchers with opportunities to use linked EMR and claims data to better understand ED risk factors not captured in administrative data alone. In this study, EMR data provided richer insight into enrollees' morbidity through problem list (PL) entries, which revealed associations between depression, tobacco use, asthma, and congestive heart failure and the risk of ED utilization.

Area-based measures, such as median income and percent of residents living in poverty, are publicly available and have been used in many prior analyses to provide insight into individuals' socioeconomic status.[134-137,153] These measures are not only a reasonable proxy of individual situations, but also measure contextual (social/environmental) predisposing and enabling factors in the residential environment that are relevant in Andersen's model.

Our enhanced models incorporated multiple domains from Andersen's behavioral model: payor and provider characteristics (contextual enabling factors); neighborhood characteristics (contextual predisposing factors); and obesity and tobacco use from enrollees' medical records (health behaviors). Other important aspects—such as enrollees' individual socioeconomic circumstances, adherence to treatment, satisfaction with their PCP, and health beliefs—could not be measured in this study. If the explanatory power of survey data could be added to our models, they would almost certainly be able to predict with greater accuracy. ED utilization is not driven solely by medical need. Although sicker enrollees are more likely to use the ED than healthier enrollees, other factors are also important.

ED risk models allow managed care organizations to set targets for expected PCS use for panels of patients against which actual PCS use can be judged. Such approaches are largely untested. Future research is needed to understand whether these tools can safely reduce PCS ED utilization.

CHAPTER V. DISCUSSION AND CONCLUSIONS

Summary

Individuals are using emergency departments (EDs) more today than ever before.[11] Approximately half of all ED visits are primary-care sensitive (PCS) – meaning that they could potentially be avoided with timely, effective primary care.[48,50,93,125,126] Reducing nonessential ED visits is important: EDs are overcrowded, and care in the ED, compared to care in a primary care setting, is usually uncoordinated, lacks follow-up, and costs more. To reduce PCS ED use, we must be able to define, measure, and predict such use in a population.

In this study, we introduce a measure of ED use that estimates the number of ED visits that are potentially sensitive to primary care, based on the NYU ED algorithm. Typically, most analysts measure ED utilization either with a simple binary indicator for any ED visit during a set period, or by counting the number of ED visits in a set period to define “frequent visitors”. These measures count necessary and undesirable ED use equally. Moreover, measuring only whether a person had any ED use does not allow us to examine magnitudes and variations in intensity.

To measure PCS ED use, researchers in prior studies using administrative data have typically applied thresholds (such as 0.50 or 0.75) created using the New York University (NYU) ED algorithm visit “scores” (probabilities) to decide when to call ED visits PCS. This method is also problematic. Few visits are 100% likely to have been PCS, so a binary measure is of limited use; applying a threshold requires the researcher to make an arbitrary choice, and that choice affects the results. Secondly, rounding visits with scores below the threshold down to 0 and those with scores above it up to 1 loses important information.

In this study, we introduce a measure of ED use that examines the subset of ED visits that are potentially sensitive to primary care, according to the NYU ED algorithm. Our approach sums the probabilities the algorithm generates for each person across all of that person's ED visits to create a PCS ED use outcome that is an estimate of the total number of PCS ED visits. We found that measuring PCS ED use in this way allows us to predict PCS ED use with high specificity, which is a critical step toward the goal of reducing unnecessary ED use. In demonstrating this method, we find that 1) most practices in our sample had lower-than-expected PCS ED visits; and 2) much of the variation in observed ED use (but not the observed-to-expected ratio) can be explained by variation within the population, as reflected in the estimated expected use. Thus, our results help make the case for the importance of setting risk-adjusted targets for ED utilization.

We have developed models to predict any ED visit and the estimated number of PCS ED visits during a 6-month period, using only claims and administrative data, in a large, nationwide sample of commercially insured individuals. We find that models predicting the number of PCS ED visits have a higher R^2 than models predicting any ED visit. Among the 0.5% of the population at greatest predicted risk of ED use (the top risk group), about 40-50% had any overall or PCS ED use in the prediction period.

Using an enhanced dataset, including factors related to enrollees’ providers, payors, problem list conditions from the electronic medical record (EMR), and neighborhoods, we predicted ED utilization in a smaller sample of managed-care network (MCN) enrollees. The enhanced models more accurately predicted any ED use and PCS ED use than models using only administrative data. With the enhanced data, we have demonstrated that neighborhood-level poverty predicts ED use, even after adjusting for common risk factors such as age, sex, and morbidity. People in lower-income neighborhoods are more likely to go to the ED, have more ED visits, and are more likely to use the ED for PCS conditions. Models predicting ED use to set targets should therefore incorporate publicly available neighborhood-level variables (i.e., contextual enabling factors) such as percent in poverty, when available. Otherwise, targets for ED use—even if adjusted for traditional "case mix" variables—may be unfair.

Our models also benefited from incorporating data on conditions in EMR problem lists (PLs). We find that, all else being equal, asthma, congestive heart failure, depression, and tobacco use are among the strongest predictors of all 3 measures of ED use. As more physicians transition to EMRs, this type of data will become more commonly used in research and quality improvement.

In our literature review, we found a few characteristics reported in more than one study to predict PCS use after adjusting for other characteristics, including being female, over 65, African American, and covered by Medicaid. Our results are largely consistent with the prior literature, although we did not find a consistent or strong effect of sex. Additionally, our study samples contained few individuals over age 65 and no Medicaid recipients. We also found that black race was significantly associated with increased ED utilization on all 3 measures, but the effect was less than that of age, morbidity, neighborhood-level poverty, or prior ED use (as determined by comparing standardized coefficients, or z-scores). Higher PCP quality scores were associated with a reduced risk of ED utilization.

Strengths and Limitations

Patient-centered medical home demonstrations, and other practice and payment reform systems, have targeted ED utilization for reduction. However, some ED visits are both necessary and desirable. Measuring, predicting, and attempting to reduce PCS ED use, rather than overall ED use, may be a better approach; however, no prior studies have evaluated this question. This study’s primary innovation was to explore the relative merits of measuring overall ED use versus the subset of PCS ED use. We evaluated different models' abilities to predict different types of use, and reported the robustness of different measures of ED use, in a real-world, managed-care setting.

This study contributes to our understanding of the prevalence and predictors of ED use in 2 populations: a managed-care population in Massachusetts; and a nationwide sample of commercially insured individuals. These data sources, which include claims, EMR, and survey data, contribute to the study’s innovation. However, our MCN study was limited by the geographically constrained population, and it included only those persons insured by one of 4 commercial insurers. Our MCN sample population was mostly white and relatively healthy, with lower ED use than average. Massachusetts is also a relatively wealthy state, with 98% of the population covered by health insurance. Our MCN findings may not be generalizable to other populations.

Another important limitation of our data is the lack of information about certain factors that the literature suggests may influence ED use. These factors include patient preferences, convenience, satisfaction with one's usual source of care, problems accessing timely primary care, and lack of education about other options for accessing care.[6,19,21,29-31,63,83] Measuring most of these aspects of patient experience requires primary data collection (i.e., surveys), which was outside of the scope of this study, as was surveying the MCN's primary care practices about operating hours and availability of same-day appointments.

We used neighborhood characteristics as covariates to improve the predictive power of our models, finding a strong effect of neighborhood poverty. Although a person's individual income (which is usually unknown in real-world managed-care settings) may differ substantially from the neighborhood average, the average is a reasonable proxy for individual income and is a good measure, in its own right, of an aspect of his or her residential environment.[134]

Our eligibility inclusion criteria required only 1 month of eligibility in the base period and 1 month in the prediction period. This approach has several advantages: it includes those who were born or died during either period, it maximizes the sample size, and it is consistent with an implementation-oriented approach using real-world data. If we had required individuals to be eligible for the entire base and prediction periods, we would not have learned important information about the heightened risks associated with infants under 1 year old. In prior research, others have found that risk adjustment models have improved predictive ability when enrollees with partial data are included, compared to requiring full-year eligibility.[89] Moreover, the expected users of these models and methods are analysts using the same type of partially complete data as we used in this study, and having demonstrated somewhat predictive results in these data shows that our results are robust to the types of partial data other analysts would be expected to encounter themselves.

For retrospective models used to set benchmarks for providers, treating people with partial-year observations the same as those with full-year observations could bias the model's benchmarks for providers with more individuals with partial-month data. If we were implementing a performance measure, we would want to perform a partial-eligibility data analysis so that we could calculate annualized rates using the number of months of eligibility divided by 12 as a weight (i.e., eligibility fractions).[110]

For using predictive modeling to support case management efforts, having missing months of eligibility in the base year presents a problem when calculating risk scores. If we are missing months of observation, we may be biasing the risk scores downward (underestimating morbidity, and thus risk of future expenditures and/or utilization). We are not aware of any published research that provides a solution for appropriately dealing with partial-year eligibility in the base year.

