uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Jenna Wong

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Epidemiology)

GRADE / DEGREE

Department of Epidemiology and Community Medicine

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Derivation and Validation of a Time-Dependent Risk Prediction Model for In-Hospital Mortality

TITRE DE LA THÈSE / TITLE OF THESIS

Carl van Walraven

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

Alan Forster

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

Yue Chen

Bernard Choi

Gary W. Slater

Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

DERIVATION AND VALIDATION OF A TIME-DEPENDENT RISK PREDICTION MODEL FOR IN-HOSPITAL MORTALITY

JENNA CHUN-LAY WONG

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies in partial fulfillment of the requirements for the M.Sc. degree in Epidemiology

Epidemiology and Community Medicine

Faculty of Medicine

University of Ottawa

Your file Votre référence
ISBN: 978-0-494-74162-7
Our file Notre référence
ISBN: 978-0-494-74162-7

Abstract

Accurate risk prediction models for in-hospital mortality are important for unbiased comparisons of hospital performance (by producing risk-adjusted mortality rates) and improved patient outcomes (by identifying high-risk patients in need of special medical attention). No previous risk prediction models have properly used post-admission information to predict risk of death in-hospital.

In this study, we used administrative and laboratory data to derive and internally validate a Cox regression model (the "*Escobar*$^{+}$" model) that predicts the risk of in-hospital death at any point during the admission. The model had excellent discrimination (*c*-statistic 0.895, 95% confidence interval [CI] 0.889-0.902) and calibration.

The *Escobar*$^{+}$ model is a powerful risk-adjustment methodology that can be used in studies where the start of observation occurs post-admission. The model could also improve the quality and timeliness of patient care by providing health care workers with highly specific and accurate estimates of in-hospital death risk during the patient's stay.

Acknowledgements

First and foremost, I would like to thank God for giving me the opportunity to undertake this challenge and my family for their love and support during the completion of this project. Mom and dad – thanks for always encouraging and believing in me. This thesis adds to my list of accomplishments that would not have been possible without your support.

Thank you to my thesis supervisors: Drs. Carl van Walraven, Alan Forster, and Monica Taljaard. In addition to being fantastic people, you are all such accomplished and knowledgeable researchers. I am very fortunate to have had the opportunity to work with you. Special thanks go out to my main supervisor, Carl – first, for choosing a research project for me that was both interesting and challenging; second, for always being available to answer my many questions; and third, for pushing me to get my thesis done by asking me *daily*, "Is it done yet?", to which I can finally say, "Yes!"

I would like to acknowledge the help of various members from The Ottawa Hospital Data Warehouse team: Natalie – for cutting my study dataset (x100), helping me become good friends with SAS, and making time (often after work hours) to chat about my thesis dilemmas; Deanna and Josée – for providing me with support and advice on data quality issues; and Steve – for answering my statistical questions and lending me a chapter from his book (which was key in guiding me through the model-building step of this project). I would also like to thank Phil Soublière (from Decision Support) and Rina Marcantonio (from Health Records) for answering my questions about coding practices at The Ottawa Hospital.

I would also like to thank Dr. Yaning Wang from the U.S. Food and Drug Administration, who I contacted after coming across one of his postings on an online statistical forum. Dr. Wang graciously sent me his SAS code, which helped me understand

how to obtain estimates of the baseline hazard function from a time-dependent Cox regression model in SAS (this was key in allowing me to assess the calibration of the risk prediction model).

Finally, I would like to acknowledge the Canadian Institutes for Health Research for funding this research project.

Table of Contents

List of Tables

List of Figures

1. Introduction

1.1 Background

Between April 2004 and March 2007, over 250 000 patients died in Canadian hospitals (excluding Quebec) (1). For some of these patients, death was inevitable due to the severity of their underlying illness. For other patients, however, death may have been prevented (or delayed) by providing timelier, higher quality, or safer (i.e. error-free) medical care. These potentially preventable in-hospital deaths have driven governments, hospital administrators, and health care professionals to focus on in-hospital death as an important outcome measure of health care quality. Death is an advantageous outcome measure because it is a definite and objective event that is accurately and completely measured in hospital records (1). In recent years, in-patient mortality has become one of the most commonly used performance indicators for hospitals and other health care institutions (2).

Statistical models that predict the risk of in-hospital death can be used in several ways to improve patient care and hospital outcomes. First, models that use information on patients' baseline characteristics and physiological status to predict mortality risk can be used to risk-adjust outcomes. Such models account for baseline differences in individual patients to allow for fair comparisons between health care providers and institutions (2-4). Unbiased comparisons can help improve patient safety and care by reliably identifying institutions and providers whose mortality rates deviate significantly from expected rates. The hospital-standardized morality ratio (HSMR), currently used in the United Kingdom, the United States, and Canada, is one such performance measure that employs a statistical model to adjust for differences in patient factors (1).

Second, risk prediction models can be used to identify risk factors for in-hospital death and quantify their effect. By taking the antilog of a covariate's parameter estimate

(i.e. e^{β}), one can describe the nature (i.e. positive or negative) and estimate the magnitude of the covariate's effect on the risk of in-hospital death. Modifiable factors with a strong, independent association with in-hospital death could be the target of interventions to improve hospital outcomes.

Third, physicians and other health care professionals can use these risk prediction models to provide timelier and higher quality patient care to improve outcomes. By producing a quantitative risk of death for each patient, these models can identify high-risk patients and prompt physicians to respond quickly with appropriate medical attention and treatments.

1.2 Measures of predictive accuracy

A model's predictive accuracy is comprised of two components: *discrimination* and *calibration*. Discrimination refers to the model's ability to distinguish between subjects who do and do not experience the event of interest. A model with good discrimination consistently predicts a higher risk for subjects who experience the event. Calibration refers to the accuracy of the model's predictions. A model with good calibration is one that achieves close agreement between the number of predicted and observed events. When assessing the predictive accuracy of a model, discrimination takes precedence because a model with poor calibration but excellent discrimination can be calibrated to improve its accuracy. A model with poor discrimination, however, cannot be adjusted to improve its accuracy (5).

Discrimination is often measured using the concept of concordance. Concordance describes how often the model produces "correct" predictions (i.e. a higher probability for subjects who experience the event). Concordance is quantified by the *c*-statistic, which represents the proportion of all possible data pairs in which the model-based probability is higher in the subject who had the event (6). Only "informative" data pairs (i.e. pairs in

which sufficient information is available to determine if the model predicts correctly) are considered in the calculation. The exact definition of a correct prediction differs depending on the type of model being considered (i.e. logistic or Cox regression).

Calibration is often measured by dividing subjects into groups (usually deciles) based on predicted risk and comparing the number of observed versus expected events in each group. This method of measuring calibration reveals how accurately the model predicts in different risk groups, and can identify groups in which the model's predictions are particularly inaccurate.

1.3 Review of literature

I searched the scientific literature to identify all generic (i.e. disease non-specific) risk prediction models for in-hospital mortality published to date, and to assess their quality and predictive performance.

1.3.1 Search strategy

I performed the literature search on 4 July 2009 using the Ovid MEDLINE (R) database (1950 to July Week 1 2009) and the following search strategy: ("in-hospital mortality" *or* "inpatient mortality") + ("risk prediction model" *or* "risk adjustment"). I included studies in the literature review if they were written in the English, specified in-patient mortality as one of the study outcomes, and presented a multivariable risk prediction model. Because I was looking for models that could be applied to a generic patient population, I excluded studies where the risk prediction model was derived on a disease- or condition-specific patient population.

1.3.2 Search results

The search strategy retrieved 970 articles. After reviewing the article abstracts, I excluded most studies because they included patient populations that were disease- or

condition-specific. For example, many studies included only patients with certain health conditions such as: cardiovascular disease, renal failure, and gastrointestinal or lung diseases; trauma patients; intensive care unit (ICU) patients; or patients admitted to hospital for medical procedures such as coronary artery bypass grafting, abdominal aortic aneurysm repair, or percutaneous coronary intervention.

I identified seven studies (7-13) that presented a risk prediction model derived on a disease non-specific patient population. I excluded one of these studies (14) because it presented a univariable model, leaving six studies for inclusion in the review (Table 1.1). The six studies included patient populations from the United States, Japan, England, Canada, and Brazil. The study populations ranged in size from 3 733 to 259 699 patients. All studies obtained patient data from administrative databases or digitized information from a data warehouse. Only one study (10) made use of administrative *and* clinical data, which is advantageous because clinical databases contain more detailed information about disease severity than administrative databases alone (15).

All models were derived using logistic regression methods and included "baseline measurable" variables (i.e. variables whose values could be determined at the time of admission). For example, all models included patient age, and the majority of models included gender, admission urgency, admission diagnosis, and an indicator of patient comorbidity (such as the Charlson Index). Two models (9;11) included "baseline immeasurable" variables (i.e. variables whose values could not be determined at admission), such as length of stay, blood transfusion, and admission to the intensive care unit (ICU). In these models, the baseline immeasurable variables were not analyzed as time-dependent covariates. As a result, these models are invalid because time-dependent bias can occur when a baseline immeasurable variable (whose status may be associated with the outcome) is treated as a time-independent covariate rather than a time-dependent covariate (16).

Table 1.1 Details of risk prediction models identified by the literature search

Author and study citation	Population	N	Year	Number of hospitals	Model Covariates*	Variable Type(s)	ICD-9 or ICD-10	Statistical method	Internal/External Validation	Data Source	Discrimination/Calibration
Bottle and Aylin (Health Serv Res 2008 February 43(1, Part I) 10-31)	England	Not specified	1996-2005	All National Health Service (NHS) hospitals in England	- Age - Gender - Admission urgency - Socioeconomic status (quintile of area-level deprivation score) - Primary diagnosis**	Pre-admission	ICD-10	Logistic regression	NONE	Administrative	- *c*-statistic >0 7 for all case-mix groups - Calibration not reported
Miyata *et al* (BMC Health Services Research 2008 8 229)	Japan	224 207	1 July 2002-31 October 2003	82	- Age - Gender - Major diagnostic category (9 categories) - Admission urgency - Use of ambulance at admission (Y/N) - **Length of hospital stay** - Charlson comorbidity index score	Pre- and post-admission	ICD-10	Logistic regression	Internal validation (split-sample)	Administrative	- *c*-statistic 0 869 (95% CI 0 860-0 879) - Close agreement between predicted and observed mortality rates across all deciles of patient risk
Escobar *et al* (Med Care 2008 Mar,46(3) 232-239)	Northern California US	259 699	January 2002-June 2005	17	- Age - Gender - Admission urgency and service (emergent surgical, emergent non-surgical, elective surgical elective non-surgical) - Admission diagnosis - *L*aboratory-based *A*cute *P*hysiology *S*core (LAPS) - *CO*morbidity *P*oint *S*core (COPS)	Pre-admission	ICD-9	Logistic regression	- Internal validation (split-sample) - External validation in a patient population from Ottawa, Canada (17)	Administrative and clinical (laboratory)	Internal validation - *c*-statistic 0 88 - *p*-value of Hosmer-Lemeshow statistic 0 66 - Excellent calibration except among patients with risk ≥60% External validation† - *c*-statistic 0 915 (95% CI 0 912-0 918) - *p*-value of Hosmer-Lemeshow statistic 0 02 - Excellent calibration among all risk strata and most risk deciles

(cont d on next page)

(Table 1 1 cont'd)

Author and study citation	Population	N	Year	Number of hospitals	Model Covariates*	Variable Type(s)	ICD-9 or ICD-10	Statistical method	Internal/External Validation	Data Source	Discrimination/Calibration
Ramianna *et al* (Rev Saude Publica 2008 Aug,42(4) 590-597)	Rio de Janeiro Brazil	3 733	January 2001-January 2003	1	- Age - **Length of hospital stay** - **Blood transfusion (Y/N)** - **ICU admission (Y/N)** - Charlson-like comorbidity index score	Pre- and post-admission	ICD-10	Logistic regression	NONE	Administrative	- *c*-statistic 0 86 (95% CI 0 83-0 89) - Calibration not reported
John-Baptiste *et al* (Journal of General Internal Medicine 2004 March 19(3) 221-228)	Toronto Canada	56 994	1 April 1993-31 December 1999	3	- Age - Charlson comorbidity score - English proficiency (proficient/limited proficiency)	Pre-admission	ICD-9	Logistic regression	NONE	Administrative	Not reported
Glenn and Jijon (Journal of Rural Health 1999 15(1) 94-107)	Tennessee US	72 295	1992	154	- Age - Gender - Diagnosis-related group - Presence of 2° diagnosis (Y/N) - Cancer as a 2° diagnosis (Y/N) - Risk of death from principal diagnosis - Risk of death from principal procedure - Risk of death from 2° diagnosis with highest risk - Number of 2° diagnoses where risk of death > risk of diagnosis-related group	Pre- and post-admission	ICD-9	Logistic regression	NONE	Administrative	Mean absolute error‡ 0 57

*Covariates in bold are baseline immeasurable variables that should have been (but were not) expressed as time-dependent covariates
**Primary diagnoses were categorized using 3- or 4-character code groupings in the model
†Using coefficients from the derivation set in the external study population
‡The average *z*-score across all hospitals where $z = (|\text{expected-observed}|) / (\sqrt{\text{expected}})$

Only two models (9;10) were internally validated using a split-sample validation approach. Model validation is important because a model's predictive accuracy may be overestimated when tested on the derivation set only (18). Only one model (10) was externally validated in a different patient population (17). External validation is an essential step before employing risk prediction models in clinical practice, since it ensures model stability in distinct populations (19). This step ensures the model does not result from some particular characteristic that is unique to the center where the model was derived.

The predictive accuracy (i.e. discrimination and calibration) of the models was not assessed in all studies. Discrimination was assessed in four studies (8-11) with the reported *c*-statistic ranging from 0.7 to 0.915 (indicating good to excellent discrimination). Calibration was assessed in only two studies (9;10), with the two models performing well across different deciles of patient risk and strata of predicted mortality.

1.3.3 Conclusions

After reviewing the scientific literature, I identified six published risk prediction models for in-hospital mortality that had been developed on a generic patient population. All of these models were derived using logistic regression methods and administrative data (one also used clinical laboratory data). Two models were at risk of time-dependent bias, two had been internally validated, and only one had been externally validated. Discrimination was assessed in four models, all of which had good discrimination. Calibration was assessed in only two models, both of which had excellent calibration.

1.4 The Escobar model

Of the risk prediction models identified by the literature review, the model derived by Escobar *et al.* (10) (the "Escobar" model) was superior for several reasons: it was derived on a large patient population; it was the only model that made use of administrative *and* clinical data; it included only baseline measurable variables (thus avoiding time-

dependent bias); it was internally and externally validated; and it was shown to have excellent discrimination and good calibration in two different patient populations.