Implications

Emergency department services account for approximately 5-10% of all healthcare expenditures in the US.[2] In general, an ED visit is more expensive than comparable care received in other ambulatory settings, such as hospital outpatient departments or physician offices.[26] Thus, policymakers and payors have concerns about both the cost and health implications of overuse and inappropriate use of EDs, particularly for persons with limited access to other ambulatory care. Many are also concerned about the potential impact of national health reform on ED use. Reviewing the effects of reform in Massachusetts on ED use provides a preview of what may lie in store for the US after 2014.

In pre-reform Massachusetts (Fall 2006), 34% of adults age 18-64 visited the ED in the prior year, and 16% said that their most recent ED visit was for a nonemergency condition. Post-reform (Fall 2009), there were no significant differences.[51] When asked why they visited the ED for a nonemergency, 55% reported going because they were unable to get an appointment as soon as one was needed.[32] Difficulty in getting a timely PCP appointment may partially explain why there was no change in the percentage of people using hospital emergency departments for nonemergencies after health reform in Massachusetts.

As in Massachusetts, although the Affordable Care Act (ACA) covers more people with public and private insurance, it is likely to lead to more ED use, not less. Prior studies have found that privately insured individuals and those enrolled in Medicaid have the highest ED use rates, whereas the uninsured have the lowest.[4,11,52,154] Nationwide shortages of PCPs will likely continue to restrict access to timely primary care as the demand from the newly insured is added to an already strained system.[155]

In our current, fragmented, fee-for-service payment system, stakeholders generally lack the incentives, tools, or ability to reduce ED use. Hospitals, in particular, have little incentive to reduce ED use, since the ED is frequently a profit center for the organization.[156] PCPs often do not have the data or analytic capacity to identify which of their enrollees are using the ED for care. Moreover, they rarely have any incentive not to refer patients to the ED. Physicians may refer patients to the ED because of the desire to avoid longer office hours, reluctance to take on complicated cases, lack of diagnostic equipment in their offices, or concerns about malpractice liability.[1] PCPs may also appropriately refer to the ED when their patients have complex or serious conditions that warrant emergency care. In one study, referral by PCPs to a pediatric ED was significantly and independently associated with illness severity and higher resource use, such as diagnostic testing, intravenous fluids, and medications.[157]

Payors have few levers for influencing ED utilization. They can refuse to pay for certain visits, set limits on visits, or increase patient cost-sharing; or they can attempt to change behavior through educational campaigns, toll-free hotlines, or case management. Legislatures and regulators in Washington, Tennessee, Iowa, New Hampshire, and Illinois have considered or enacted measures that would limit payment for nonemergency ED visits by Medicaid enrollees, based on discharge diagnosis.[47] In 2011, Washington State's Medicaid agency proposed a 3-visit limit for any of about 700 conditions, and when that proposal was blocked by a judge,[158] they went even further. In 2012, they proposed to refuse to pay for any visits for about 500 conditions deemed nonemergent.[159] After further negotiations with providers, the state legislature passed a bill that attempts to reduce ED use through improving technology and information sharing, educating patients, identifying frequent users, better coordinating care, and soliciting provider feedback.[73]

Both public and private payors can and do raise copayments for ED visits, and the result is generally a decrease in ED visits.[65,160,161] Arizona, Oregon, Illinois, Iowa, Nebraska, North Carolina, and New Mexico have recently implemented or considered implementing some level of copayment requirement for nonemergency use of the ED by Medicaid enrollees, despite recent evidence suggesting that the strategy may be ineffective.[47,162] Moreover, the strategy carries a risk of patients' not going to the ED for a true emergency because they are unable to afford the copayment. However, evidence from research in a commercially insured population shows that raising copayments can reduce ED visits without increasing hospitalizations, intensive care unit admissions, or mortality.[38] Ideally, such cost-shifting would be accompanied by an increase in same-day PCP availability, extended (evening and weekend) hours, or both.

Case management (also called care coordination) is a strategy for delivering more comprehensive and coordinated care to at-risk persons, who often have multiple chronic diseases or other risk factors.[114] Typically, case management attempts to address patient needs through a multidisciplinary approach involving medical, nursing, social work, and mental health providers. Although there is limited evidence of its efficacy in preventing ED utilization, this may be related to difficulty identifying persons at risk.[113] Our hope is that improved methods will lead to more accurate identification of people at highest risk and avert unnecessary ED use by providing better care at a lower cost.

In the December 2013 issue of Health Affairs, Kellermann and colleagues wrote that the safest and surest way to reduce ED use is to improve patients' access to primary care.[12] If PCPs (physicians and their physician assistant and nurse practitioner colleagues) can expand after-hours access, do a better job of managing chronic conditions and educating patients about when to use the ED, reduce ED use, and get rewarded for it, both patients and providers could benefit.

Our research was designed to support performance improvement efforts in a managed-care environment to reward PCPs for reducing their patients' use of the ED. In future implementations of these methods, we expect that providers will weigh the potential rewards they will get if they lower ED use in their panel against the costs inherent in expanding after-hours and same-day access to their practice, educating patients about when to seek care in the ED, referring fewer patients away (and, in turn, taking risks associated with dealing with acute care), possibly seeing more patients, or trying new approaches such as electronic and group visits.

The issue of whether to risk-adjust provider performance measures for socioeconomic status (SES), as we have done in this dissertation, is controversial. As of 2014, the National Quality Forum (NQF) measure evaluation criteria indicate that factors related to disparities in care should not be included in risk adjustment models for outcome performance measures.[163] There are at least two different views on adjusting for differences in SES, race, and ethnicity:

1) Adjustment may obscure potential problems in equitable care and outcomes, so analyses should instead be stratified to identify disparities in care.

2) Adjustment is essential for fair comparisons among providers to account for factors beyond their control that influence patient outcomes.

In our research, we have shown that some providers would be judged differently if neighborhood-level poverty were omitted from our models. The extent to which these differences in judgment would affect practice patterns is unknown, but penalizing providers simply because their patients are poor is surely an unappealing outcome. Additionally, there are concerns that providers might avoid serving disadvantaged populations to prevent being labeled a poor performer, which would affect access to care for those populations; in addition, bonus and award payments might shift from those who serve the disadvantaged to those who care for the affluent, leaving safety net providers with fewer resources to care for their vulnerable patients. In a world in which provider quality is publicly reported, consumers might also avoid providers who serve disadvantaged populations if they are labeled as poor performers.

In Conclusion

It is possible to identify patients at high risk for ED use, thereby providing a target for efforts to reduce the number of ED visits. However, the distinction between reducing overall ED use and reducing undesirable ED use is key. If providers and payors can accurately evaluate the near-term risk of PCS ED use (i.e., undesirable use) in a population, they can target high-risk patients with educational and care management programs to try to prevent unnecessary ED visits. In addition, ED risk models allow managed care organizations to set targets for expected PCS use for panels of patients against which actual PCS use can be judged. Such approaches are largely untested. Future research is needed to understand whether these tools can safely reduce PCS ED utilization.

REFERENCES

1. IOM Committee on the Future of Emergency Care in the US Health System. Hospital-Based Emergency Care: At the Breaking Point. 2007; 424-424. Available at: http://www.iom.edu/reports/2006/emergency-medical-services-at-the-crossroads.aspx. Accessed Feb. 5, 2014.

2. Lee MH, Schuur JD, Zink BJ. Owning the cost of emergency medicine: beyond 2%. *Ann Emerg Med.* Nov 2013;62(5):498-505 e493.

3. Delia D, Cantor JC. Emergency department utilization and capacity. *Synth Proj Res Synth Rep.* 2009;17(July):45929.

4. Cunningham PJ, May J. Insured Americans Drive Surge in Emergency Department Visits. *Med Care.* 2003(70):1-6.

5. Weber EJ, Showstack Ja, Hunt Ka, et al. Are the uninsured responsible for the increase in emergency department visits in the United States? *Ann Emerg Med.* 2008;52(2):108-115.

6. Cunningham PJ. What accounts for differences in the use of hospital emergency departments across U.S. communities? *Health Aff.* 2006;25(5):w324-336.

7. Tang N, Stein J, Hsia RY, Maselli JH, Gonzales R. Trends and Characteristics of US Emergency Department Visits, 1997-2007. *JAMA.* 2010;304(6):664-670.

8. Cheung PT, Wiler JL, Lowe RA, Ginde AA. National Study of Barriers to Timely Primary Care and Emergency Department Utilization Among Medicaid Beneficiaries. *Ann Emerg Med.* Mar 12 2012.

9. Rust G, Ye J, Baltrus P, Daniels E, Adesunloye B, Fryer GE. Practical Barriers to Timely Primary Care Access: Impact on Adult Use of Emergency Department Services. *Arch Intern Med.* 2008;168(15):1705-1710.

10. Capp R, Rooks SP, Wiler JL, Zane RD, Ginde AA. National study of health insurance type and reasons for emergency department use. *Journal of general internal medicine.* Apr 2014;29(4):621-627.

11. National Center for Health Statistics. National Hospital Ambulatory Medical Care Survey Emergency Department Summary Tables. 2013; http://www.cdc.gov/nchs/ahcd/web_tables.htm. Accessed Jan. 12, 2014.