The Escobar model was derived and validated on a cohort of 259 699 patients (comprising 409 305 hospitalizations) served by 17 Northern California Kaiser Permanente Medical Care Program (KPMCP) hospitals. The study population included patients admitted from 1 January 2002 to 31 July 2003 and 1 October 2003 to 30 June 2005. Patients were excluded from the study if they were initially transferred from a non-KPMCP hospital (interhospital transfer admissions within the KPMCP system were still included), were less than 15 years of age at the time of admission, or were admitted for childbirth. The final logistic regression model included the following predictor variables: sex, splined age^2 (expressed as three nonlinear terms), admission type (emergent surgical/non-surgical, elective surgical/non-surgical), a *L*aboratory-based *A*cute *P*hysiology *S*core (LAPS), a *CO*morbidity *P*oint *S*core (COPS), and several interaction terms (splined age^2 x LAPS, splined age^2 x COPS, LAPS x COPS). Splined age^2 was a restricted cubic spline with four knots specified at the 5th, 35th, 65th, and 95th percentiles.

The LAPS was a continuous variable that summarized the acuity of a patient's condition at admission based on pre-admission laboratory abnormalities. Specifically, the LAPS integrated results from 14 different laboratory tests (performed within the 24 hours before hospitalization) into a single score that was positively associated with risk of in-hospital death. The LAPS could range in value from a minimum of zero to a theoretical maximum of 256. Escobar *et al.* derived and validated the LAPS on a cohort of 854 139 hospitalizations that occurred in the 17 KPMCP hospitals between 1 November 1999 and 30 June 2005. The final LAPS model had a *c*-statistic of 0.73 in the validation set, indicating fairly good discrimination.

The COPS was a continuous variable that summarized the extent of a patient's chronic medical condition at admission. Using the same study population used to derive

the Escobar model, Escobar *et al.* derived the COPS using Diagnostic Cost Groups (DxCG) software – a software program used by the KPMCP hospitals to scan in-patient and outpatient data for diagnoses. To calculate the COPS, diagnoses from encounters in the previous year were assigned points if they fit into one of 41 comorbidity groups. The COPS could range in value from a minimum of zero to a theoretical maximum of 701. In the validation set, the discrimination of the final COPS model was only 0.67, indicating fair discrimination. Further details of the derivation and validation of both the LAPS and COPS have been described elsewhere (20).

Escobar *et al.* found that the risk prediction model performed best when admissions were divided into 44 *Primary Condition* groups and a separate logistic regression model was derived for each group. Admissions were assigned to a Primary Condition group based on the ICD-9 admission diagnosis code.

The risk prediction model had very good discrimination (*c*-statistic 0.88). The model predicted mortality quite accurately across most risk deciles and strata of predicted mortality. The predictions were less accurate among patients with an extremely high predicted mortality risk (≥60%).

The Escobar model was subsequently validated in a cohort of 188 724 patients admitted to The Ottawa Hospital (TOH) between January 1998 and April 2002 (17). Patients discharged after April 2002 were not included in the study because the diagnostic coding system at TOH changed from ICD-9-CM to ICD-10-CA (Escobar *et al.* used ICD-9 diagnosis codes to assign patients to a Primary Condition group). In this patient population using coefficients determined from the study data, the Escobar model had very good discrimination (*c*-statistic 0.915, 95% CI 0.912-0.918). The Escobar model also performed equally well when the COPS was replaced by other indicators of patient comorbidity, including the Elixhauser Index and the Charlson Index (17). Using these alternative comorbidity indicators is advantageous because special (DxCG) software is required to

calculate the COPS. The models that used the Elixhauser and Charlson Index had excellent discrimination, with a *c*-statistic of 0.901 (95% 0.898-0.904) and 0.894 (0.891-0.897), respectively. Both models also had very good calibration in all strata of predicted mortality and almost all risk deciles.

1.5 Improving the Escobar model with time-dependent covariates

The Escobar model used only data available at the time of admission to predict risk of death throughout the hospitalization. However, events that occur post-admission could certainly change patients' risk of death. Such events could include admission to the ICU, development of acute renal failure, or an increase in white blood cell count.

The addition of post-admission events to the Escobar model would likely improve its ability to predict risk of death throughout the hospitalization. However, because post-admission events are not measurable at the time of admission, the model is susceptible to bias if such events are analyzed as time-independent covariates (16). To avoid this bias, these events can be analyzed as time-dependent covariates in a proportional-hazards (Cox) regression model, in which the values of all time-dependent covariates are constructed as a function of time (21).

Although not commonly recognized or used in the medical literature, the analysis of baseline immeasurable variables using time-dependent covariate analysis is an important tool for accurate statistical modeling (16). To our knowledge, no risk prediction models for in-hospital mortality have included baseline immeasurable variables appropriately using time-dependent covariate analysis. This analysis can be technically challenging, but if done properly, the resulting model could predict risk of in-hospital death with greater accuracy.

1.6 Utility of a time-dependent risk prediction model for in-hospital mortality

The Escobar model was developed as a risk-adjustment methodology for comparing hospital outcomes in quality of care analyses. For this purpose, the model was time-independent and appropriately included only pre-admission variables to adjust for patients' medical condition at the time of admission. Post-admission variables were not included in the model to avoid adjusting for post-admission events that often reflect treatment decisions and quality of care provided by the hospital.

A time-dependent risk prediction model for in-hospital mortality will be able to predict risk of death at any time during hospitalization and account for the effect of both pre- and post-admission events. Such a model will be able to risk-adjust more accurately for analyses in which the observation start point occurs *during* the hospitalization, rather than right at admission. For example, in quality of care analyses, one may want to compare mortality between hospitals after an operation, or after discharge from the intensive care unit. In research studies, one may want to determine the association between post-admission events and the risk of in-hospital death (risk factor analysis). In these analyses, a time-dependent model would adjust for differences in patient risks at the analytical (post-admission) baseline more accurately than a time-independent model, thus reducing the chance of obtaining biased results.

Researchers can also use a time-dependent risk prediction model to explore different mechanisms by which post-admission risk factors may influence risk of in-hospital death (causal model evaluation). These models are ideal for examining causal pathways because they can properly adjust for post-baseline events and allow one to study the temporal relationship between risk factors, potential mediators, and outcomes.

Finally, by accurately predicting patients' risk of death at any time during the hospitalization, a time-dependent prediction model can improve patient outcomes by

identifying high-risk patients and enabling physicians to provide them with additional and appropriate care in a timely manner.

2. Study Objectives

2.1 Primary objective

The primary objective of this study was to increase the utility of the Kaiser Permanente hospital mortality risk-adjustment methodology (the "Escobar" model) by deriving and validating a Cox regression model for in-hospital mortality that included post-admission events expressed as time-dependent covariates (hereafter referred to as "*Escobar*$^{+}$").

2.2 Secondary objectives

There were two secondary objectives in this study:

i. To compare the predictive accuracy of the Escobar model fit using logistic regression versus proportional-hazards (Cox) regression.
ii. To compare the predictive accuracy of the *Escobar*$^{+}$ model versus the original Escobar model (fit using Cox regression instead of logistic regression).

3. Methods

3.1 Study design

This study was a secondary analysis, retrospective cohort study of all hospitalizations (including same-day surgeries) at The Ottawa Hospital (TOH) where the patient was discharged between 1 April 2004 and 31 March 2009. The unit of analysis in this study was the hospitalization. Using a split-sample validation approach, I derived the

risk prediction model on a randomly chosen 2/3 of admissions. I validated the model by applying the model coefficients to other 1/3 of admissions.

3.2 Study setting

This study took place at TOH, a tertiary-care teaching hospital located in Ottawa, Canada. The hospital operates within a publicly funded health care system and serves a population of approximately 1.5 million people in Ottawa and Eastern Ontario. The hospital is comprised of three main campuses (the General, Civic, and Riverside), receives all trauma patients for the region, and provides oncological care to most of the region. In-patient services are offered at two of the campuses (the General and Civic), while same-day surgeries are offered at all three campuses.

This study also included admissions at the University of Ottawa Heart Institute (UOHI). Located at the Civic campus of TOH, UOHI (a TOH partner) is Canada's largest cardiovascular health centre that provides care for more than 80 000 patients annually.

3.3 Study population

I replicated the inclusion and exclusion criteria used to derive the Escobar model (10). As in the original study, I included all in-patient and same-day surgery admissions. I excluded admissions of patients younger than 15 years of age, admissions where patients were transferred from or to other acute care hospitals (other than another TOH campus), and childbirth-related admissions. If a patient was transferred between TOH campuses, I linked the hospitalizations and considered them as a single admission.

3.4 Data source

All data in this study came from The Ottawa Hospital Data Warehouse (TOHDW). TOHDW is a repository of clinical, laboratory, and administrative data that originates from the hospital's major operational information systems. These information systems include

the patient registration system (Invision 26, Siemens Medical Systems USA Inc), electronic health record system (vOacis v.r7.2.0_20080402, Emergis Inc), and health records abstracting system (WinRecs version 2.8.6.003, MED2020 Health Care Software Inc).

The patient registration system is the hospital's primary system for creating and managing patient profiles, including data on patient demographics (i.e. age, sex, address, city of residence, etc.).

The electronic health record system contains details about patient "encounters" and services performed during each encounter. Such "encounters" include in-patient admissions, same-day surgeries, outpatient and emergency room visits, and use of outpatient laboratory or radiologic services. The patient registration system and the electronic health record system are updated by administrative, ward, or other hospital service staff during each encounter.

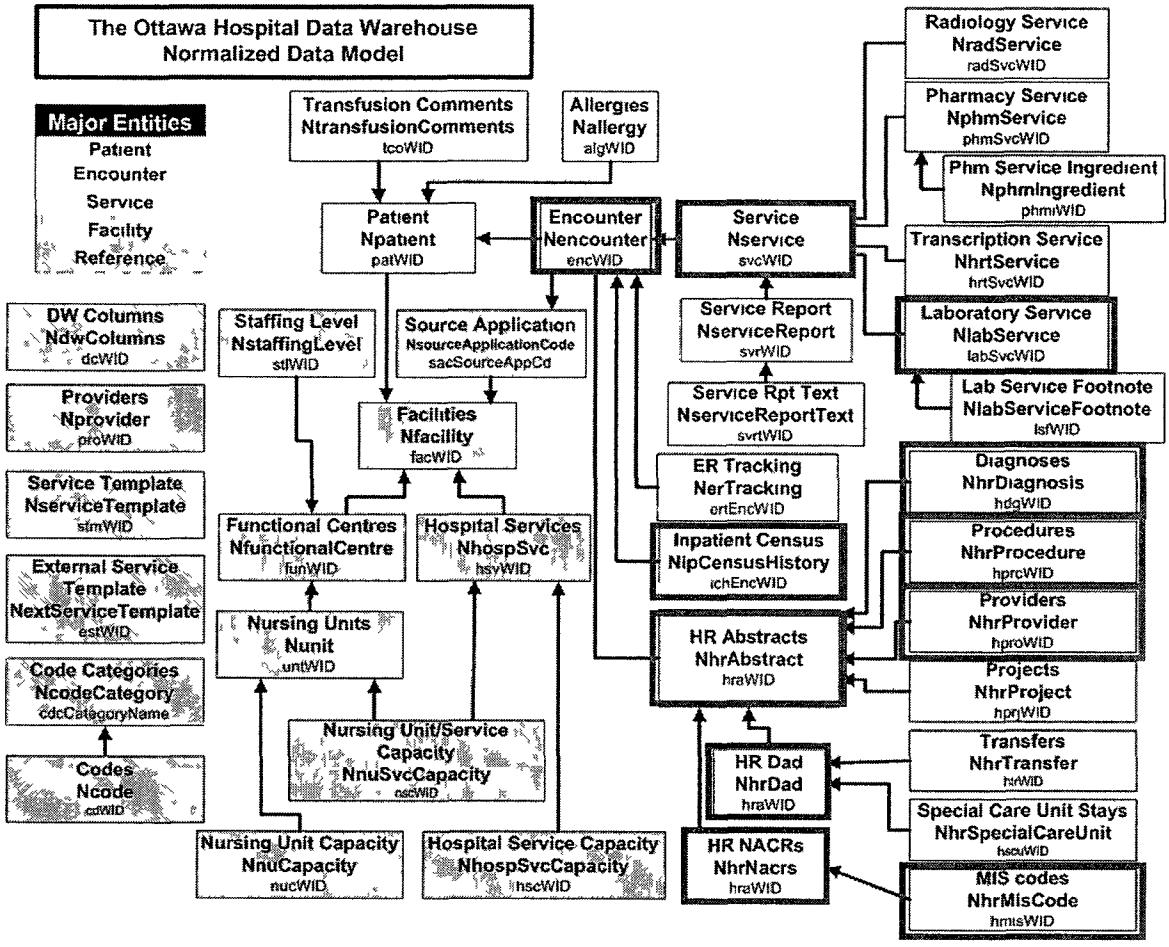
The health records abstracting system is comprised of two databases: the Discharge Abstract Database (DAD) and the National Ambulatory Care Reporting System (NACRS). The DAD contains information on in-patient and overnight same-day surgery admissions, while NACRS contains information on non-overnight same-day surgery admissions and emergency room encounters. Both systems are maintained by health records technologists who review the patient chart retrospectively (i.e. post-discharge) and record information about the admission, including health care providers involved, special care units used, procedures performed, and patient diagnoses.

In TOHDW, data are organized around five central tables, called major "entities" (Figure 3.1). These major entities branch off into numerous tables (or "sub-entities") that contain more specific information about a particular aspect of the major entity. Each data warehouse table contains at least one unique warehouse identifier (WID) that allows users to link to other tables to retrieve all data associated with an episode of care. WIDs can

refer to individual patients (patWIDs), unique encounters (encWIDs), or unique services performed during an encounter (svcWIDs).

In this study, I used data from two major entities ("Encounter" and "Service") and nine of their associated sub-entities (as indicated in Figure 3.1) to create the study cohort, determine the study outcomes, and create the model covariates (Table 3.1). Because of data availability in the "Inpatient Census" table, I could only include admissions with a discharge date later than 31 March 2004 in this study.

Figure 3.1 Map of The Ottawa Hospital Data Warehouse



The 9 data warehouse tables I used in this study are outlined in red

Table 3.1 Data Warehouse tables used in the study

Table	Entity type	Hospital source system	Earliest data availability*
Encounter	Major	Patient registration system and electronic medical record system	January 1996 (General only) August 1999 (Civic added) December 2001 (Riverside added)
Service	Major	Electronic medical record system	January 1996 (General only) July 2002 (Civic and Riverside added)
HR Dad**	Sub-entity	Health records abstracting system	April 1995 (General and Civic only)
HR NACRs	Sub-entity	Health records abstracting system	April 2002 (all campuses)
HR Abstracts	Sub-entity	Health records abstracting system	Contains further details of admissions included in "HR Dad" and "HR NACRs"
Diagnoses	Sub-entity	Health records abstracting system	Contains further details of admissions included in "HR Dad" and "HR NACRs"
Procedures	Sub-entity	Health records abstracting system	Contains further details of admissions included in "HR Dad" and "HR NACRs"
Providers	Sub-entity	Health records abstracting system	Contains further details of admissions included in "HR Dad" and "HR NACRs"
MIS codes	Sub-entity	Health records abstracting system	April 2002 (all campuses)
Inpatient Census	Sub-entity	Patient registration system	April 2004 (all campuses)
Laboratory Service	Sub-entity	Electronic medical record system	January 1996 (General only) July 2002 (Civic and Riverside added)

*Dates refer to the discharge date of admissions

3.5 Creation of the study cohort

In this study, I included all in-patient and same-day surgery admissions at TOH where the patient was greater than 15 years of age, not transferred from or discharged to a non-TOH acute care hospital, and not hospitalized for obstetrical reasons.