12. Kellermann AL, Hsia RY, Yeh C, Morganti KG. Emergency care: then, now, and next. *Health Affairs.* 2013;32(12):2069-2074.

13. Kaskie B, Obrizan M, Cook EA, et al. Defining emergency department episodes by severity and intensity: A 15-year study of Medicare beneficiaries. *BMC Health Serv Research.* 2010;10:173-173.

14. General Accounting Office. Hospital Emergency Departments: Crowded Conditions Vary among Hospitals and Communities. 2003; 1-71. Available at: www.gao.gov/new.items/d03460.pdf. Accessed Feb. 5, 2014.

15. Wilson MJ, Siegel B, Williams M. Perfecting Patient Flow. 2005; http://www.rwjf.org/en/research-publications/find-rwjf-research/2005/01/perfecting-patient-flow.html. Accessed Feb. 5, 2014.

16. Carrier E, Yee T, Holzwart RA. Coordination between emergency and primary care physicians. 2011; 1-11. Available at: http://www.nihcr.org/ED-Coordination.html. Accessed Feb. 5, 2014.

17. Billings JC, Parikh N, Mijanovich T. Emergency Department Use in New York City: A Survey of Bronx Patients. 2000; http://www.commonwealthfund.org/Content/Publications/Issue-

Briefs/2000/Nov/Emergency-Department-Use-in-New-York-City--A-Survey-of-Bronx-Patients.aspx. Accessed Feb. 5, 2014.
18. DeSalvo A, Rest SB, Nettleman M, Freer S, Knight T. Patient education and emergency room visits. *Clin Perform Qual Health Care.* 2000;8(1):35-37.
19. Gill JM, Riley AW. Nonurgent use of hospital emergency departments: urgency from the patient's perspective. *J Fam Pract.* 1996;42(5):491-496.
20. Redstone P, Vancura JL, Barry D, Kutner JS. Nonurgent use of the emergency department. *J Ambul Care Manage.* 2008;31(4):370-376.
21. Young GP, Wagner MB, Kellermann AL, Ellis J, Bouley D. Ambulatory visits to hospital emergency departments: Patterns and reasons for use. *JAMA.* 1996;276(6):460-465.
22. McWilliams A, Tapp H, Barker J, Dulin M. Cost analysis of the use of emergency departments for primary care services in Charlotte, North Carolina. *NC Med J.* 2011;72(4):265-271.
23. Baker LC, Baker LS. Excess cost of emergency department visits for nonurgent care. *Health Aff.* 1994;13(5):162-171.
24. Mehrotra A, Liu H, Adams JL, et al. Comparing costs and quality of care at retail clinics with that of other medical settings for 3 common illnesses. *Ann Intern Med.* 2009;151(5):321-328.
25. Weinick RM, Burns RM, Mehrotra A. Many emergency department visits could be managed at urgent care centers and retail clinics. *Health Aff.* 2010;29(9):1630-1636.
26. New England Healthcare Institute. A Matter of Urgency: Reducing Emergency Department Overuse. 2010; http://www.nehi.net/publications/55/a_matter_of_urgency_reducing_emergency_department_overuse. Accessed Feb. 5, 2014.
27. Baker DW, Stevens CD, Brook RH. Determinants of emergency department use by ambulatory patients at an urban public hospital. *Ann Emerg Med.* 1995;25(3):311-316.
28. Matteson KA, Weitzen SH, Lafontaine D, Phipps MG. Accessing care: use of a specialized women's emergency care facility for nonemergent problems. *J Womens Health (Larchmt).* 2008;17(2):269-277.
29. Northington WE, Brice JH, Zou B. Use of an emergency department by nonurgent patients. *Am J Emerg Med.* 2005;23(2):131-137.
30. Fuda KK, Immekus R. Frequent users of Massachusetts emergency departments: a statewide analysis. *Ann Emerg Med.* Jul 2006;48(1):9-16.
31. Lowe RA, Fu R, Ong ET, et al. Community characteristics affecting emergency department use by Medicaid enrollees. *Med Care.* 2009;47(1):15-22.
32. Long SK, Masi PB. Access and affordability: an update on health reform in Massachusetts, fall 2008. *Health Aff.* 2009;28(4):w578-587.
33. Marte J. 10 things emergency rooms won't tell you: Why it takes so long and costs so much to get care in the E.R. *MarketWatch*. New York City, NY: Wall Street Journal; 2013.
34. Tuers DM. Defensive medicine in the emergency department: increasing health care costs without increasing quality? *Nursing administration quarterly.* 2013;37(2):160-164.
35. Engel KG, Heisler M, Smith DM, Robinson CH, Forman JH, Ubel PA. Patient comprehension of emergency department care and instructions: are patients aware of when they do not understand? *Annals of emergency medicine.* 2009;53(4):454-461. e415.
36. Adeoye O, Hornung R, Khatri P, Kleindorfer D. Recombinant Tissue-Type Plasminogen Activator Use for Ischemic Stroke in the United States A Doubling of Treatment Rates Over the Course of 5 Years. *Stroke.* 2011;42(7):1952-1955.
37. Lowe RA, Fu R, Gallia CA. Impact of policy changes on emergency department use by Medicaid enrollees in Oregon. *Medical care.* 2010;48(7):619-627.

38. Hsu J, Price M, Brand R, et al. Cost Sharing for Emergency Care and Unfavorable Clinical Events: Findings from the Safety and Financial Ramifications of ED Copayments Study. *Health services research.* 2006;41(5):1801-1820.

39. LaCalle E, Rabin E. Frequent users of emergency departments: the myths, the data, and the policy implications. *Annals of emergency medicine.* 2010;56(1):42-48.

40. Doupe MB, Palatnick W, Day S, et al. Frequent Users of Emergency Departments: Developing Standard Definitions and Defining Prominent Risk Factors. *Annals of emergency medicine.* 2012;60(1):24-32.

41. Weber EJ. Defining Frequent Use: The Numbers No Longer Count. *Ann Emerg Med.* Mar 22 2012.

42. Althaus F, Paroz S, Hugli O, et al. Effectiveness of interventions targeting frequent users of emergency departments: A systematic review. *Ann Emerg Med.* 2011;58(1):41-52.e42.

43. Blank FSJ, Li H, Henneman PL, et al. A descriptive study of heavy emergency department users at an academic emergency department reveals heavy ED users have better access to care than average users. *J Emerg Nurs.* 2005;31(2):139-144.

44. Billings J, Raven MC. Dispelling An Urban Legend: Frequent Emergency Department Users Have Substantial Burden Of Disease. *Health Affairs.* December 1, 2013 2013;32(12):2099-2108.

45. Garan Lucow Miller PC. Frequently Asked Questions about the Emergency Medical Treatment and Active Labor Act (EMTALA). 2012; http://www.emtala.com/faq.htm. Accessed 5/31/12.

46. Rosenbaum S. The enduring role of the Emergency Medical Treatment And Active Labor Act. *Health Affairs.* 2013;32(12):2075-2081.

47. Raven MC, Lowe RA, Maselli J, Hsia RY. Comparison of Presenting Complaint vs Discharge Diagnosis for Identifying "Nonemergency" Emergency Department Visits. *JAMA.* 2013;309(11):1145-1153.

48. Utah Office of Health Care Statistics. Primary Care Sensitive Emergency Department Visits in Utah, 2001. 2004; http://health.utah.gov/hda/report/ER.php. Accessed Feb. 5, 2014.

49. Blue Cross Blue Shield of Michigan. Patient-centered Medical Homes. 2012; http://www.bcbsm.com/portal/pcmh/newsletter.shtml. Accessed 5/12/12.

50. Division of Health Care Finance and Policy. Preventable/Avoidable Emergency Department Use in Massachusetts Fiscal Years 2004 to 2008. 2010; http://www.measureandpredict.com/wp-content/uploads/2014/04/MA-preventable_avoidable_ed_use_2004_2008.pdf. Accessed 4/1/14.

51. Long SK, Stockley K. Sustaining health reform in a recession: an update on Massachusetts as of fall 2009. *Health Aff.* 2010;29(6):1234-1241.

52. Cunningham PJ, Clancy CM, Cohen JW, Wilets M. The use of hospital emergency departments for nonurgent health problems: a national perspective. *Med Care Res Rev.* 1995;52(4):453-474.

53. Sarver JH, Cydulka RK, Baker DW. Usual source of care and nonurgent emergency department use. *Acad Emerg Med.* 2002;9(9):916-923.

54. Lowe RA, Schull M. On easy solutions. *Ann Emerg Med.* Sep 2011;58(3):235-238.

55. Centers for Medicare and Medicaid Services. National health expenditure data. 2014; http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical.html. Accessed Jan. 14, 2014.

56. Institute of Medicine. *Rewarding Provider Performance: Aligning Incentives in Medicare.* Washington, DC: Institute of Medicine;2006.

57. Stange KC, Nutting PA, Miller WL, Jaén CR, Gill JM. Defining and measuring the patient-centered medical home. *Journal of general internal medicine.* 2010;25(6):601-612.

58. TransforMed. The TransforMED Patient-Centered Model. 2009; http://www.transformed.com/pdf/TransforMEDMedicalHomeModel-letter.pdf. Accessed Feb. 5, 2014.

59. Grumbach K. Outcomes of Implementing Patient Centered Medical Home Interventions: A Review of the Evidence from Prospective Evaluation Studies in the United States. 2010. http://forwww.pcpcc.net/files/evidence_outcomes_in_pcmh_2010.pdf. Accessed Feb. 5, 2014.

60. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? *J Health Social Behavior.* 1995;36(1):1-10.

61. Andersen RM, Davidson PL. Improving access to care in America. *Changing the US health care system: key issues in health services policy and management. 3a. edición. San Francisco: Jossey-Bass.* 2007:3-31.

62. McCusker J, Karp I, Cardin S, Durand P, Morin J. Determinants of Emergency Department Visits by Older Adults: A Systematic Review. *Acad Emerg Med.* 2003;10(12):1362-1370.

63. Padgett DK, Brodsky B. Psychosocial factors influencing non-urgent use of the emergency room: a review of the literature and recommendations for research and improved service delivery. *Soc Sci Med.* 1992;35(9):1189-1197.

64. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health.* Jun 1998;52(6):377-384.

65. Wharam JF, Landon BE, Galbraith AA, Kleinman KP, Soumerai SB, Ross-degnan D. Emergency Department Use and Subsequent Hospitalizations Among Members of a High-Deductible Health Plan. *JAMA.* 2007;297(10):1093-1102.

66. NYU Center for Health and Public Service Research. The NYU ED Algorithm. 2011; http://wagner.nyu.edu/faculty/billings/acs-algorithm. Accessed Feb. 5, 2014.

67. Bindman AB, Grumbach K, Osmond D, et al. Preventable hospitalizations and access to health care. *JAMA.* 1995;274(4):305-311.

68. Ballard DW, Price M, Fung V, et al. Validation of an algorithm for categorizing the severity of hospital emergency department visits. *Med Care.* 2010;48(1):58-63.

69. Garcia TC, Bernstein AB, Bush MA. Emergency department visitors and visits: Who used the emergency room in 2007? 2010; http://www.cdc.gov/nchs/data/databriefs/db38.htm. Accessed 5/12/12.

70. Jones K, Paxton H, Hagtvedt R, Etchason J. An Analysis of the New York University Emergency Department Algorithm's Suitability for Use in Gauging Changes in ED Usage Patterns. *Med Care.* 2011;00(00):1-10.

71. Lowe RA, Fu R. Can the emergency department algorithm detect changes in access to care? *Acad Emerg Med.* 2008;15(6):506-516.

72. Lines LM, Ash AS. Predictors of Primary Care Sensitive Emergency Department Visits: A Systematic Review. 2012. http://works.bepress.com/lisa_lines/4/. Accessed Feb. 5, 2014.

73. Washington State Hospital Association. The NYU Classification System for ED Visits: WSHA Technical Concerns. 2010. http://www.wsha.org/files/169/NYU_Classification_System_for_EDVisits.pdf. Accessed Feb. 5, 2014.

74. Lowe RA, Localio AR, Schwarz DF, et al. Association between primary care practice characteristics and emergency department use in a Medicaid managed care organization. *Med Care.* 2005;43(8):792-800.

75. Chiou S-J, Campbell C, Myers L, Culbertson R, Horswell R. Factors influencing inappropriate use of ED visits among type 2 diabetics in an evidence-based management programme. *J Eval Clin Practice.* 2010;16(6):1048-1054.

76. Wolinsky FD, Liu L, Miller TR, et al. Emergency department utilization patterns among older adults. *J Gerontol A Biol Sci Med Sci.* 2008;63(2):204-209.

77. Agency for Healthcare Research & Quality. National Quality Measures Clearinghouse. 2012; http://www.qualitymeasures.ahrq.gov/. Accessed Feb. 5, 2014.

78. Lowe RA, Bindman AB. Judging who needs emergency department care: a prerequisite for policy-making. *Am J Emerg Med.* 1997;15(2):133-136.

79. O'Brien GM, Shapiro MJ, Woolard RNW, et al. "Inappropriate" emergency department use: a comparison of three methodologies for identification. *Acad Emerg Med.* 1996;3(3):252-257.

80. Liu T, Sayre MR, Carleton SC. Emergency Medical Care: Types, Trends, and Factors Related to Nonurgent Visits. *Acad Emerg Med.* 1999;6(11):1147-1152.

81. Grumbach K, Keane D, Bindman A. Primary care and public emergency department overcrowding. *Am J Public Health.* 1993;83(3):372-378.

82. Petersen LA, Burstin HR, O'Neil AC, Orav EJ, Brennan TA. Nonurgent emergency department visits: the effect of having a regular doctor. *Med Care.* 1998;36(8):1249-1255.

83. Pilossoph-Gelb S, Mower WR, Ajaelo I, Yang SC. Psychosocial difficulties and emergency department use. *Acad Emerg Med.* 1997;4(6):589-592.

84. Schwartz MP. Office or emergency department: what's the difference? *South Med J.* 1995;88(10):1020-1024.

85. Wuerz RC, Milne LW, Eitel DR, Travers D, Gilboy N. Reliability and validity of a new five-level triage instrument. *Acad Emerg Med.* 2000;7(3):236-242.

86. Pelletier LR. *Information-Enabled Decision-Making in Health Care: EHR-Enabled Standardization, Physician Profiling and Medical Home*, Worcester Polytechnic Institute; 2010.

87. Briesacher BA, Quittner AL, Fouayzi H, Zhang J, Swensen A. Nationwide trends in the medical care costs of privately insured patients with cystic fibrosis (CF), 2001–2007. *Pediatric Pulmonology.* 2011;46(8):770-776.

88. Truven Health Analytics. MarketScan Studies: Abbreviated Bibliography. 2012; http://sites.truvenhealth.com/bibliography/2012_Thomson_Reuters_MarketScan_Bibliography.pdf. Accessed Feb. 21, 2014.

89. Liu C-F, Pope GC, Ellis RP. *Prospective DCG Risk Adjustment Using Partial Year Data*: Report to the Health Care Financing Administration under Order No. HCFA-96-0558;1997.

90. Pines JM, Newman D, Pilgrim R, Schuur JD. Strategies for integrating cost-consciousness into acute care should focus on rewarding high-value care. *Health Aff (Millwood).* Dec 2013;32(12):2157-2165.

91. Patient-Centered Primary Care Collaborative. PCMH Pilots and Demonstrations in the United States. 2012; http://www.pcpcc.net/pcpcc-pilot-projects. Accessed May 12, 2012.

92. Billings JC, Parikh N, Mijanovich T. Emergency Department Use in New York City: A Substitute for Primary Care? 2000; http://www.commonwealthfund.org/Content/Publications/Issue-Briefs/2000/Mar/Emergency-Department-Use-in-New-York-City--A-Substitute-for-Primary-Care.aspx. Accessed 5/12/12.

93. Billings JC, Parikh N, Mijanovich T. Emergency Room Use: The New York Story. 2000; http://www.commonwealthfund.org/Content/Publications/Issue-Briefs/2000/Nov/Emergency-Room-Use--The-New-York-Story.aspx. Accessed Feb. 5, 2014.

94. Lowe RA. Comment on Ballard DW, Price M, Fung V, et al. Validation of an Algorithm for Categorizing the Severity of Hospital Emergency Department Visits. *Med Care.* 2012;50(4):361-362.

95. Blue Cross Blue Shield of Massachusetts. Making quality health care affordable: The Alternative Quality Contract. 2012; http://bluecrossma.com/visitor/pdf/alternative-quality-contract.pdf. Accessed Feb. 5, 2014.

96. Ash AS, Ellis RP. Risk-adjusted payment and performance assessment for primary care. *Medical care.* 2012;50(8):643-653.

97. Ellis RP, Pope GC, Iezzoni L, et al. Diagnosis-based risk adjustment for Medicare capitation payments. *Health care financing review.* Spring 1996;17(3):101-128.

98. Ash AS, Posner MA, Speckman J, Franco S, Yacht AC, Bramwell L. Using claims data to examine mortality trends following hospitalization for heart attack in Medicare. *Health Serv Res.* Oct 2003;38(5):1253-1262.

99. Zhao Y, Ellis RP, Ash AS, et al. Measuring population health risks using inpatient diagnoses and outpatient pharmacy data. *Health Serv Res.* Dec 2001;36(6 Pt 2):180-193.

100. Zero-inflated negative binomial regression. http://www.ats.ucla.edu/stat/stata/dae/zinb.htm. Accessed April 18, 2014.