3.5.1 Identification of in-patient and overnight same-day surgery encounters

To identify in-patient and overnight same-day surgery encounters, I accessed the "HR Dad" and "HR Abstracts" tables to obtain the discharge date-time, abstract status,

abstract type, entry code, admission category, and discharge disposition of all encounters. I included admissions with a discharge date-time between 1 April 2004 and 31 March 2009, a completed abstract status, and an in-patient abstract type. I excluded admissions where the discharge disposition code indicated that the patient was a cadaver or stillbirth.

I classified admissions as in-patient unless they met specific criteria for overnight same-day surgery encounters: length of stay of one day, a specific nursing station code ("C3SI"), a "direct" entry code, and an "elective" admission category.

3.5.2 Identification of non-overnight same-day surgery encounters

To identify non-overnight same-day surgery encounters, I accessed the "HR NACRs", "HR Abstracts", and "MIS codes" tables to obtain the discharge date-time, abstract status, Management Information Systems (MIS) code, and discharge disposition code of all encounters. I included encounters with a discharge date-time between 1 April 2004 and 31 March 2009, a completed abstract status, and one of three MIS codes used by the hospital for same-day surgery admissions (MIS codes are functional centre account codes used by the hospital for billing purposes). I excluded admissions where the discharge disposition code indicated that the patient died upon arrival to the hospital.

3.5.3 Exclusion of invalid and incomplete records

3.5.3.1 Same-day surgery consultations

At TOH, same-day surgery patients undergo a pre-operative consultation before the day of the surgery. The consultation and actual surgery are recorded as two separate encounters in the hospital's databases. To avoid including consultations in the study, I linked each same-day surgery encounter to the "Procedures" table to determine if any procedures were performed during the encounter, and if so, the operating room number of each procedure. If a same-day surgery encounter did not have a procedure performed in a

main operating room, day-surgery unit, or night care unit, I classified the encounter as a pre-operative consultation and excluded it from the study.

3.5.3.2 Incorrectly linked encounters

In this study, I linked patient and encounter information from various Data Warehouse tables that originated from different hospital source systems (Figure 3.1). Occasionally in the Data Warehouse, encounter information from different source systems is incorrectly linked. To avoid including such encounters, I compared the admission and discharge dates from the "HR Abstract" table (which originates from the health records abstracting system) with those in the "Encounter" table (which originates from the patient registration and electronic medical record system). I excluded encounters where the admission or discharge date differed by more than 24 hours from these tables. This criterion to identify incorrectly linked encounters has been previously validated (and is presently used) by TOHDW analysts.

3.5.2.3 Incomplete records

I excluded encounters which had no information in the "Diagnosis" or "Inpatient Census" tables because these encounters were incomplete.

3.5.4 Exclusion of same-day surgery admissions that turned into in-patient encounters

In some (rare) cases, a same-day surgery patient was subsequently re-admitted to hospital as an in-patient due to surgery-related complications. To avoid including such "related" admissions as separate encounters in the study, I identified all cases in which a same-day surgery patient was admitted as an in-patient before midnight of the same day, and retained only the in-patient admission.

3.5.5 Exclusions as per the Escobar model

3.5.5.1 Age at admission

I excluded encounters where the patient was less than 15 years of age at the time of admission. To determine the patient's age at admission, I calculated the difference (in years) between the admission date and the patient's date of birth (obtained from the "HR Abstract" table). This algorithm was recommended by TOHDW analysts as the most reliable method to calculate age at admission.

3.5.5.2 Transfers to or from another acute care hospital

I excluded admissions where the patient was transferred to or from another acute care hospital *other than* another TOH campus. I excluded these admissions because information about the patient's entire hospitalization was missing from TOHDW.

I used information from the "HR Abstract" table to identify encounters where the patient was transferred to or from another acute care hospital. Such encounters had a "master number" for an acute care hospital present in the fields that indicate whether the patient was transferred to or from another health care institution (the master numbers correspond to the Ministry of Health and Long-Term Care Master Numbering System). I did not exclude admissions where the master number was for one of the TOH campuses.

3.5.5.3 Admissions for childbirth

I excluded admissions where the main patient service in the "HR Abstract" table was one of the following services: "obstetrics delivered", "obstetrics undelivered," "obstetrics aborted," or "stillbirth." The main patient service is that deemed by the health records analyst to be most responsible for the patient's care during the admission. I also excluded all admissions in the "Pregnancy" Primary Condition group (20).

3.5.6 Exclusion of admissions with a missing admission diagnosis

In the Escobar model, admissions were assigned to a Primary Condition group using the ICD-9 admission diagnosis of the hospitalization. At TOH, however, admission diagnoses were not recorded by health records analysts during the study period. I therefore consulted with Dr. van Walraven, TOHDW analysts, and Health Records staff to devise a logical algorithm to approximate the admission diagnosis (Appendix A).

The algorithm compared up to four diagnosis types: type 6 ("proxy most responsible diagnosis"), type M ("most responsible diagnosis"), type 1 ("pre-admit comorbidity"), and type 2 ("post-admit comorbidity") (Appendix A). The type 6 diagnosis, if present, was deemed the admitting diagnosis. Otherwise, if a type 6 diagnosis was not present, the type M diagnosis was used *unless* the type M diagnosis was also coded as a type 2 diagnosis. In these cases, because the type M diagnosis could not be used (since the condition arose post-admission), one of the type 1 diagnoses was used (randomly chosen among all type 1 diagnoses if more than one was present for the admission). For >99% of admissions, the type M diagnosis was deemed the admitting diagnosis (Appendix A). For 157 admissions (0.06% of candidate admissions), the algorithm could not assign an admission diagnosis because a type 6 diagnosis was not present, the type M diagnosis was also a type 2 diagnosis, and a type 1 diagnosis was not present. I excluded these admissions because I could not assign them to a Primary Condition group without an admission diagnosis.

3.5.7 Concatenation of interhospital transfer admissions

I used an algorithm validated by TOHDW analysts (Appendix B) to identify admissions linked to a previous encounter at another TOH campus ("interhospital transfer admissions"). I linked these interhospital transfer admissions to the "index" admission and considered them collectively as a single encounter.

3.6 Study outcome

In this study, the outcome was time to in-hospital death. To determine if a patient died during the admission, I obtained the discharge disposition code from the "HR Dad" table (for in-patient and overnight same-day surgery admissions) or the "HR NACRs" table (for non-overnight same-day surgery admissions), and determined whether the code was for death. I assumed that all other patients were discharged alive. Although I did not validate the death indicator in this study, this indicator has been validated in previous studies (22).

3.7 Model covariates

I created the primary condition groups and time-independent covariates as per the Escobar model.

3.7.1 Primary condition group

In the original study (10), admissions were grouped into 44 different primary condition categories based on the ICD-9 admission diagnosis code (20). During the study period, however, the hospital recorded diagnoses using the ICD-10-CA diagnostic coding system. To convert the admission diagnoses from ICD-10-CA to their ICD-9-CM equivalent, I used fiscal-year specific conversion tables, which I obtained from the Canadian Institute for Health Information. I then assigned admissions to a primary condition group using the converted ICD-9 admission diagnosis code.

For two primary condition groups ("Chest Pain" and "Gynecology"), there were no observed deaths in the derivation set. Rather than assign all admissions in these groups a predicted risk of zero, I re-assigned these admissions to another primary condition group that was biologically similar and had the most comparable mortality rate ("Other Cardiac Conditions" and "Gynecologic Cancers," respectively).

3.7.2 Time-independent covariates

3.7.2.1 Sex

I determined the sex of all patients directly from the "HR Abstract" table.

3.7.2.2 Age at admission

As previously mentioned, I determined the patient's age at admission by calculating the difference (in years) between the admission date and the patient's date of birth. In addition, as in the Escobar model, I expressed the square of the patient's age as a restricted cubic spline (i.e. a cubic spline constrained to be linear in the tails). To fit the spline, I expanded the square of the patient's age to three non-linear terms, specifying four knots (as specified in the Escobar model) at the 5^{th}, 35^{th}, 65^{th}, and 95^{th} percentiles in the derivation set (23). These knots corresponded to age 25, 52, 69, and 85, respectively.

3.7.2.3 Admission type

I expressed admission type as a categorical variable with four levels: "emergent surgical," "emergent non-surgical," "elective non-surgical," and "elective surgical." I chose elective surgical admissions as the reference in all statistical models.

To determine the urgency (emergent or elective) of in-patient admissions, I accessed the "HR Dad" table and obtained the admission category. I classified all patient admissions as elective unless its admission category was "emergency" or "urgent." I classified all same-day surgeries as elective admissions.

I determined the service type (surgical or non-surgical) of in-patient admissions using the service of the most responsible provider (i.e. physician) for the admission, which I obtained from the "NhrProvider" table. If the provider was from one of nine surgical services (general surgery, cardiac surgery, neurosurgery, oral surgery, orthopaedic surgery, plastic surgery, thoracic surgery, vascular surgery, or cardiothoracic surgery), I

classified the admission as surgical. I classified all same-day surgeries as surgical admissions.

3.7.2.4 Patient comorbidity

As previously mentioned, the Escobar model performed equally well when either the Charlson Index or Elixhauser Index was used instead of COPS to summarize patient comorbidity (17).

The Charlson Index (24) is a weighted index that defines 17 comorbidities and assigns patients a single comorbidity score based on the number and seriousness of their comorbidities. The Elixhauser Index (25) is weighted point system that summarizes 30 comorbidities as a single number rather than 30 binary variables, as in the original Elixhauser comorbidity system (26). Both indices can be used with ICD-10-CA diagnostic codes (27).

I expressed patient comorbidity using the Elixhauser Index because in a previous study (17), the discrimination of the Escobar model using the Elixhauser Index (C-statistic 0.901, 95% CI 0.898-0.904) was slightly better than the Charlson Index (0.894, 0.891-0.897). A number of studies (28-30) have also shown that the Elixhauser system is statistically slightly superior to the Charlson Index.

To identify patient diagnoses for the Elixhauser score, I used the "Diagnoses" table to obtain ICD-10 diagnosis codes and their associated diagnosis type for all index admissions as well as previous TOH admissions where the patient was in hospital during the last seven years. I chose a lookback window of seven years to ensure that all admissions had the same availability of previous diagnosis data from TOHDW. In calculating the Elixhauser score, I excluded type 2 diagnoses from index admissions because they represented conditions that arose post-admission.

3.7.2.5 LAPS at admission

I used the "Service" and "Laboratory Service" tables to obtain test results for all laboratory tests required to calculate LAPS: serum albumin; serum chloride; arterial pH, $PaCO_2$, and PaO_2; bicarbonate; total serum bilirubin; blood urea nitrogen; serum creatinine; serum glucose; serum sodium; serum troponin I; hematocrit; and total white blood cell count. I obtained only tests that were performed in the 24-hour time period before admission.

Using the scoring scheme developed by Escobar *et al.* (20), I assigned points to the 13 components of LAPS and summed the points to obtain a final score. If a test was performed multiple times within the pre-admission period, I used the test result that assigned the greatest points to LAPS (i.e. the most physiologically abnormal result).

With the exception of three lab tests (arterial pH, serum troponin I, and total white blood cell count), I assumed the patient had normal test results if the test was not actually conducted. This method was used in previous studies (10;17), and is the usual severity of illness scoring method used for handling missing data (10). For arterial pH, serum troponin I, and total white blood cell count, the points I assigned to patients who did not have the test done depended upon whether they were classified as "high risk" or "low risk" based on a simple logistic regression model. This model [called the Laboratory Score Preliminary Model (Appendix C)], included five variables and incorporated results from five commonly performed laboratory tests: blood urea nitrogen, serum creatinine, serum chloride, serum bicarbonate, and serum sodium. If a laboratory test in the preliminary model was not performed, I assigned the patient a normal test result. For low risk patients (i.e. those with an expected mortality risk <6% based on the preliminary model), I assigned zero points each for the missing arterial pH, serum troponin I, and total white blood cell count result. For high risk patients (i.e. those with an expected mortality risk ≥6%), I assigned patients 11, 9, and 23 points for their unconducted arterial pH, serum troponin I, and total white

blood cell count test, respectively. Escobar *et al.* assigned different points for high and low risk patients because an association existed between the absence of certain laboratory tests under certain conditions and an increased risk of in-hospital mortality (20).

I attempted to standardize laboratory test results using a two-step standardization process (Appendix D), since the analytical machinery and methods used at TOH to measure laboratory test results may have varied over time and between campuses. Because I found no difference in the discrimination and calibration of the Escobar model using standardized versus non-standardized lab results (Table 3.2), I used non-standardized lab results to calculate LAPS.

Table 3.2 Performance of the Escobar model in the validation set using standardized versus non-standardized lab results

	Standardized results	Non-standardized results
C-statistic (95% CI)	0.926 (0.923-0.930)	0.927 (0.924-0.931)
Hosmer-Lemeshow statistic (*p* value)	21.92 (0.005)	22.34 (0.004)

3.7.3 Time-dependent covariates

I identified post-admission factors likely associated with risk of in-hospital death through an informal consensus process with Drs. van Walraven and Forster over the course of several meetings. We limited eligible post-admission factors to those that could be measured using data from TOHDW. This criterion automatically excluded physiological measures such as vital signs, since such data are not captured in TOHDW.

Our final list of eligible covariates consisted of five post-admission factors: daily LAPS, admission to the intensive care unit (ICU), transfer to an acute monitoring area (AMA), change from active care to an alternative level of care (ALC), and performance of any therapeutic procedure(s) significantly associated with risk of in-hospital mortality.

3.7.3.1 Daily LAPS

Using the same methodology as that used to calculate LAPS at admission, (Section 3.7.2.5), I calculated patients' LAPS on each day of the admission using tests performed in the previous 24 hours. If a test had not been performed in the previous 24 hours, but had been performed previously during the admission, I used the result of the most recent test.

For the three LAPS components (arterial pH, serum troponin I, and total white blood cell count) where high risk patients were assigned points for missing data, I made one slight modification: I calculated the patient's provisional death risk (from the Laboratory Score Preliminary Model) *on that day* to determine whether the patient was high or low risk.

3.7.3.2 Admission to the ICU

The "Inpatient Census" table contains details of all in-patient "transactions" that occur during a hospitalization. Such transactions include admissions, discharges, and transfers between hospital services, nursing stations, or beds.