101. Diehr P. Methods for analyzing health care utilization and costs. *Annu Rev Public Health.* 1999;20:125-144.

102. Dobson AJ. *An Introduction to Generalized Linear Models, Second Edition*. Boca Raton, FL: CRC Press; 2002.

103. Schwartz M, Ash AS. Empirically Evaluating Risk Adjustment Models. In: Iezzoni L, ed. *Risk Adjustment for Measuring Healthcare Outcomes*. 4th ed. Chicago, IL: Health Administration Press; 2012.

104. Cohen ME, Ko CY, Bilimoria KY, et al. Optimizing ACS NSQIP modeling for evaluation of surgical quality and risk: patient risk adjustment, procedure mix adjustment, shrinkage adjustment, and surgical focus. *Journal of the American College of Surgeons.* Aug 2013;217(2):336-346 e331.

105. Bennett ML, Morath JM, Yarbrough D, Sinard R, Netterville J, Eavey RD. The mortality observed-to-expected ratio in otolaryngology. *Otolaryngology--head and neck surgery : official journal of American Academy of Otolaryngology-Head and Neck Surgery.* Jan 2013;148(1):59-63.

106. Hoot NR, Zhou C, Jones I, Aronsky D. Measuring and forecasting emergency department crowding in real time. *Ann Emerg Med.* Jun 2007;49(6):747-755.

107. Gill JM. Nonurgent use of the emergency department: appropriate or not? *Ann Emerg Med.* 1994;24(5):953-957.

108. Murphy AW. Inappropriate attenders at accident and emergency departments I: definition, incidence, and reasons for attendance. *Fam Pract.* 1998;15(1):23-32.

109. Raven MC, Billings JC, Goldfrank LR, Manheimer ED, Gourevitch MN. Medicaid patients at high risk for frequent hospital admission: real-time identification and remediable risks. *J Urban Health.* 2009;86(2):230-241.

110. Zhao Y, Ash AS, Ellis RP, et al. Predicting pharmacy costs and other medical costs using diagnoses and drug claims. *Med Care.* Jan 2005;43(1):34-43.

111. National Quality Measures C. Ambulatory care sensitive conditions: age-standardized acute care hospitalization rate for conditions where appropriate ambulatory care prevents or reduces the need for admission to the hospital per 100,000 population younger than age 75 years. 2014; http://www.qualitymeasures.ahrq.gov/content.aspx?id=47604. Accessed 4/1/2014.

112. Billings J. ACS Algorithm. 2014; http://wagner.nyu.edu/faculty/billings/acs-algorithm. Accessed Apr. 1, 2014.

113. Kumar GS, Klein R. Effectiveness of case management strategies in reducing emergency department visits in frequent user patient populations: a systematic review. *J Emerg Med.* 2012;44(3):717-729.

114. Rosen A, Teesson M. Does case management work? The evidence and the abuse of evidence-based medicine*. *Australian and New Zealand Journal of Psychiatry.* 2001;35(6):731-746.

115. Saleh SS, Vaughn T, Hall J, Levey S, Fuortes L, Uden-Holmen T. The effect of case management in substance abuse on health services use. *Care Manag J.* 2003;4(2):82-87.

116. Ash AS, Ellis RP, Pope GC, et al. Using diagnoses to describe populations and predict costs. *Health care financing review.* Spring 2000;21(3):7-28.

117. Verisk Health. DxCG Medical Classification System White Paper. 2011. http://www.veriskhealth.com/resources/library?display=whitepaper#whitepaper (registration required). Accessed Feb. 5, 2014.

118. Zhao Y, Ellis RP, Ash AS, et al. Measuring population health risks using inpatient diagnoses and outpatient pharmacy data. *Health services research.* 2001;36(6 Pt 2):180-193.

119. Billings J. NYU ED Algorithm. 2014; http://wagner.nyu.edu/faculty/billings/nyued-background. Accessed Apr. 1, 2014.

120. Basu J, Friedman B, Burstin H. Primary care, HMO enrollment, and hospitalization for ambulatory care sensitive conditions: a new approach. *Medical care.* 2002;40(12):1260-1269.

121. Ash A, Schwartz M. R^2: A useful measure of model performance when predicting a dichotomous outcome. *Statist Med.* 1999;18:375-384.

122. Burgess J, Zhou C, Fishman P, Wang L, Sloan K. Analyzing Health Care Cost Data in the Presence of Heterogeneous Patients: Wouldn't Three Pieces Be Better than One? *iHEA 2007 6th World Congress: Explorations in Health Economics*. Copenhagen, Denmark: Social Science Research Network; 2007.

123. Duncan IG. *Healthcare Risk Adjustment and Predictive Modeling*: Actex Publications; 2011.

124. Chen J, Ellis RP, Toro KH, Ash AS. Medicare Risk Adjustment Models: DxCG vs. CMS-HCC. *Randall P. Ellis Blog.* 2014. http://blogs.bu.edu/ellisrp/files/2014/01/ChenEllisToroAsh_Medicare-Risk-Adjustment-Models_20131231_2.pdf. Accessed April 18, 2014.

125. Durand A-C, Gentile S, Devictor B, et al. ED patients: how nonurgent are they? Systematic review of the emergency medicine literature. *Am J Emerg Med.* 2011;29(3):333-345.

126. Begley CE, Vojvodic RW, Seo M, Burau K. Emergency room use and access to primary care: evidence from Houston, Texas. *Journal of health care for the poor and underserved.* Aug 2006;17(3):610-624.

127. Lurie N, Gresenz CR, Blanchard J, et al. Assessing Health and Health Care in the District of Columbia: Phase 2 Report. 2008; http://www.rand.org/pubs/working_papers/WR579.html. Accessed Feb. 5, 2014.

128. Bohland J. Neighborhood variations in the use of hospital emergency rooms for primary care. *Social Science & Medicine.* 1984;19(11):1217-1226.

129. National Quality Forum. National Discussion on Risk Adjustment. 2014; http://www.qualityforum.org/National_Discussion_on_Risk_Adjustment.aspx. Accessed Apr. 25, 2014.

130. US Census Bureau. American Community Survey. 2014; https://www.census.gov/acs/www/. Accessed Apr. 25, 2014.

131. Harrell F, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine.* 1996;15:361-387.

132. Agency for Healthcare Research & Quality. Priority Conditions. 2001; http://meps.ahrq.gov/mepsweb/data_stats/MEPS_topics.jsp?topicid=41Z-1. Accessed May 22, 2012.

133. Turner KS. What are regular expressions and how can I use them in Stata? 2005; http://www.stata.com/support/faqs/data-management/regular-expressions/. Accessed Apr. 25, 2014.

134. Geronimus AT, Bound J, Neidert LJ. On the validity of using census geocode characteristics to proxy individual socioeconomic characteristics. *J Am Stat Assoc.* 1996;91(434):529-537.

135. Krieger N, Chen JT, Waterman PD, Rehkopf DH, Subramanian SV. Race-ethnicity, gender, and monitoring socioeconomic gradients in health: A comparison of area-based socioeconomic measures--The Public Health Disparities Geocoding Project. *Am J Public Health.* 2003;93(10):1655-1671.

136. Krieger N, Chen JT, Waterman PD, Soobader MJ, Subramanian SV, Carson R. Choosing area based socioeconomic measures to monitor social inequalities in low birth weight and childhood lead poisoning: The Public Health Disparities Geocoding Project (US). *J Epidemiol Community Health.* 2003;57:186-199.

137. Diez Roux AV, Kiefe CI, Jacobs DR, et al. Area Characteristics and Individual-Level Socioeconomic Position Indicators in Three Population-Based Epidemiologic Studies. *Ann Epidemiol.* 2001;11(6):395-405.

138. US Census Bureau. When to use 1-year, 3-year, or 5-year estimates. 2013; https://www.census.gov/acs/www/guidance_for_data_users/estimates/. Accessed Dec. 7, 2013.

139. Bazargan M, Bazargan S, Baker RS. Emergency department utilization, hospital admissions, and physician visits among elderly African American persons. *Gerontologist.* 1998;38(1):25-36.

140. Doty MM, Holmgren AL. Health care disconnect: gaps in coverage and care for minority adults. Findings from the Commonwealth Fund Biennial Health Insurance Survey (2005). 2006; 1-12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16892521. Accessed 5/12/12.

141. Falik M, Needleman J, Wells BL, Korb J. Ambulatory care sensitive hospitalizations and emergency visits: experiences of Medicaid patients using federally qualified health centers. *Med Care.* 2001;39(6):551-561.

142. Hong R, Baumann BM, Boudreaux ED. The emergency department for routine healthcare: race/ethnicity, socioeconomic status, and perceptual factors. *J Emerg Med.* 2007;32(2):149-158.

143. Kirby J, Berdahl T. Indicators of Health Care Quality by Income and Insurance Status among Individuals with a Usual Source of Care, 2009. 2012. http://meps.ahrq.gov/mepsweb/data_files/publications/st367/stat367.shtml. Accessed Mar. 24, 2014.

144. Weber EJ, Showstack Ja, Hunt Ka, Colby DC, Callaham ML. Does lack of a usual source of care or health insurance increase the likelihood of an emergency department visit? Results of a national population-based study. *Ann Emerg Med.* 2005;45(1):4-12.

145. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ.* 1995-01-21 08:00:00 1995;310(6973):170.

146. Sharma V, Simon SD, Bakewell JM, Ellerbeck EF, Fox MH, Wallace DD. Factors influencing infant visits to emergency departments. *Pediatrics.* 2000;106(5):1031-1039.

147. Oster A, Bindman AB. Emergency department visits for ambulatory care sensitive conditions: insights into preventable hospitalizations. *Med Care.* 2003;41(2):198-207.

148. Bigby J, Dunn J, Goldman L, et al. Assessing the preventability of emergency hospital admissions: a method for evaluating the quality of medical care in a primary care facility. *The American journal of medicine.* 1987;83(6):1031-1036.

149. Lenhardt RO, Catrambone CD, Walter J, McDermott MF, Weiss KB. The asthma emergency department visit: treating a crisis in the midst of uncontrolled disease. *Ann Allergy Asthma Immunol.* Mar 2008;100(3):237-243.

150. Coventry JA, Weston MS, Collins PM. Emergency room encounters of pediatric patients with asthma: cost comparisons with other treatment settings. *J Ambul Care Manage.* 1996;19(2):9-21.

151. Ford JD, Trestman RL, Steinberg K, Tennen H, Allen S. Prospective association of anxiety, depressive, and addictive disorders with high utilization of primary, specialty and emergency medical care. *Soc Sci Med.* Jun 2004;58(11):2145-2148.

152. US Department of Health and Human Services. Doctors and hospitals' use of health IT more than doubles since 2012. 2013; http://www.hhs.gov/news/press/2013pres/05/20130522a.html. Accessed Jan. 18, 2014.

153. Chien AT, Williams TR, Yanagihara D, Casalino LP. Do physician organizations located in lower socioeconomic status areas score lower on pay-for-performance measures? *Journal of general internal medicine.* 2012;27(5):548-554.

154. Machlin SR. Expenses for a Hospital Emergency Room Visit, 2003. 2006; http://www.meps.ahrq.gov/mepsweb/data_files/publications/st111/stat111.pdf. Accessed Feb. 5, 2014.

155. Bodenheimer T, Pham HH. Primary care: current problems and proposed solutions. *Health Aff.* 2010;29(5):799-805.

156. Henneman PL, Lemanski M, Smithline HA, Tomaszewski A, Mayforth JA. Emergency department admissions are more profitable than non-emergency department admissions. *Ann Emerg Med.* 2009;53(2):249-255.

157. Rinderknecht AS, Ho M, Matykiewicz P, Grupp-Phelan JM. Referral to the emergency department by a primary care provider predicts severity of illness. *Pediatrics.* 2010;126(5):917-924.

158. PR Newswire Association. Judge Rules Against Medicaid ER Visit Limit. 2011; http://www.acep.org/Content.aspx?id=82576&LangType=1033. Accessed Feb. 5, 2014.

159. Kellermann AL, Weinick RM. Emergency departments, Medicaid costs, and access to primary care -- understanding the link. *NEJM.* 2012;366(23):2141-2143.

160. Reed M, Fung V, Brand R, et al. Care-seeking behavior in response to emergency department copayments. *Medical care.* 2005;43(8):810-816.

161. Selby JV, Fireman B, Swain BE. Effect of a copayment on use of the emergency department in a health maintenance organization. *NEJM.* 1996;334:635-641.

162. Mortensen K. Copayments did not reduce Medicaid enrollees' nonemergency use of emergency departments. *Health Affairs.* 2010;29(9):1643-1650.

163. National Quality Forum. Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors. 2014;

http://www.qualityforum.org/ProjectDescription.aspx?projectID=73517. Accessed Apr. 25, 2014.

APPENDIX A. VISITS FLAGGED AS UNCLASSIFIABLE BY THE NYU ED ALGORITHM

The NYU ED Algorithm was developed on a sample of approximately 6,000 ED visits to New York University hospitals in the mid-late 1990s. Although it is the only method currently available for categorizing ED utilization using ICD-9-CM diagnosis codes, some diagnoses are not classifiable according to the algorithm. These include codes that occurred to infrequently in the original sample, and new codes that have been introduced since the algorithm was originally developed. The algorithm has not been updated by the original developers since 2003. However, a version that included some code updates was developed by the Massachusetts Center for Health Information and Analysis (CHIA), with input from the original developer and assistance from an emergency medicine physician, in 2009.[50] This updated version is not publicly available, but was obtained by the authors for use in this study through personal correspondence with CHIA. The updated algorithm is available from the authors. In our data, the updated algorithm reduced the proportion of unclassifiable visits from 14.8% to 10.1%.

In **Table A-1** below, we provide a list of all the diagnosis codes in the 2010 Managed Care Network (MCN) data that were associated with 10 or more ED visits and flagged as unclassifiable by the algorithm.

Table A - 1. Diagnosis codes associated with 10 or more ED visits that were flagged as unclassifiable by the CHIA-updated version of the NYU ED Algorithm, 2010

ICD-9-CM Diagnosis Code	Description	Number of Cases	Percent of Cases
780.60	Fever, unspecified	77	0.9%
276.51	Dehydration	58	0.7%
338.19	Other acute pain	50	0.6%
V015	Contact with or exposure to rabies	37	0.4%
564.00	Constipation, unspecified	33	0.4%

ICD-9-CM Diagnosis Code	Description	Number of Cases	Percent of Cases
345.90	Epilepsy, unspecified, without mention of intractable epilepsy	20	0.2%
V58.32	Removal of sutures	20	0.2%
415.1	Pulmonary embolism and infarction	14	0.2%
888.1	Cerebral Thermography	12	0.1%
338.29	Other chronic pain	12	0.1%
788.20	Retention of urine, unspecified	12	0.1%
338.18	Other acute postoperative pain	11	0.1%
599.70	Hematuria, unspecified	11	0.1%
349.0	Reaction to spinal or lumbar puncture	10	0.1%
453.42	Acute venous embolism and thrombosis of deep vessels of distal lower extremity	10	0.1%
584.9	Acute kidney failure, unspecified	10	0.1%
726.33	Olecranon bursitis	10	0.1%

APPENDIX B: DETAILED DESCRIPTION OF MCN PROVIDER QUALITY SCORE

The purpose of this appendix is to provide further details on the provider quality score used to measure PCP quality in the MCN analysis. The provider quality score was an average of 3 z-scores (from 2009, 2010, and 2011 – the study period) that were provided by the MCN as part of their data on the characteristics of the providers and practices included in this study. The MCN calculated these scores using a method developed by the MCN as part of a dissertation research project.

The measures included were evaluated on whether they: 1) were feasible to measure in the data available; 2) were used by other pay-for-performance measurement models; 3) were publically reported (e.g., by other physician quality programs); 4) were externally vetted; 5) had been proven to be associated with provider quality (i.e., evidence-based); 6) were clinically relevant in the MCN; 7) had benchmarks available for comparison; and 8) were case-mix adjustable. Forty-four measures were evaluated against the selection criteria, and 21 were retained. **Table B-1** describes each of the 21 measures included in the score, all of which were also Healthcare Effectiveness Data and Information Set (HEDIS) measures except for the generic prescribing measure.

Table B - 1. Quality measures included in the MCN provider quality score

Measure	**Population**	**Service Received/Not Received or Test Result**
Well Child Visit	Individuals aged 0-11 years	Received the following: • Health and development history (physical and mental) • Physical exam • Health education/ anticipatory guidance

Measure	Population	Service Received/Not Received or Test Result
Well Adolescent Visit	Individuals aged 12-21 years	Received the following: • Health and development history (physical and mental) • Physical exam • Health education/ anticipatory guidance
Pharyngitis	Children who had an outpatient visit or ED encounter with only a diagnosis of pharyngitis Excludes: • Encounters with > 1 diagnosis • Children with a history of antibiotic Rx within 30 days of encounter	Were dispensed an antibiotic and also received a Group A streptococcus test 3 days before or 3 days after the prescription.
Upper Respiratory Infection	Children 3 months to 18 years who were given a diagnosis of upper respiratory infection (URI) Excludes: • Encounters with > 1 diagnosis • Children with a history of antibiotic Rx within 30 days of encounter	Were NOT dispensed an antibiotic prescription within 3 days of the URI diagnosis
Chlamydia Ages 16-20	Women identified as presumed sexually active by pharmacy Rx data, or claims data indicating potential sexual activity Excludes: • Women who had a pregnancy test followed within 7 days by either a prescription for Accutane (isotretinoin) or an X-ray.	Received a screening test for chlamydia yearly
Chlamydia Ages 21-24	Women identified as presumed sexually active by pharmacy Rx data, or claims data indicating potential sexual activity Excludes: • Women who had a pregnancy test followed within 7 days by either a prescription for Accutane (isotretinoin) or an X-ray.	Received a screening test for chlamydia yearly
Diuretics	Members 18 years of age and older who received at least 180 treatment days of ambulatory medication therapy with a diuretic	Received at least one therapeutic monitoring event for the therapeutic agent