To determine if and when a patient was admitted to or discharged from the ICU, I used the "Inpatient Census" table to obtain the date and time of all transfers to and from an ICU nursing station. I compared these date-times to the admission date-time to determine the admission days on which the patient was in the ICU. If there was no transaction to indicate that the patient was transferred out of the ICU, I assumed the patient was in the ICU until the end of follow-up.

3.7.3.3 Transfer to an acute monitoring area

Acute monitoring areas (AMA) are nursing station units where patients are monitored in more detail than on the regular wards. Most of these areas have the capability to continuously monitor patients through telemetry.

To determine if and when a patient was transferred to or from an AMA, I used the "Inpatient Census" table to obtain the date and time of all transfers to and from an AMA nursing station unit. As with ICU admissions, I compared the date-times to the admission date-time to determine the admission days on which the patient was in an AMA. If there was no transaction to indicate that the patient was transferred out of the AMA, I assumed the patient was in the AMA until the end of follow-up.

3.7.3.4 Change from active care to an alternative level of care

"Alternative level of care" (ALC) patients are patients who no longer receive acute medical services, but remain in-hospital while awaiting placement in a more appropriate setting such as a rehabilitation or long-term care facility. Although ALC days usually occur at the end of a hospitalization, they can occur at any time during the admission (31).

The "Encounter" table contains the start and end date of up to four ALC "episodes" per admission (more than four ALC episodes per admission is extremely rare). If an admission had ALC dates present in this table, I determined the admission days on which the patient was ALC. If an admission had no ALC dates present, I assumed no ALC days occurred during the hospitalization.

3.7.3.5 Performance of in-hospital procedures

I searched the Ovid MEDLINE (R) database (1950 to April Week 2 2010) for studies that quantified (i.e. using a metric such as an index) the risk of in-hospital death associated with various in-hospital procedures in a generic (i.e. disease non-specific) patient population. Among 800 studies retrieved by the search, six studies (32-37) developed an index to quantify the risk of death following a range of surgeries. Four of these indices (38-41), however, were derived using a disease-specific patient population, and all indices included clinical variables (i.e. blood pressure, pulse rate, functional status, etc.) that could not be determined using data from TOHDW.

I therefore derived and internally validated the Procedure Independent Mortality Risk (PIMR) Index (Appendix E) (42). The PIMR index assigns scores based on the type and urgency of therapeutic procedures performed in-hospital. The advantages of this index are that it can be calculated using administrative data, it includes a broad range of therapeutic medical procedures (not limited to surgeries), and the risk scores quantify the independent influence of procedures *after* adjusting for all of the covariates in the Escobar model.

To calculate the PIMR score, I obtained the admission urgency status (used as a proxy for the procedure urgency) from the "HR Dad" table From the "Procedures" table, I obtained the date and time of all procedures performed during each admission and the Canadian Classification of Interventions (CCI) code of each procedure. I calculated the total PIMR score on each day by summing the points of all scored procedures that were performed during that day.

3.7.3.6 Functional form of time-dependent covariates

Since the value of time-dependent covariates can change over time, the method in which values are assigned to each day (called the functional form of the covariate) needs to be considered carefully. The choice of an incorrect form can lead to bias, but trying too many functional forms or extremely complex ones can lead to great overfitting of the data (21).

Upon discussion with Drs. Van Walraven and Forster, we decided to use a step function for all time-dependent covariates except in-hospital procedures (the PIMR score). With a step function, we used the most recent information available *prior* to the beginning of each day to assign a value that remained constant throughout the day. We choose this functional form because it was the most intuitive, it was simple and easy for others to understand, and it reduced the chance of overfitting the data. We agreed, however, to

explore other functional forms if the step function approach produced results that were contrary to our expectations.

For in-hospital procedures, we used a different approach because we knew the effect of procedures (particularly high-risk ones such as surgeries) extended beyond the day of the procedure. We hypothesized that the effect of PIMR procedures was strongest immediately following the procedure and gradually diminished over time. Given this *a priori* hypothesis, we decided to use two time-dependent covariates to model the effect of PIMR procedures over time. The first covariate was simply the PIMR score, where we changed the PIMR score the day *after* a PIMR episode occurred and kept the score constant until the day after the next PIMR episode or the end of follow-up, whichever occurred first (essentially a step function). The second covariate was "number of days since the last PIMR episode." We included this covariate so that the data would determine how the effect of the PIMR score changed over time (since we had no hypothesis about the rate or shape with which the risk attenuated over time).

3.8 Statistical analysis

3.8.1 Statistical software

I performed all data manipulation and statistical analyses in this study using SAS, Version 9.2 (Cary, NC) and SAS Enterprise Guide, Version 4.2 (Cary, NC).

3.8.2 Descriptive statistics

I used basic descriptive statistics to describe the characteristics of admissions in the derivation and validation sets. To describe the distribution of continuous variables in each set, I reported medians and interquartile ranges (IQR) because the distribution of most continuous variables (i.e. Elixhauser, LAPS, and PIMR) was skewed. To describe the categorical variables, I reported the proportion of admissions in each category.

3.8.3 Assessment of the Escobar model fit using logistic regression

Using PROC LOGISTIC and the BY statement for primary condition groups, I fit a separate logistic regression model containing the covariates in the Escobar model for each primary condition group.

Using the coefficients obtained from the derivation set, I calculated the probability of death for each admission in the validation set. Using these probabilities, I assessed the model discrimination by calculating the *C*-statistic. Informative pairs included all pairs where only one subject died. Pairs in which both subjects died or survived were not considered in the calculation because they could not be classified as discordant or concordant (termed "non-informative" pairs). Since this was a logistic regression model, a concordant pair was one in which the predicted risk was higher for the subject who died. Pairs in which both subjects had the same predicted risk were considered "tied" and counted as half a concordant and half a discordant pair. I determined the *c*-statistic and 95% confidence interval using a macro (%ROC) written by Mithat Gönen (43).

I assessed the model calibration using the Hosmer-Lemeshow goodness-of-fit test. I arranged admissions in order of increasing predicted risk, divided admissions into risk deciles, and calculated the Hosmer-Lemeshow test statistic according to the equation

$$\chi^2_{(df=8)} = \sum_{J=1}^{10} \frac{(O_J - E_J)^2}{E_J(1 - E_J / n_J)}, \tag{3.1}$$

where O_J represented the number of observed events (deaths) in decile j, E_J represented the number of expected events in decile j, and n_J represented the number of admissions in decile j. To determine the p-value of the test statistic, I compared it to a chi-square distribution with eight degrees of freedom.

3.8.4 Derivation of Escobar[+]

3.8.4.1 Counting process approach

The counting process approach is extremely useful when working with survival data (44), particularly data containing time-dependent covariates since the value of such covariates (and consequently the model-based predicted risk) can change over time. In this study, I used the counting process approach to derive the time-dependent risk prediction model and obtain *daily* estimates of the model-based predicted risk (i.e. the linear predictor) and two diagnostic residuals (the scaled score residual and the Martingale residual) for each admission.

To use the counting process approach in SAS, I had to arrange the study dataset in the counting process style of input. In this dataset, I created a row for each day of each admission (since this was the shortest time interval over which a patient's set of covariate values could remain constant) and specified the corresponding covariate values and outcome status for that day. To identify which day each row belonged to, I defined two variables, t_1 and t_2, which represented a semi-closed time interval $(t_1, t_2]$. For example, the values for day 1 were represented by the row where $t_1 = 0$ and $t_2 = 1$.

3.8.4.2 BY versus STRATA for primary condition groups

In PROC PHREG, I could use either the BY or STRATA statement to account for the different primary condition groups. Using the BY statement would fit a separate Cox regression model (and set of parameter estimates) for each primary condition group. Using the STRATA statement would fit a single set of parameter estimates for the entire cohort, but allow the baseline hazard estimate to vary between primary condition groups.

I compared the discrimination of a Cox regression model containing the Escobar covariates using BY versus STRATA to account for the primary condition groups. In the validation cohort, the model discrimination was significantly better with the STRATA

statement (*c*-statistic 0.839, 95% confidence interval [CI] 0.830-0.847]) than the BY statement (0.765, 95% CI 0.755-0.776]). As a result, I used the STRATA statement to account for the different primary condition groups in all Cox regression models.

3.8.4.3 Exploratory analyses

Using admissions in the derivation set only, I created bar charts, histograms, and box-and-whisker plots to visualize the distribution of covariate values and to identify extreme or erroneous values. I also performed a simple bivariable analysis of the association between each covariate and the risk of in-hospital death by comparing the proportion of deaths in each covariate category, where I grouped continuous covariates into categories containing at least 250 admissions per category. For binary time-dependent covariates (ICU, ALC, and AMA), I used the *daily* covariate and outcome status to calculate the mortality rate (and 95% confidence interval) for each covariate level for the first 22 days of admission (the 95th percentile of length of stay in the derivation set). I calculated 95% confidence intervals using exact methods (45).

To investigate the nature of the association between each continuous covariate and the log hazard, I categorized each continuous variable into equidistant cut-points (ensuring at least 250 admissions per group) and plotted the parameter estimate for each group from a univariable Cox regression model containing the categorized continuous covariate (i.e. a plot where the parameter estimates formed a straight line would suggest a linear association). I also determined the best form for each continuous covariate in a univariable Cox regression model using the standard fractional polynomial function selection procedure described by Sauerbrei *et al.* (46). Because the function selection procedure required covariates to have a domain greater than zero, I transformed covariates whose domain included values ≤ 0 by shifting all values up by the theoretical minimum value + 1. To keep the Cox model simple and reduce the number of covariates required to test the

significance of interactions (i.e. interactions with time to assess the proportional hazards assumption), I limited the most complex function to a first-degree fractional polynomial. I restricted the choice of possible powers for each covariate to the usual set of values: p = {-2, -1, -0.5, 0, 0.5, 1, 2, 3} for x^p, where x is a positive continuous covariate, p is the fractional polynomial value used to transform x, and, in this case, x^0 denotes log(x). For time-independent covariates only, I ran a univariable model and checked graphically for covariate misspecification using the ASSESS statement and VAR= option in PROC PHREG to compare the observed cumulative Martingale residual process against 1000 simulated paths from the null distribution (47). Close agreement between the observed and simulated paths suggested good covariate specification, as did a non-significant p-value from the Kolmogorov-type supremum test (which I obtained using the RESAMPLE option in the ASSESS statement) (47). I could not use the ASSESS statement options in PROC PHREG to check for covariate misspecification among time-dependent covariates because these options were not available in SAS for models containing time-dependent covariates.

I did not model continuous variables with a spline in order to keep the final risk prediction model simple and reduce the number of covariates required to test the significance of interactions. However, for age only, I evaluated age^2 expressed as a restricted cubic spline because this is how age was expressed in the Escobar model. To decide which form of age to use in the *Escobar*$^+$ model, I used the value of Akaike's information criterion (AIC) to compare the fit of a univariable Cox model containing splined age^2 versus the form recommended by the fractional polynomial function selection procedure.

For time-independent covariates only, I assessed the proportional hazard assumption using four different methods. First, I created a log-negative-log plot [i.e. log(-log(S(t)))] for each covariate level. For continuous variables, I created the plots by

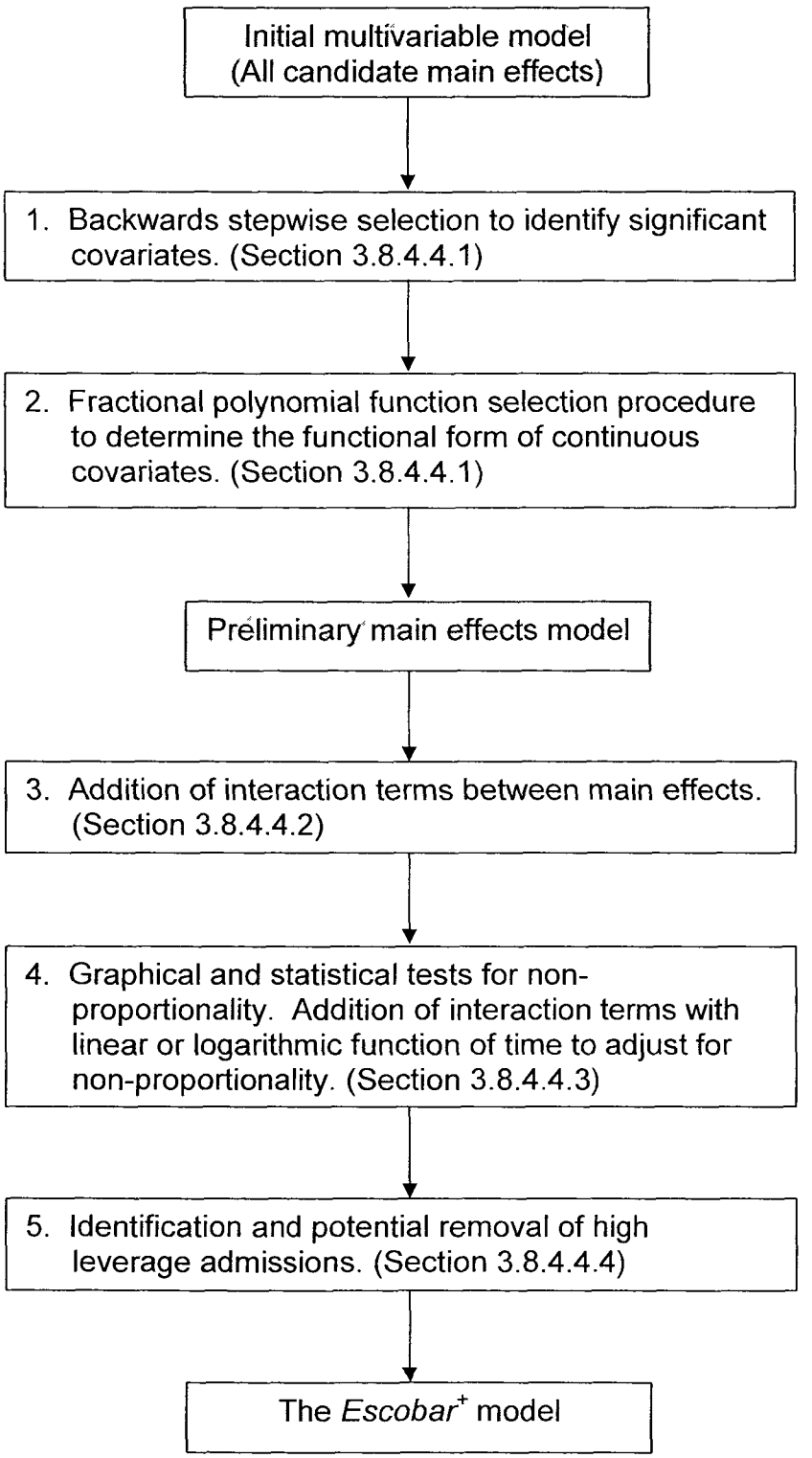
categorizing the values into four levels using the recommended form from the fractional polynomial function selection procedure. A plot with fairly parallel lines suggested that the hazard between covariate levels was proportional over time. Second, I ran a univariable model and used the ASSESS statement and PH option in PROC PHREG to compare the observed score process (also called the cumulative sum of the weighted Schoenfeld residuals) against 200 simulated processes created under the assumption of proportional hazards (47;48). Close agreement between the observed and simulated processes suggested proportionality, as did a non-significant *p*-value from the Supremum test (which I obtained using the RESAMPLE option with the ASSESS statement and PH option) (47). Third, I plotted the weighted Schoenfeld residuals (defined for uncensored admissions only) from a univariable model and used PROC REG to test for a linear trend over time. A non-significant trend suggested proportional hazards, whereas a significant increasing (decreasing) trend suggested an increasing (decreasing) hazard ratio over time (49). Finally, I tested the significance of adding a time-dependent interaction term between the covariate and two functions of time (linear and logarithmic) using the likelihood ratio test. I did not test more complex functions of time because simpler functions are usually preferred (50). A significant time-dependent interaction with time suggested non-proportionality.

I did not assess the proportional hazard assumption for time-dependent covariates because the methods used to assess proportionality for time-independent covariates could not be applied to time-dependent covariates. Furthermore, there are currently no established methods to test the assumption of proportionality for time-dependent covariates because such covariates are usually used to adjust for non-proportionality in time-independent covariates.

3.8.4.4 Model-building process

To build the *Escobar*$^{+}$ model, I followed the methods described by Hosmer and Lemeshow (44) (depicted in Figure 3.2).

Figure 3.2 Model-building approach used to derive the *Escobar*$^{+}$ model



3.8.4.4.1 Preliminary main effects model

To derive the preliminary main effects model, I started with an initial multivariable model containing all candidate main effects. For continuous variables, I used the functional form recommended by the fractional polynomial function selection procedure in the exploratory analysis. For age, I used either splined age^2 or the functional form recommended by the function selection procedure, depending on which form produced the model with the lowest AIC in the exploratory analysis.

Using a backwards stepwise selection approach and a significance level of 0.05, I removed non-significant covariates one at time in order of decreasing *p* value. I used the *p*-value of the partial likelihood ratio test to confirm that the deleted covariate was not significant. I also checked whether the non-significant covariate was an important confounder by determining the percent change in coefficients between the "reduced" and "full" models. I excluded the covariate if all coefficients changed by less than 10%. I then added all previously excluded covariates back into the model to confirm they were not significant predictors or important confounders. I repeated this process until only significant covariates and important confounders remained.

Next, I used the methods described by Sauerbrei and Royston (51) to identify the best first-degree fractional polynomial transformation for all continuous covariates in a *multivariable* model containing all significant main effects. I limited the most complex function to a first-degree fraction polynomial to simply the model and reduce the number of covariates required to test the significance of interactions (i.e. interactions between main effects and with time). To implement the function selection procedure, I used a SAS macro (%MFP8) to perform the multivariable fractional polynomial algorithm (52). In % MFP8, I entered all continuous covariates in their original, untransformed state and used the same set of restricted powers as in the exploratory analysis.

3.8.4.4.2 Interactions between main effects

I automatically considered interaction terms from the Escobar model for inclusion in the *Escobar*$^{+}$ model To identify additional candidate interaction terms, I met with Drs. van Walraven and Forster to come up with a list of biologically plausible interactions. Given the large size of the derivation cohort, I chose not to test a wide assortment of interaction terms because it was likely that some would be significant by chance.

I first tested the significance of each candidate interaction term separately in the preliminary main effects model using the partial likelihood ratio test. I then added all significant interaction terms ($p<0.05$) jointly to the main effects model and removed non-significant interactions in order of decreasing p-value.

To visualize the effect of each significant interaction on the risk of in-hospital death, I created "risk score plots." To create each plot, I used the model coefficients to calculate the risk score (i.e. the linear predictor, $\Sigma\beta x$) when the value of one covariate was allowed to vary by different values of the other covariate. For the other covariates not involved in the interaction, I used the median or reference value to calculate the risk score.

3.8.4.4.3 Tests for non-proportionality

For the time-independent main effects and interaction terms only, I created plots of the weighted Schoenfeld residuals for each covariate and assessed the proportional hazard assumption graphically by testing for the presence of a linear trend. Because the model contained time-dependent covariates, I could not obtain graphs of the observed score process versus simulated processes (as done in the exploratory analysis). I also tested the significance of adding interaction terms between each time-independent main effect and two functions of time (linear and logarithmic) to the model using the partial likelihood ratio. I decided whether the interaction with the linear or logarithmic function of time was better by determining which function produced the model with the lowest AIC.

After testing the significance of each interaction term separately in the model, I added all significant interactions jointly to the model and removed non-significant interactions in order of decreasing *p*-value.

As per the exploratory analysis, I did not test for non-proportionality among the time-dependent main effects and interaction terms.

3.8.4.4.4 Identification of outliers and influential admissions

I assessed the influence (or leverage) of each admission on the model coefficients using the scaled score residuals. The scaled score residuals represent the approximate change in each coefficient (β) when each admission is omitted. Using the counting process approach and the OUTPUT statement and DFBETA option in PROC PHREG, I obtained a *daily* residual for each admission-covariate combination and calculated the total scaled score residual for each admission by summing the component daily residuals.

I then created plots of the total scaled score residuals for each covariate. I determined the study ID of admissions whose residual appeared to deviate significantly from the rest. After identifying all potentially influential admissions from the plots, I refit the model excluding these admissions and calculated the percent change in the coefficients using the "reduced" versus "full" cohort. I consulted with Drs. van Walraven and Forster to decide whether to include or exclude each influential admission from the final model. We excluded admissions that we felt were exceptional or clinically implausible.

3.8.4.5 Model assessment

3.8.4.5.1 Discrimination

In a Cox regression model, the concept of concordance considers not only whether the event occurs, but also the time at which the event occurs. With survival data, a concordant pair is one in which the model-based predicted risk is higher for the subject

who dies *earlier.* Pairs in which both subjects either die at the same time or have the same model-based predicted risk can be counted as "ties" (i.e. half a concordant and half a discordant pair). Informative pairs therefore include all pairs where both subjects die, or only one subject dies and the other (censored) subject is followed-up beyond the event time of the first subject. Non-informative pairs include pairs where both subjects are censored, or one subject dies and the other subject is censored before the event time of the first subject.

I used the methods described by Kremers (6) to measure concordance for the Escobar model (fit using Cox regression) and the *Escobar*$^{+}$ model. This method allows the inclusion of time-dependent covariates and ties in prediction and/or event times, and can be used to compare models with and without time-dependent covariates. To calculate the *C*-statistic and approximate 95% confidence interval for all Cox regression models, I used the SAS macro, %SurvCsTD (53). The macro produced two estimates of the *c*-statistic: one which counted ties in risk score only, and another that counted ties in both risk score and/or event time. I made an *a priori* decision to use the latter estimate for all models.

3.8.4.5.2 Calibration

I assessed calibration using Martingale residuals, which I obtained from PROC PHREG using the OUTPUT statement and RESMART option. The Martingale residual (M_i) represents the difference between the observed and expected number of events (as predicted by the Cox regression model) for a particular subject *i* (54). Using the *daily* Martingale residuals (which represent the difference between the observed and *conditional* expected number of events on each day), I calculated the daily expected number of events for each admission as

$$\text{“expected}_{ij}\text{”} = \text{“observed}_{ij}\text{”} - M_i, \qquad (3.2)$$

where "expected$_{ij}$" represented the number of expected events for subject *i* on day *j*, "observed$_{ij}$" represented the vital status of subject *i* at the end of day *j*, and M_{ij} represented the Martingale residual for subject *i* on day *j*.

I compared the number of expected to observed events at a common time point for all admissions. I chose to assess calibration at four different time points during the admission: days 1, 2, 6, and 22 (corresponding to the 25th, 50th, 75th and 95th percentile of length of stay in the derivation cohort). At each time point, I identified all admissions where the patient was still in hospital and determined the patient's risk score (i.e. the linear predictor, $\Sigma\beta x$) on that day. Using these risk scores, I divided admissions into deciles and summed the number of expected and observed events on that day. To assess the level of agreement between the number of observed and expected events in each decile, I calculated the *z*-statistic and its associated *p*-value [using the methods described by May and Hosmer (54)], where

$$z = (\text{observed-expected})/(\sqrt{\text{expected}}). \quad (3.3)$$

3.8.5 Validation of Escobar^{+}

To validate the *Escobar*$^{+}$ model, I assessed the model's predictive accuracy when the coefficients obtained from the derivation set were applied to the validation set. I assessed the model discrimination in the validation set by calculating the *c*-statistic and associated 95% confidence interval with the same methods used in the derivation set.

To measure calibration, I could not calculate the daily number of expected events as in the derivation set because I could only obtain Martingale residuals in SAS for admissions used in the model derivation. According to May and Hosmer (54), however, the expected number of events for an admission can be calculated using the equation

$$\text{"expected}_i\text{"} = -\exp\left(\left[\sum \beta x\right]_i\right) * \log\left(S_0\left(t_{(i)}\right)\right), \quad (3.4)$$

where "expected$_i$" represents the number of expected events for subject *i*, $[\Sigma\beta x]_i$ represents the risk score for subject *i* (calculated by summing the product of each model coefficient and corresponding covariate value for subject *i*), and $S_0(t_{(i)})$ represents the baseline survivor function estimate at the end of follow-up for subject *i*. A simple rearrangement of equation 3.4 proves that the expected number of events can be calculated using the cumulative baseline hazard function instead, since

$$\begin{aligned} \text{"expected}_i\text{"} &= -\exp\left(\left[\sum \beta x\right]_i\right) * \log\left(S_0\left(t_{(i)}\right)\right) \\ &= \exp\left(\left[\sum \beta x\right]_i\right) * -\log\left(S_0\left(t_{(i)}\right)\right) \\ &= \exp\left(\left[\sum \beta x\right]_i\right) * H_0\left(t_{(i)}\right), \end{aligned} \tag{3.5}$$

where $H_0(t_{(i)})$ represents the cumulative baseline hazard function at the end of follow-up for subject *i*.

Note that equations 3.4 and 3.5 refer to a time-independent Cox regression model since it assumes that the risk score stays constant for each subject over the entire follow-up period. Since the *Escobar*$^+$ model was time-dependent, however, the risk score for each subject could change on a daily basis. I therefore calculated the *conditional daily* expected number of events (or equivalently the daily hazard estimate) using the equation

$$\text{"expected}_{ij}\text{"} = \left(\exp\left[\sum \beta x\right]_{ij}\right) * \left(H_0\left(t_j\right) - H_0\left(t_{j-1}\right)\right), \tag{3.6}$$

where "expected$_{ij}$" represented the conditional number of expected events for subject *i* on day *j*, $[\Sigma\beta x]_{ij}$ represented the risk score for subject *i* on day *j*, and $H_0(t_j)$ represented the cumulative baseline hazard estimate on day *j*. I used primary condition-specific estimates of the cumulative baseline hazard function in equation 3.6, which I obtained in SAS [following the methods of Wang (55)] by: a) appending "fake" baseline records (i.e. rows with the value of each covariate set to the baseline or reference value and the outcome status set to missing) for each primary condition group and day of interest (i.e. day 1, 2, 6, and 22) to the study dataset; b) using the OUTPUT statement and SURVIVAL option in

PROC PHREG to obtain estimates of the baseline survivor function; and c) converting the survivor function estimates to cumulative hazard estimates using equation 3.7:

$$H(t) = -\log(S(t)) \tag{3.7}$$

I validated this method of calculating the daily expected number of events by comparing the number of expected events calculated using equation 3.6 to the number of expected events calculated using the Martingale residuals in the derivation cohort. I found that both methods produced identical results.

3.8.6 Sensitivity analyses

3.8.6.1 Informative censoring

In this study, censoring was not random because censored patients (i.e. patients discharged alive) were almost certainly at a lower risk of in-hospital death (otherwise they would not have been discharged). Because I could not adjust for informative censoring using standard methods, I performed a sensitivity analysis to investigate the impact that informative censoring may have had on the study results (56). To do this, I re-ran the model changing the censoring time of all censored patients to one day greater than the largest event time in the derivation set (425 days), corresponding to the extreme assumption that all censored patients, had they not been discharged from hospital, would have survived past the largest event time. I then compared the discrimination of the *Escobar*$^{+}$ model in the validation set when the observed versus extreme censoring times were used to derive the model coefficients.

3.8.6.2 Clustering of admissions

In this study, all admissions were not independent of each other. In the derivation set, approximately 27% of patients contributed more than one admission to the set (18.5% of patients contributed two admissions and 5% contributed three admissions). I therefore

performed a sensitivity analysis where I adjusted for clustering using the marginal approach (57) and compared the adjusted parameter estimates and standard errors [obtained using the COV(AGGREGATE) option in the PROC PHREG statement] with those from the unadjusted model.

3.8.6.3 Irregular measurement intervals

To correct for irregularly collected observations in their Cox regression model, De Bruijne *et al.* (58) added a time-dependent covariate representing "time elapsed since the person was last assessed." Their hypothesis was that a "fresh" or newly updated value was more informative than an "old" one. De Bruijne *et al.* found that this time-dependent covariate was an important predictor of death in addition to time-dependent white blood cell counts among patients with chronic myeloid leukemia.

In this study, the laboratory tests used to calculate LAPS were not performed at regular time intervals, and the frequency with which tests were performed was likely associated with risk of in-hospital death (sicker patients likely had tests performed more frequently than healthier patients). Assuming that (in the majority of cases) LAPS would have changed if enough laboratory test results were updated, I performed a sensitivity analysis to correct for irregularly performed lab tests by testing the significance of a time-dependent covariate representing "time elapsed since the last change in LAPS."

Since the other post-admission events were recorded immediately when they occurred, I did not test the significance of "time elapsed since the last observation" for these covariates because they were measured at all times throughout the admission.

4. Results

4.1 Creation of the study cohort

Between 1 April 2004 and 31 March 2009, 278 630 individual admissions met the study criteria (Figure 4.1). After concatenation of interhospital transfer admissions, a total of 275 460 admissions involving 172 396 patients remained. Of these, 159 794 (58%) were in-patient admissions, 113 175 (41%) were non-overnight same-day surgery admissions, and 2491 (0.9%) were overnight same- day surgery admissions.

I randomly divided all admissions into a derivation and validation set. The derivation set contained 183 640 admissions with 130 049 unique patients, and the validation set contained 91 820 admissions with 75 224 unique patients. The distribution of characteristics in the derivation and validation set was very similar (Table 4.1).