Measure	Population	Service Received/Not Received or Test Result
ACE and ARBs	Members 18 years of age and older who received at least 180 treatment days of ambulatory medication therapy with an ACE/ARB	Received at least one therapeutic monitoring event for the therapeutic agent
CAD LDL Control	Members 18-75 years of age who were discharged alive for AMI, CABG or PCI in the year prior to the MY, or who had a diagnosis of IVD during the MY and the year prior to the MY	Received LDL-C screening
CAD LDL Testing	Members 18-75 years of age who were discharged alive for AMI, CABG or PCI in the year prior to the MY, or who had a diagnosis of IVD during the MY and the year prior to the MY	Had LDL-C <100mg/dL
Diabetes Nephropathy	Members 18-75 years of age with diabetes (type 1 or 2)	Received medical attention for nephropathy
Diabetes LDL Control	Members 18-75 years of age with diabetes (type 1 or 2)	Had LDL-C <100mg/dL
Diabetes LDL Testing	Members 18-75 years of age with diabetes (type 1 or 2)	Received LDL-C screening
Diabetes A1C < 7 (Good)	Members 18-75 years of age with diabetes (type 1 or 2)	Had HbA1C < 7
Diabetes A1C >9 (Poor)	Members 18-75 years of age with diabetes (type 1 or 2)	Had HbA1C > 9
Diabetes Testing (2/yr)	Members 18-75 years of age with diabetes (type 1 or 2)	Received HbA1c testing
Cervical Cancer Screening	Women 21-64 years of age	Received a PAP smear
Breast Cancer Screening	Women 50-74 years of age	Received a mammogram
Colorectal Cancer Screening	Adults 50–75 years of age	Received a colon cancer screening test (fecal occult blood test, sigmoidoscopy, or colonoscopy)
Lower Back Pain	Members with a primary diagnosis of low back pain	Did not receive an imaging study (plain X-ray, MRI, CT scan) within 28 days of diagnosis
Generic Prescribing	All patients	Percentage of all prescriptions that a physician writes for generic drugs

APPENDIX C. TOP-CODED VARIABLES

In our analysis of managed care network (MCN) data (discussed in Chapters II and IV), we top-coded all continuous variables—both the dependent and predictor variables—in order to reduce the effects of outliers on our statistical models. In **Table C-1** below, we provide details on the variables we top-coded and the values of the 99.5th percentiles.

Table C - 1. Top-coded variables and values, MCN data, 2009-11

Description	Number of observations	Value of 99.5th Percentile
Adult body mass index	10,954	44.3
Driving distance from home to the closest ED (miles)	107,449	14.5
Number of ED visits in the base year	107,449	3.0
Mean diastolic blood pressure reading	36,102	100.0
Mean systolic blood pressure reading	36,102	161.1
Median census tract home value	107,437	607,600
Median census tract family income ($)	107,442	152,375
Median census tract monthly rent ($)	105,944	1,774
Prospective DxCG morbidity score	107,448	15.3
Concurrent DxCG morbidity score	107,448	24.9
Normalized prospective DxCG morbidity score	107,448	10.9
Normalized concurrent DxCG morbidity score	107,448	17.2
Percent of households below poverty level in census tract	107,442	41.0
Driving distance from home to PCP office (miles)	107,305	52.9
Mean travel time to work in census tract (minutes)	107,449	35.9
Number of ED visits in the prediction year	107,449	3.0
Number of ED visits in the prediction year, development sample only	53,112	3.0
Number of PCS ED visits in the prediction year	107,449	1.87
Number of PCS ED visits in the prediction year, development sample only	53,112	1.88
Mean PCP quality score (2009-11)	107,293	2.96

APPENDIX D. CONCORDANCE BETWEEN PROBLEM LISTS AND CLAIMS

In this appendix, we provide further details on the methods used to create condition indicators using electronic medical record (EMR) data and the results of an analysis of the concordance between the conditions identified in the EMR vs. the medical claims condition categories (CCs) generated by the DxCG software.

In **Table D-1**, we provide details on the algorithms used to define each condition from the problem lists and physical measurements. These algorithms were developed by a PhD candidate with input from an MD, an MD/PhD student, and a PhD researcher. The problem lists obtained from Allscripts included both a diagnosis field (containing an ICD-9-CM code) and a description field. Since many records were missing either the diagnosis code or description, we used a two-stage algorithm to identify cases from either field.

We identified the initial diagnosis codes to be matched using the ICD-9-CM code manual at http://www.icd9data.com/. We developed the description search terms using a software programming technique known as "regular expression matching". Regular expressions are strings of letters and special characters known as operators, which can be used to match substrings and portions of text in a text-based variable (in this case, the description field in a problem list). For example, to identify individuals with arthritis, we first flagged every record that contained any ICD-9-CM code within the range 714.00 to 716.99. We then searched the description fields for the word "Arthritis", which could be either upper- or lower-case as indicated by the brackets around an upper- and lower-case letter A in the regular expression in **Table D-1**. We then flagged and removed any records that contained the terms allergic, bacterial, bowel, infect* (using a wildcard character, *, to match any word that started with “infect”; such as infectious), reactive, and septic, to exclude other forms of arthritis. We then scanned the remaining records to ensure that the preliminary set of arthritis cases was accurate and that no other description field terms should be included or excluded.

In addition, we generated a list of enrollees who had a claim for one of the 10 conditions but had not been flagged by the problem list algorithm for that condition. We randomly selected 25 cases per condition for detailed problem list review. For each set of cases, two researchers independently reviewed all of the problem list entries for each enrollee to determine whether any diagnosis codes or description field terms should be added to the algorithm.

Table D - 1. Algorithms for creating condition indicators from problem list entries in the electronic medical record

Condition	Physical Measure	Problem List Entries			
		ICD-9 Code Matches	ICD-9 Code Does Not Match	Description Field Includes	Description Field Does Not Include
Arthritis	N/A	714-716.99		[Aa]rthritis	[Aa]llergic [Bb]acterial [Bb]owel [Ii]nfect* [Rr]eactive [Ss]eptic
Asthma	N/A	493.00-493.99		[Aa]sthma*	
Cancer	N/A	140.00-209.39		[Cc]ancer* [Ll]eukemia [Ll]ymphoma [Mm]alignant [Nn]eoplasm [Cc]arcinoma	[Hh]istory

Condition	Physical Measure	Problem List Entries: ICD-9 Code Matches	ICD-9 Code Does Not Match	Description Field Includes	Description Field Does Not Include
Chronic obstructive pulmonary disease	N/A	491.00-492.8, 496*		[Cc]hronic obstructive [Ee]mphysema [Cc]hronic [Bb]ronchitis	
Congestive heart failure	N/A	428.0-428.9		[Cc]ongestive heart	
Depression	N/A	296.2, 296.3, 311*		[Dd]epression	[Ff]racture
Diabetes	N/A	250.0-250.9		[Dd]iabet*	[Gg]estational [Ii]nsipidus [Pp]rediabetes [Ss]creening [Pp]regnancy
Hypertension	Systolic >= 140 & diastolic >=90	401*	V17.49, V81.1, 403.01, 403.10, 403.11, 403.90, 405.11, 405.91, 459.30-459.33, 642.12, 642.70, 760.0	[Hh]ypertensi*	[Oo]cular [Pp]ulmonary [Ii]ntracranial [Pp]ortal [Pp]regnancy
Overweight	(Adults only) BMI >25	278.0*		[Oo]verweight [Oo]bes*	[Ff]eel
Tobacco use	N/A	305.1, V15.82	V65.43	[Ss]moke* [Tt]obacco [Nn]icotine	[Ff]ormer [Ss]econdhand [Hh]ous* [Nn]ever [Hh]istory [Uu]nknown [Ss]topping [Rr]emission

Asterisks are "wild card" operators. For example, infect would match infection, infectious, etc. Letters in [brackets] designate letters that were allowed to be either upper- or lower-case. For example, [Tt]obacco would match either tobacco or Tobacco.

Concordance for each of the 10 conditions was calculated in 3 ways: 1) the percentage of persons with a claim for the condition who also had the condition recorded in the problem list; 2) the percentage of persons with a problem-list entry indicating a condition who also had a claim for that condition; and 3) the percentage of persons with a measurement indicating overweight or hypertension (HTN) among those who had a claim for that condition. For adults only, BMI ≥ 25 was defined as overweight (BMI was not calculated for those under age 18). Hypertension was defined as systolic ≥ 140 mmHg and diastolic ≥ 90 mmHg (Joint National Committee 8 criteria).

In **Table D-2**, we provide details on the prevalence of each of the conditions as identified in the claims data from the base year (CC) and the problem list (PL) entries from either year. In most cases, the PLs identified substantially more cases than the claims. This is not surprising, since the claims represent treatments recorded over a 12-month period, whereas the PLs represent self-reports from patients collected over a longer period (up to 3 years).