Figure 4.1 Creation of the study cohort

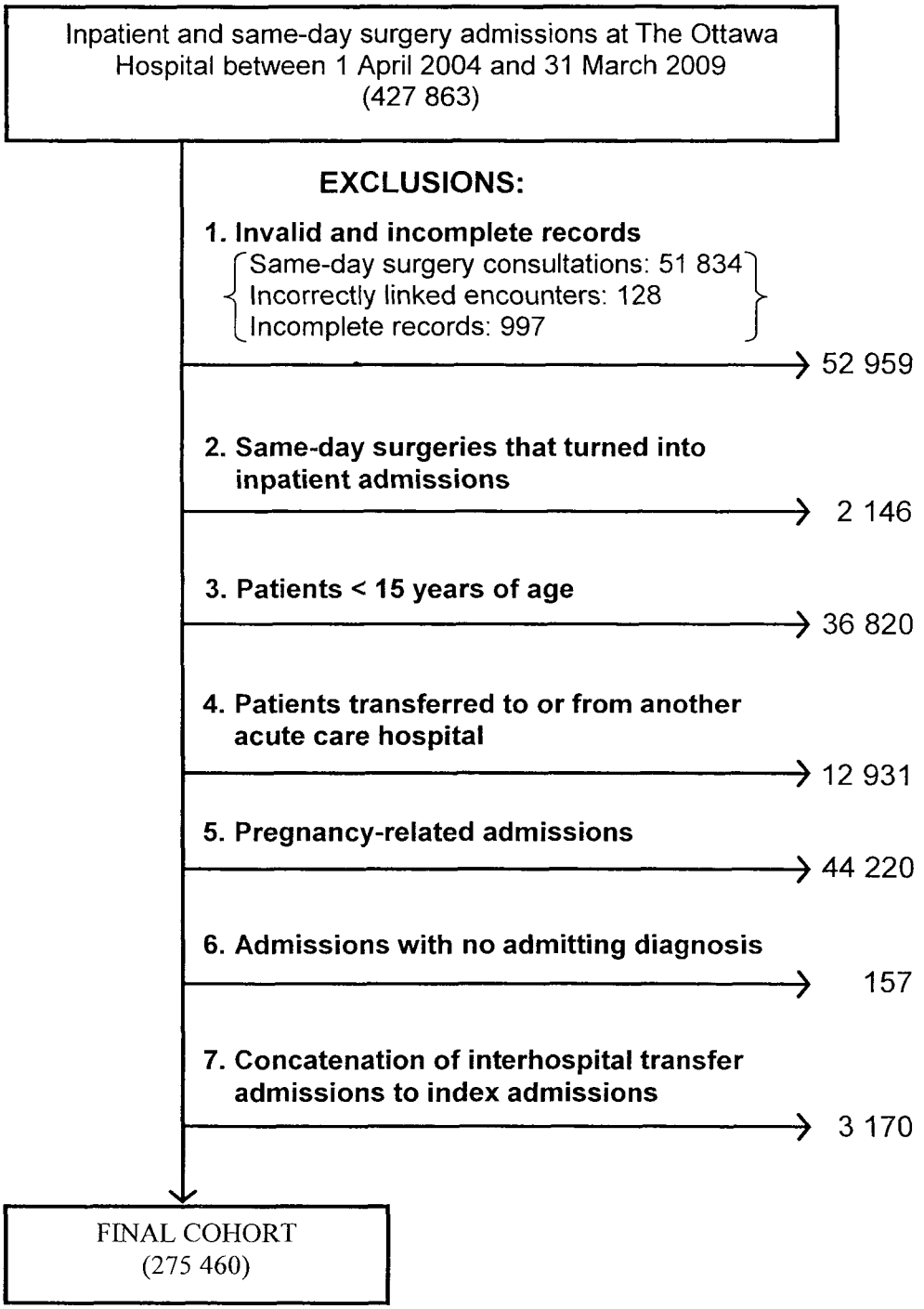


Table 4.1 Characteristics of the study cohort

Characteristic	Derivation	Validation
Patients/Hospitalizations, n*	130049/183640	75224/91820
Deaths in-hospital, n (%)	5362 (2.9)	2691 (2.9)
Length of admission in days, median (IQR*)	2 (1-6)	2 (1-6)
Male, n (%)	88164 (48.0)	44127 (48.1)
Age at admission, median (IQR)	61 (46-74)	60 (46-74)
Admission type, n (%)		
Emergent surgical	22536 (12.3)	11185 (12.2)
Emergent non-surgical	50005 (27.2)	24845 (27.1)
Elective non-surgical	14239 (7.8)	6916 (7.5)
Elective surgical	96860 (52.7)	48874 (53.2)
Elixhauser score, median (IQR)	0 (0-4)	0 (0-4)
LAPS* at admission, median (IQR)	0 (0-15)	0 (0-13)
Admissions to intensive care unit, n (%)		
0	178196 (97.0)	89173 (97.1)
1	5003 (2.7)	2455 (2.7)
2	380 (0.2)	159 (0.2)
3	50 (0.0)	25 (0.0)
4+	11 (0.0)	8 (0.0)
Transfers to acute monitoring area, n (%)		
0	176266 (96.0)	88214 (96.1)
1	6914 (3.8)	3382 (3.7)
2	399 (0.2)	185 (0.2)
3	49 (0.0)	21 (0.0)
4+	12 (0.0)	18 (0.0)
Alternative level of care episodes, n (%)		
0	178819 (97.4)	89446 (97.4)
1	4149 (2.3)	2052 (2.2)
2	527 (0.3)	245 (0.3)
3	99 (0.1)	55 (0.1)
4+	46 (0.0)	22 (0.0)
PIMR* episodes, n (%)		
0	150187 (81.8)	75361 (82.1)
1	30097 (16.4)	14838 (16.2)
2	2241 (1.2)	1076 (1.2)
3	654 (0.4)	322 (0.4)
4+	461 (0.3)	223 (0.2)
First PIMR score†, median (IQR)	-2 (-5-2)	-2 (-5-2)
Most common primary conditions, n (%)		
Neurologic problems, mental disorders, and senility (excluding seizures and drug overdoses)	38748 (21.1)	19391 (21.1)
Arthropathies and spine disorders (excluding infections and autoimmune conditions)	15033 (8.2)	7549 (8.2)
Gynecology and gynecologic cancers	12794 (7.0)	6375 (6.9)
Acute myocardial infarction	10945 (6.0)	5407 (5.9)

(cont'd on next page)

(Table 4.1 cont'd)

Characteristic	Derivation	Validation
Malignant neoplasms (excluding malignant respiratory and intrathoracic neoplasms, leukemias, non-Hodgkin's lymphomas, other histiocytic malignancies, ovarian cancer, and metastatic cancers)	9014 (4.9)	4455 (4.9)
Miscellaneous conditions (ICD-9: certain V codes, and all E codes)	8451 (4.6)	4269 (4.6)
Appendicitis, hernias, cholecystitis, and cholangitis	8352 (4.5)	4205 (4.6)
Ingestions and benign tumors	6125 (3.3)	3057 (3.3)
Miscellaneous conditions (ICD-9: 990-999)	5216 (2.8)	2606 (2.8)
Infections (excluding urinary infections, heptatitis, sepsis, meningitis)	4994 (2.7)	2435 (2.7)

*n=number; IQR=interquartile range; LAPS=Laboratory-based Acute Physiology Score; PIMR= Procedure Independent Mortality Risk

†among admissions with at least 1 PIMR episode

4.2 Cohort characteristics

At admission, the median (IQR) patient age was 61 (46-74) in the derivation set and 60 (46-74) in the validation set (Table 4.1). The patient was male in 48% of admissions. The admission types, in decreasing order of frequency, were elective surgical (53%), emergent non-surgical (27%), emergent surgical (12%), and elective non-surgical (8%). The median (IQR) Elixhauser score was 0 (0-4), indicating that the majority of patients had few comorbidities. The median (IQR) LAPS at admission was low: 0 (0-15) in the derivation set and 0 (0-13) in the validation set.

During the hospitalization, few patients were admitted to the ICU (3%), transferred to an AMA (4%), or changed to ALC status (3%). At least one PIMR procedure was performed during 18% of admissions. Most of these procedures were low-risk, as indicated by the median (IQR) score of the first PIMR episode of -2 (-5 to 2) (PIMR scores less than zero are associated with a decreased risk of in-hospital death). Few admissions (2%) had more than one PIMR episode.

Most hospitalizations were short, with a median (IQR) stay of 2 (1-6) days. Overall, the patient died in 2.9% of all admissions. During the study, the annual proportion of

deaths in-hospital decreased significantly from 3.3% (95% confidence interval [CI] 3.1-3.5%) in 2004 to 2.4% (95% CI 2.2-2.7%) in 2009.

4.3 Predictive accuracy of the Escobar model fit using logistic versus Cox regression

When I fit the Escobar model using logistic regression, the model had excellent discrimination, with a *c*-statistic in the validation cohort of 0.927 (95% CI 0.924-0.931). The model calibration was also excellent. Although the *p*-value of the Hosmer-Lemeshow test statistic in the validation cohort was significant (p=0.004), the number of observed versus expected events in the first seven risk deciles was very close (Table 4.2).

When I fit the Escobar model using Cox regression, the predictive accuracy of the model was not as good: the *c*-statistic in the validation cohort was 0.839 (0.830-0.847). Compared to the logistic model, the calibration of the Cox model appeared slightly better when I grouped admissions into risk deciles based on the predicted risk from their respective models. For the Cox model, there was a statistically significant difference between the observed and expected number of events in only one risk decile (compared to two risk deciles for the logistic model), and the *p*-value was higher (indicating closer agreement between the observed and expected number of events) in six risk deciles (Table 4.1). However, the calibration of the Cox model appeared worse when I grouped admissions using the same risk deciles (based on the predicted risk from the logistic model). The number of deaths predicted by the Cox model was less accurate than the logistic model in six risk deciles (Table 4.2).

Table 4.2 Calibration of the Escobar model fit using logistic versus Cox regression

	Logistic				Cox					
Risk decile	N	Observed # of deaths	Expected # of deaths*	*p*-value	N	Observed # of deaths	Expected # of deaths**	*p*-value	Expected # of deaths*	*p*-value
1	9168	0	0.61	0.4344	9181	1	2.81	0.2798	2.69	0.1008
2	9348	0	1.50	0.2204	9202	3	5.92	0.2305	4.72	0.0299
3	8928	3	2.57	0.7892	9170	5	9.09	0.1747	5.38	0.3048
4	9262	2	4.34	0.2620	9189	8	10.11	0.5067	5.40	0.1435
5	9204	8	8.04	0.9875	9288	7	14.06	0.0598	10.65	0.4169
6	9179	30	25.10	0.3283	9061	24	25.14	0.8202	32.81	0.6233
7	9185	57	57.55	0.9418	9183	63	62.91	0.9914	64.26	0.3650
8	9182	190	149.30	0.0009	9182	258	190.48	<.0001	168.36	0.0954
9	9182	551	499.77	0.0219	9182	557	554.12	0.9028	500.14	0.0230
10	9182	1850	1890.63	0.3501	9182	1765	1794.22	0.4903	1874.45	0.5723
Total	91820	2691	2639.42	0.3154	91820	2691	2668.86	0.6683	2668.86	0.6683

*For risk deciles based on predicted risk of in-hospital death from the logistic model
**For risk deciles based on predicted risk of in-hospital death from the Cox model

4.4 Exploratory analyses

4.4.1 Simple bivariable analyses

The mortality rate in the derivation set was similar among males (3.2%) and females (2.6%). For admission type, the mortality rate was highest among emergent non-surgical admissions (8.5%), followed by emergent surgical admissions (4%). The mortality rate was very low among elective non-surgical (0.5%) and elective surgical admissions (0.1%).

The proportion of in-hospital deaths increased rapidly as age at admission increased (Figure 4.2). Among patients aged 50-54, 70-74, and 95+ years at admission, the mortality rate was 1.5%, 4%, and 19.5%, respectively.

The mortality rate was higher among patients with greater comorbidities, indicated by a higher Elixhauser score (Figure 4.3). Among patients with a score of 15-19 and 25+, the mortality rate was 13.5% and 37%, respectively. These rates were noticeably higher than the mortality rate of 1% among patients with a score of 0-4.

The mortality rate was also higher among hospitalizations with a higher LAPS at admission (Figure 4.4). Among admissions with a score of 0-9, 60-69, and 120+, the mortality rate was 0.5%, 15%, and 51%, respectively. When I plotted the mean LAPS score by outcome on each admission day, I found that the mean scores were significantly higher among patients who died that day compared to patients who survived (Figure 4.5). In addition, among all admissions, I noticed a spike in the mean score on day two (Figure 4.5). The lower mean scores on day one were likely an artifact, since 69% of admissions had none of the 14 laboratory tests performed in the 24 hours prior to admission.

For the PIMR score, the mortality rate was 2% among admissions with a score of 0 (Figure 4.6). Comparatively, the mortality rate was lower among admissions with a

negative score (ranging from 0.06% to 0.36%), and substantially higher among admissions with a positive score (ranging from 9.5% to 66%).

Patients were admitted to the ICU in less than 3% of admissions, but the mortality rate among these patients (33%) was much higher than those where the patient was never admitted to the ICU (2%). When I plotted the daily mortality rate by daily ICU status, I also observed that the death rate was consistently higher among ICU patients compared to non-ICU patients (Figure 4.7).

The overall mortality rate among patients who were transferred to an acute monitoring area (AMA) was 8%, compared to patients who were never transferred to an AMA (2.7%). However, when I plotted the daily mortality rate by daily AMA status, the rates for AMA patients were not significantly higher, although there was an increasing trend over time in the difference between the mortality rate for AMA patients and non-AMA patients (Figure 4.8). This trend was likely because patients who are transferred to an AMA often experience complications or new developments later on in their stay.

I found that the overall mortality rate among admissions with at least one ALC episode (9.1%) was higher than the rate among admissions with no ALC episodes (2.8%). This is expected since ALC patients (who are awaiting long-term care) are usually sicker and more medically compromised than the average patient. However, when I plotted the daily mortality rate by daily ALC status (Figure 4.9), the mortality rate was higher (though not statistically significant) among ALC patients for the first 10 days only, after which the mortality rate was significantly higher among *non*-ALC patients. This can be explained because fewer patients are in hospital after day 10 (Figure 4.9), most of whom are more medically compromised than ALC patients (i.e. have a higher risk of death).

I also calculated the mortality rate by categories of length of stay. I found that the mortality rate was lowest among admissions shorter than a week (1.7%) and increased

with time, reaching a maximum of 16% for admissions between eight and nine weeks long (Figure 4.10).