Table D - 2. Prevalence of selected conditions in problem lists and claims

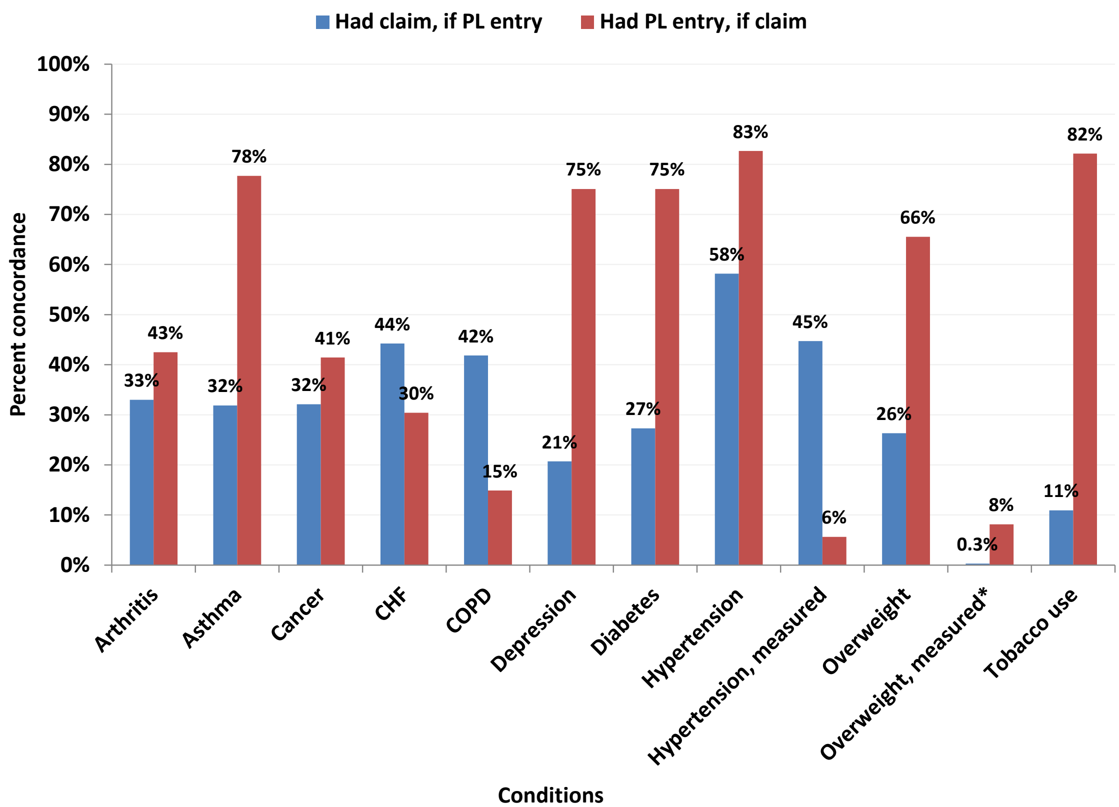
	Prevalence			
Condition	**Per Problem Lists**		**Per Claims**	
	N	**%**	**N**	**%**
Arthritis	3,824	7.2%	2,994	5.6%
Asthma	6,481	12.2%	2,659	5.0%
Cancer	11,225	21.1%	8,698	16.4%
CHF	253	0.5%	368	0.7%
COPD	215	0.4%	604	1.1%
Depression	7,093	13.4%	3,106	5.8%
Diabetes	8,541	16.1%	3,106	5.8%
Hypertension	12,655	23.8%	8,909	16.8%
Hypertension, measured	1,127	2.1%	8,909	16.8%
Overweight	5,812	10.9%	2,336	4.4%
Overweight, measured	1,949	3.7%	2,336	4.4%
Tobacco use	8,827	16.6%	1,178	2.2%

CHF: congestive heart failure; COPD: chronic obstructive pulmonary disease

As shown in **Figure D-1**, when we tabulated the proportion of persons with a claim if they had a PL entry for the condition (blue bars), we found the highest percentage agreement for hypertension, at 58%, and the lowest for measured overweight, at 0.3%. In other words, among those with hypertension in their problem list, 58% had a claim in the base year for hypertension. On the other hand, among adults with a measured BMI of 25 or more recorded in their electronic medical record, only 0.3% had a claim in the base year for overweight/obesity.

Looking at the proportion of persons with a PL entry if they had a claim (red bars), we found higher levels of agreement overall (mean=59%), with the highest, again, for hypertension (83%) and the lowest for measured overweight (8%). In other words, 8% of adults with a claim for overweight had a measured BMI of 25 or more.

Figure D - 1. Concordance between claims and problem lists for 10 priority conditions



*BMI measured in adults only

PL: problem list; CHF: congestive heart failure; COPD: chronic obstructive pulmonary disease

The two conditions for which we had physical measurements derived from the EMR, overweight and hypertension, also allowed us to calculate the proportion of individuals who had a claim for a condition, but whose measured values did not indicate the condition. Obviously, this would be expected among individuals with hypertension taking medication to keep blood pressure under control. In the case of hypertension, 20% of those with a claim in the base year had mean blood pressure values below the Joint National Committee 8 criteria (systolic ≥ 140 mmHg and diastolic ≥ 90 mmHg), and 7% of those with an overweight claim in the base year had BMIs below 25 in the EMR.

APPENDIX E. CENSUS TRACT VARIABLES, SOURCE FILES, AND EXPLORATORY FACTOR ANALYSIS

In **Table E-1** below, we list the 17 variables obtained from the US Census Bureau, with the file and claim identifiers from the Census. We also show the mean estimates for each variable in our development and validation samples and in the entire US.

Table E - 1. Census variable definitions, files, variable names, and mean values

Description	Census File ID	Census Variable Name(s)	Develop-ment Mean	Validation Mean	Entire US Mean
Median age (yrs.)	DP05	HC01_VC21	40	40	37
% female	DP05	HC03_VC05	51%	51%	51%
% Black	DP05	HC03_VC44	3%	3%	13%
% Asian	DP05	HC03_VC50	4%	4%	5%
% Hispanic	DP05	HC03_VC82	6%	6%	16%
% students (defined as % of the population aged 18+ enrolled in college or graduate school)	S1401	HC01_EST_VC22	9%	9%	10%
% unemployed (among those age 16+)	S2301	HC04_EST_VC01	7%	7%	9%
% high school graduate or higher (ages 25+)	DP02	HC03_VC93	91%	91%	85%
% foreign-born	DP02	HC03_VC134	10%	10%	13%
% speak English less than very well	DP02	HC03_VC170	6%	6%	9%
Median household income	DP03	HC01_VC85	$78,704	$78,946	$52,762
% in poverty (defined as <200% FPT)	DP03	HC03_VC166	7%	7%	14%
% owner-occupied housing	DP04	HC03_VC63	74%	74%	66%
Median monthly rent	DP04	HC01_VC185	$965	$969	$871
Median house value	DP04	HC01_VC125	$305,063	$305,888	$186,200
% vacant housing units	DP04	HC03_VC05	6%	6%	12%
Mean travel time to work (minutes)	DP03	HC01_VC36	28	28	25

FPT: Federal poverty threshold ($22,050 for a family of four in 2010). Estimates are from the American Community Survey 2011 5-year estimate files, downloaded from http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml.

Factor Analysis

In order to explore the associations between Census-tract characteristics, we performed an iterated principal axes analysis, retaining 3 factors, followed by varimax rotation. The three factors explained 100% of the total variance observed and were poorly correlated with each other. The table below lists the variables, their associations with each of the 3 factors (factor loadings), and their uniqueness.

Table E - 2. Factor loading and uniqueness for Census tract variables

Variable	Factor 1	Factor 2	Factor 3	Uniqueness
% female				**0.93**
Median age	*-0.55*			**0.62**
% Black	0.54	0.53		0.42
% Asian			0.69	0.43
% Hispanic	0.84			0.23
% high school graduates	*-0.81*			0.24
% unemployed	0.60			0.56
% students		0.35		**0.86**
Mean travel time to work (minutes)	*-0.31*	*-0.61*		0.53
Median household income	*-0.69*	*-0.44*	0.50	0.09
% foreign-born	0.45	0.68	0.56	0.02
% speak English less than very well	0.70	0.53		0.18
% vacant housing units	0.53			**0.70**
% owner-occupied housing	*-0.81*	*-0.37*		0.21
Median house value	*-0.33*	*-0.36*	0.65	0.34
Median monthly rent			0.58	**0.65**
% in poverty	0.88			0.18

Negative factor loadings are shown in *red italics*; uniqueness >.6 shown in **bold text**. Blank cells indicate factor loadings <.3.

These factor loadings suggest that median age, percent of high school graduates, mean travel time to work, median household income, % owner-occupied housing, and median house value are all negatively correlated with Factor 1, whereas % black, % Hispanic, % unemployed, % foreign-born, % who speak English less than very well, % of vacant housing units, and % in poverty are all positively correlated with Factor 1 in a particular Census tract.