Figure 4.2 Proportion of deaths by age at admission in the derivation set

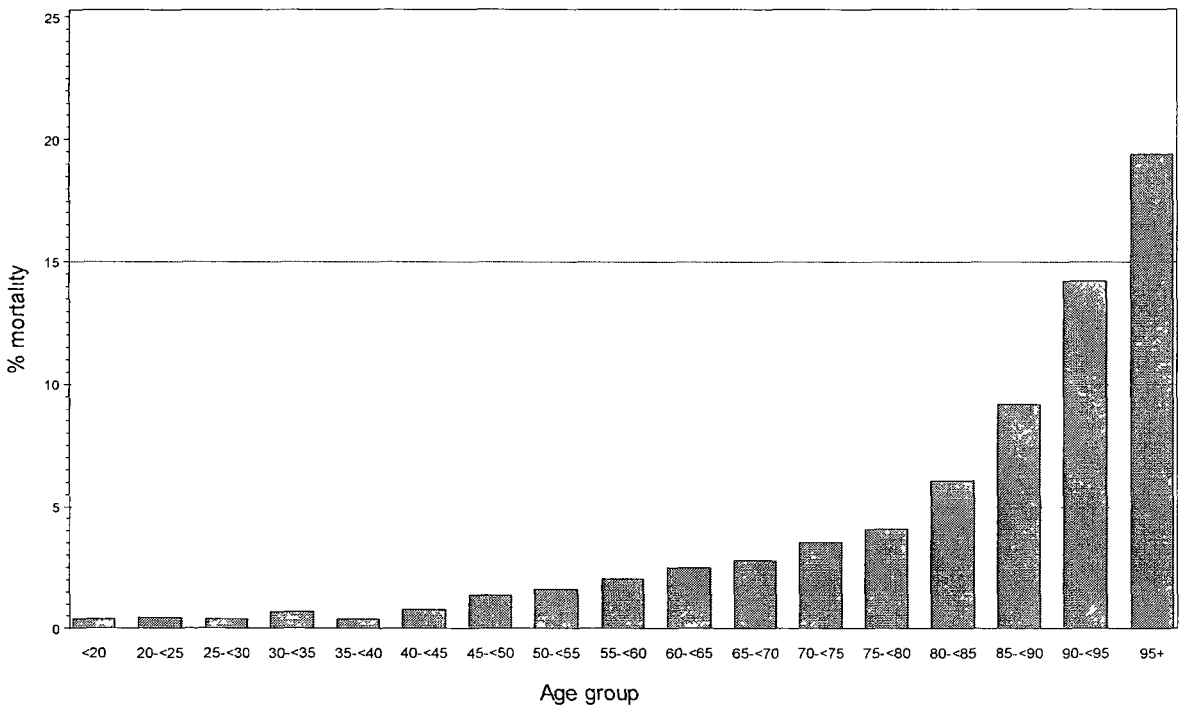


Figure 4.3 Proportion of deaths by Elixhauser score in the derivation set

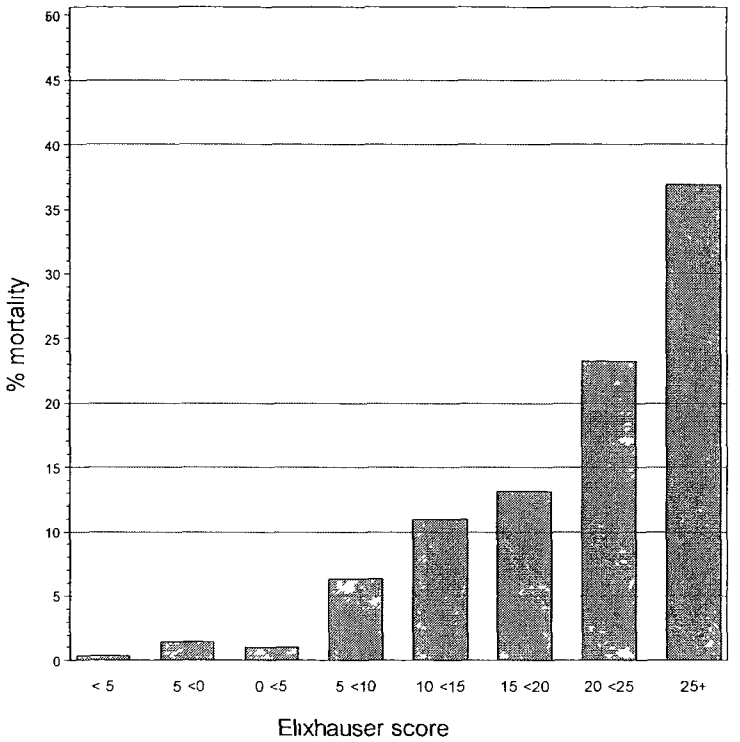


Figure 4.4 Proportion of deaths by LAPS at admission in the derivation set

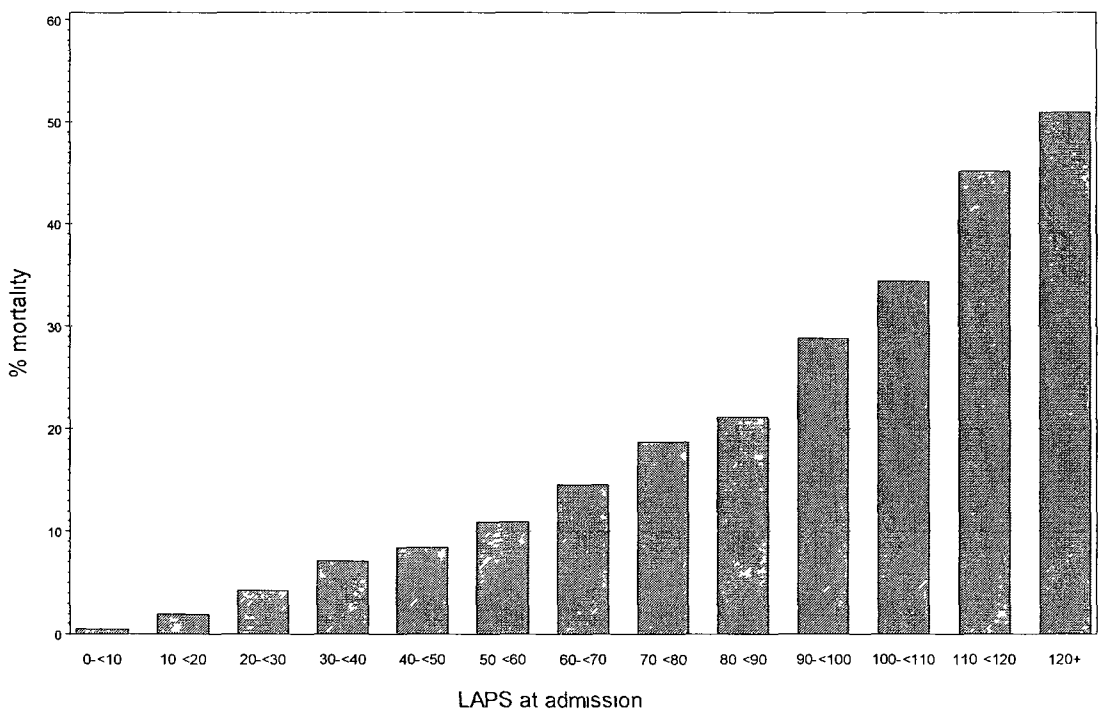
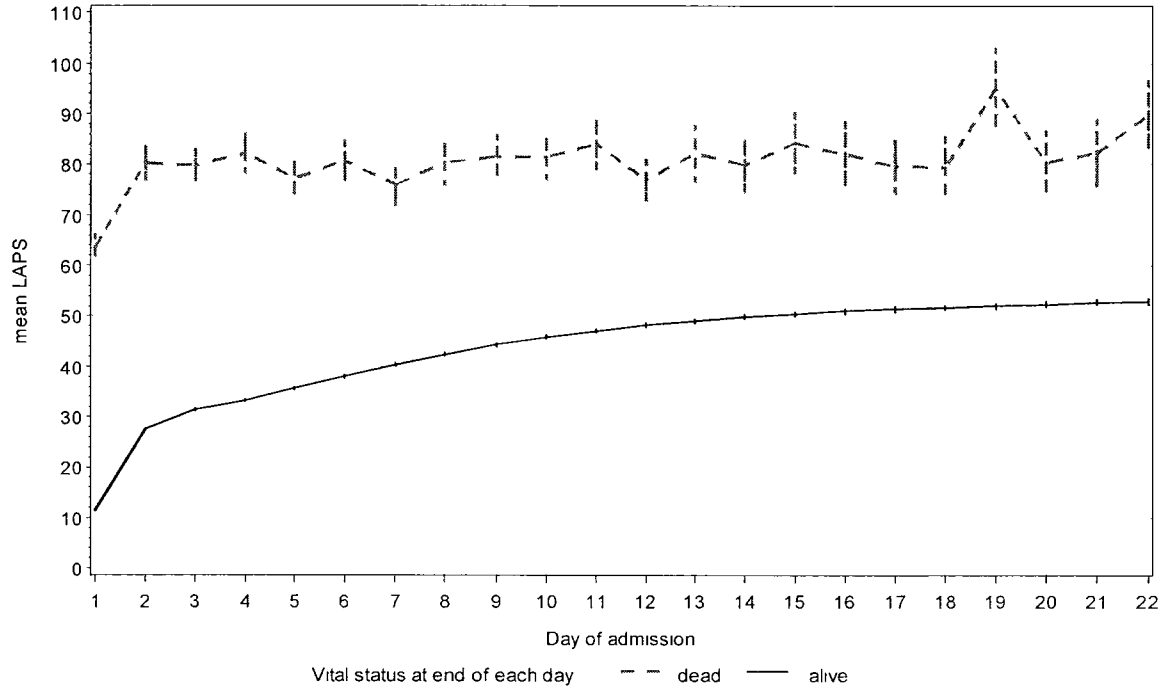
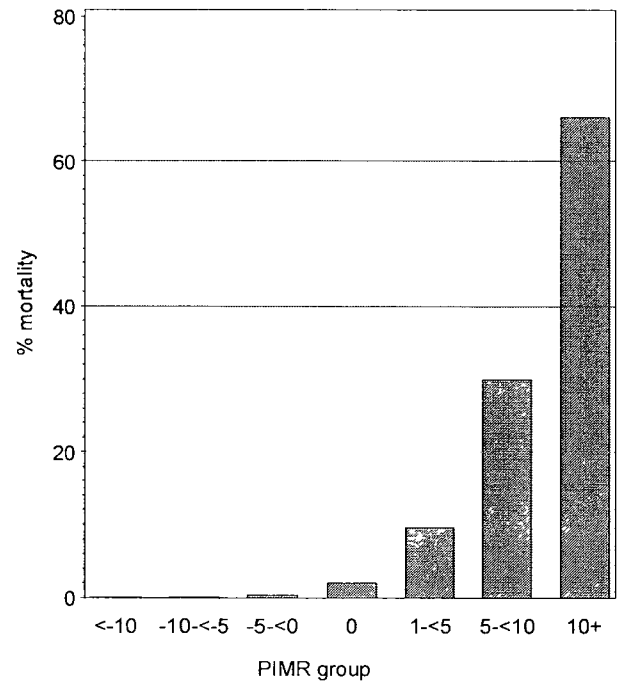


Figure 4.5 Mean LAPS by daily outcome status in the derivation set



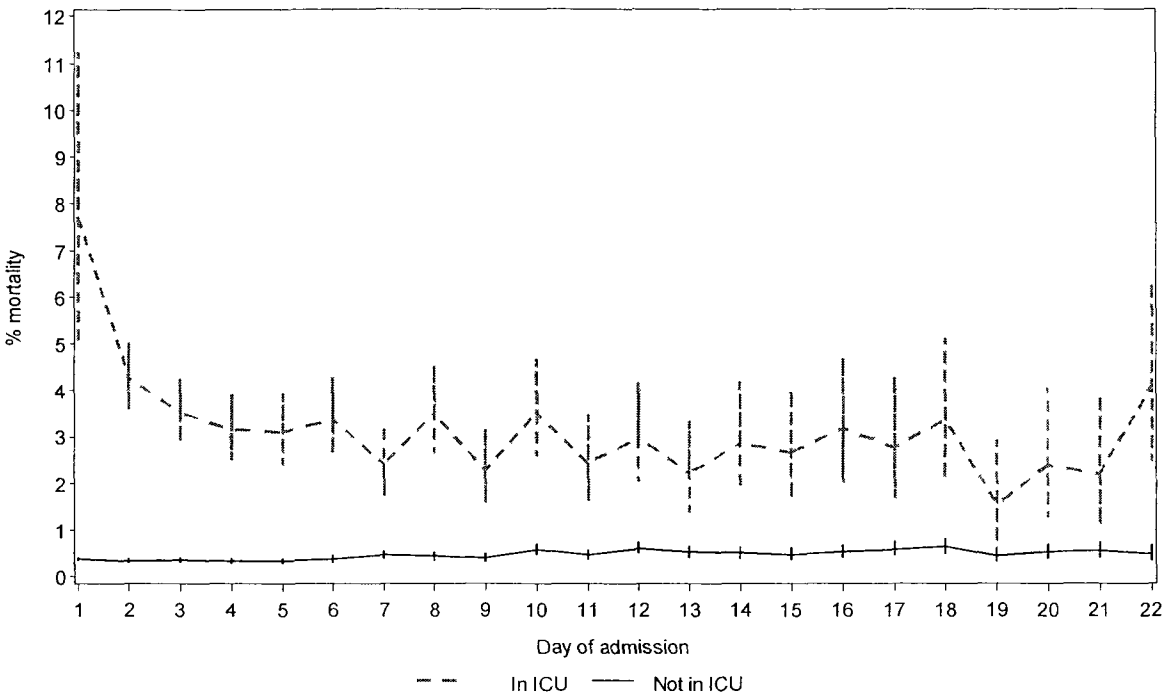
The mean LAPS is presented with the 95% confidence intervals (CI), where the confidence interval limits are calculated using the formula: mean ± (1.96*SE).

Figure 4.6 Proportion of deaths by total PIMR score in the derivation set



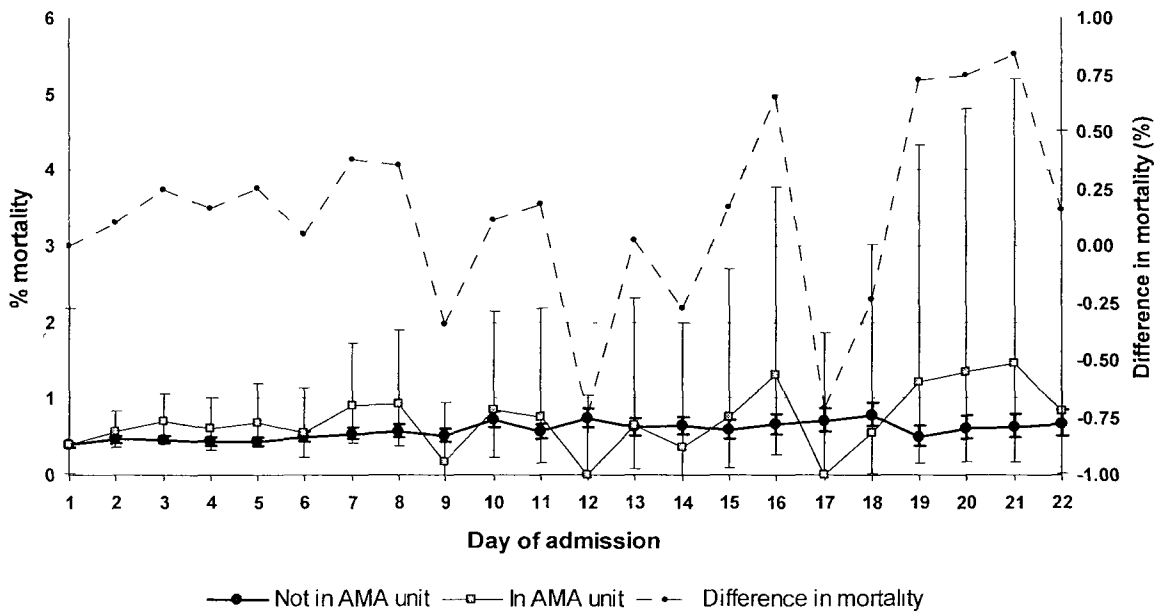
Only the first non-zero PIMR score was used if more than one PIMR episode occurred during the admission.

Figure 4.7 Daily proportion of deaths by daily ICU status in the derivation set



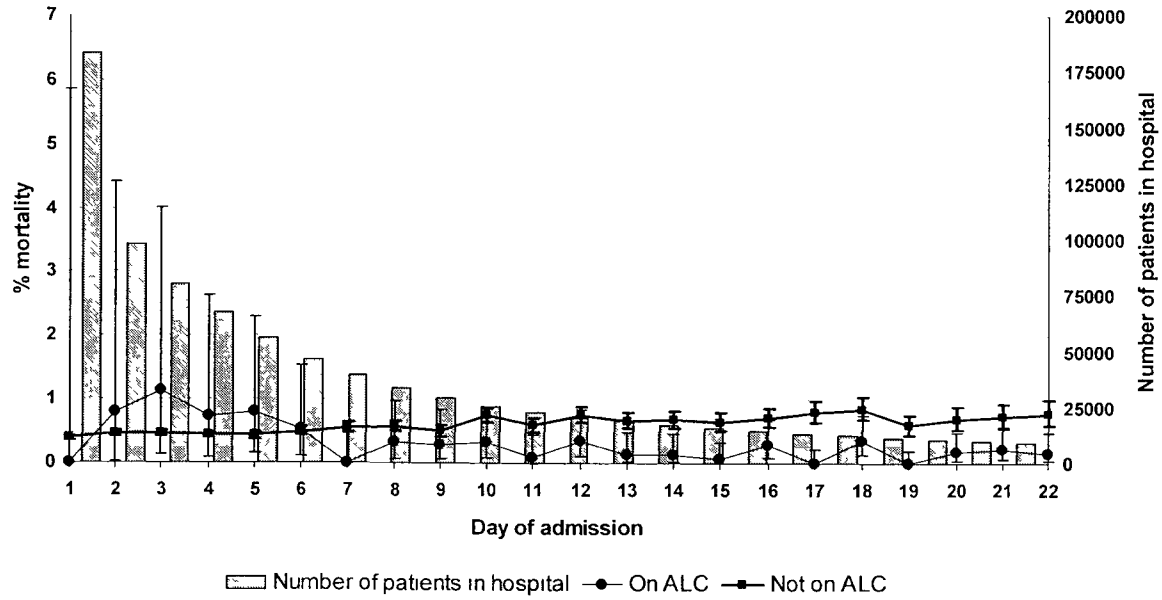
The observed mortality rates are presented with 95% confidence intervals calculated using exact methods (59).

Figure 4.8 Daily proportion of deaths by daily AMA status in the derivation set



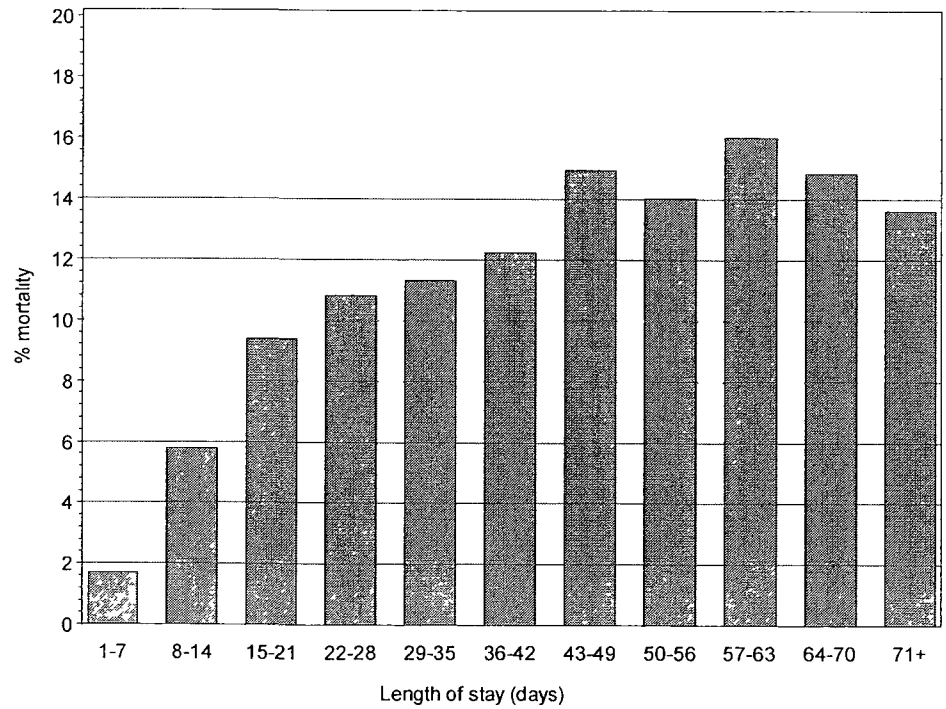
The observed mortality rate (95% confidence interval [CI]) among patients not in an AMA unit is represented by the black circles (bold line), while the rate among patients in an AMA unit is represented by the grey squares (thin line) (left vertical axis). The difference in mortality rate (right vertical axis) is represented by the dotted line and is calculated as the mortality rate among AMA patients subtracted by the rate among non-AMA patients. The 95% CIs for the observed mortality rates are calculated using exact methods (60).

Figure 4.9 Daily proportion of deaths by daily ALC status in the derivation set



The observed mortality rate (95% confidence interval [CI]) among patients on ALC is represented by the black circles (thin line), while the rate among patients not on ALC is represented by the black squares (bold line) (left vertical axis) The total number of patients in hospital on each day is presented by the bar graph (right vertical axis) The 95% CIs for the observed mortality rates are calculated using exact methods (61)

Figure 4.10 Proportion of deaths by length of admission in the derivation set



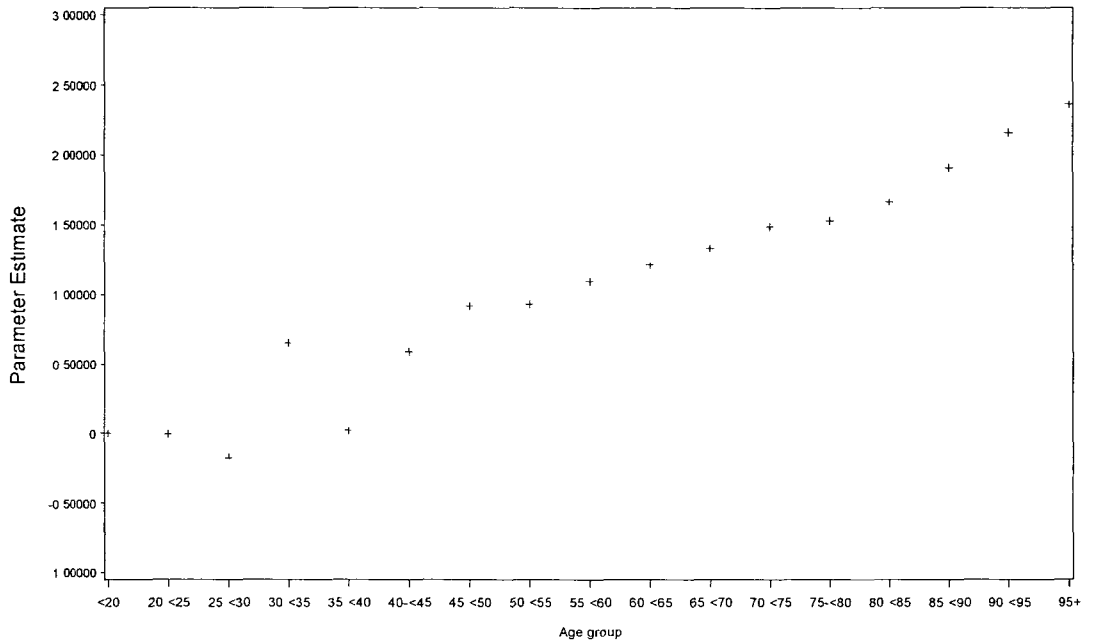
4.4.2 Functional form of continuous covariates

4.4.2.1 Age at admission

When I plotted the parameter estimates from a univariable Cox regression model with age at admission categorized into five-year age groups, the association between age and the log hazard did not appear linear, particularly in the younger age groups (Figure 4.11). The observed cumulative Martingale residual path (Figure 4.12) also deviated significantly from the simulated paths and the *p*-value of Kolmogorov-type supremum test was statistically significant (p=0.0030), thus suggesting some degree of covariate misspecification for linear age. However, the fractional polynomial function selection procedure chose the linear form as the best fit in the Cox model.

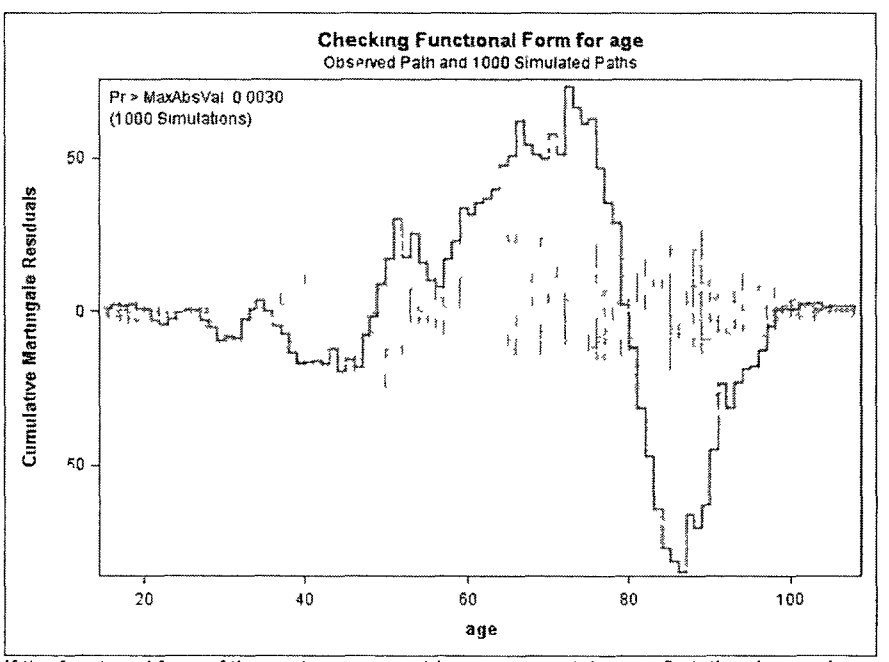
I compared the Cox model fit when age at admission was expressed as a linear term versus splined age^2. The AIC of the model with splined age^2 (54271.745) was lower than the model with linear age (54283.041), thus suggesting that splined age^2 was a better fit.

Figure 4.11 Unadjusted parameter estimates for categories of age at admission



A linear trend in the parameter estimates suggests a linear association between the continuous covariate and the log hazard The parameter estimates for categories of age at admission are not linear in the younger age groups (i e do not provide support for a linear association)

Figure 4.12 Cumulative Martingale residual plot for linear age at admission

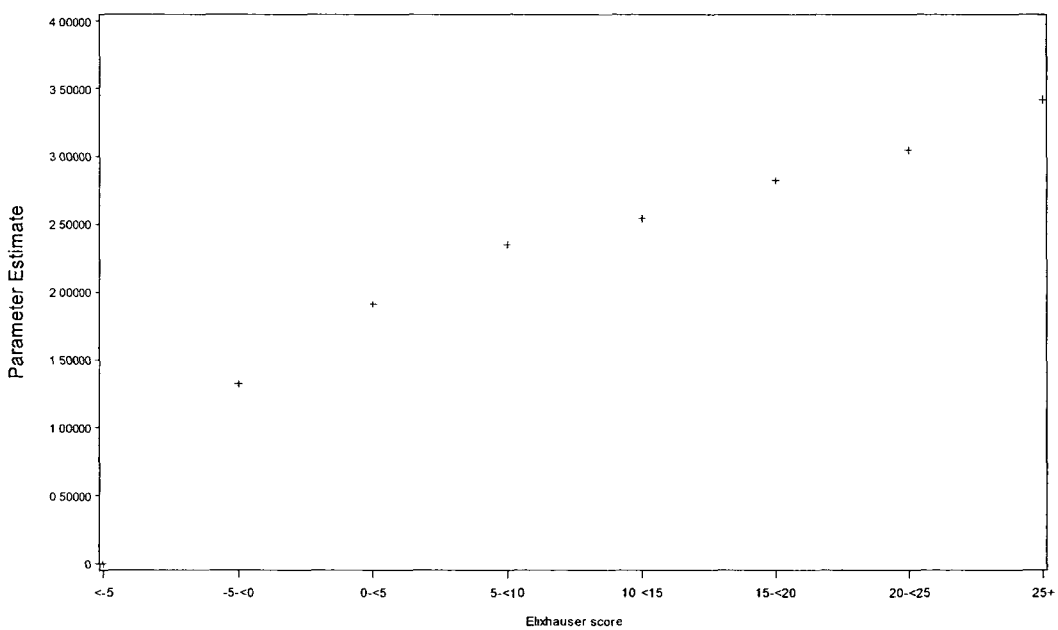


If the functional form of the continuous variable is appropriately specified the observed cumulative Martingale residual path (dark solid line) should remain within the range of the simulated residual paths (light dotted lines) and the *p*-value of the Kolmogorov-type supremum test (top left) should be non significant The plot above suggests that the linear form may be inappropriate for age since the observed path deviates from the simulated paths and the *p* value from the supremum test is significant

4.4.2.2 Elixhauser score

The parameter estimates from a univariable Cox model containing categories of the Elixhauser score suggested that the association between Elixhauser and the log hazard was fairly linear (Figure 4.13). However, the best form from the function selection procedure was $(\text{Elixhauser}+20)^{-0.5}$, where a value of 20 was added to each score because the theoretical minimum score was -19. The cumulative Martingale residual plots confirmed that this transformation was superior to the linear form, since the amount of deviation between the observed and simulated residuals paths was less for the transformation (Figure 4.14a) compared to the linear form (Figure 4.14b). The *p*-value of the Kolmogorov-type supremum test was also less significant (0.0020 for the transformation versus <.0001 for the linear form).

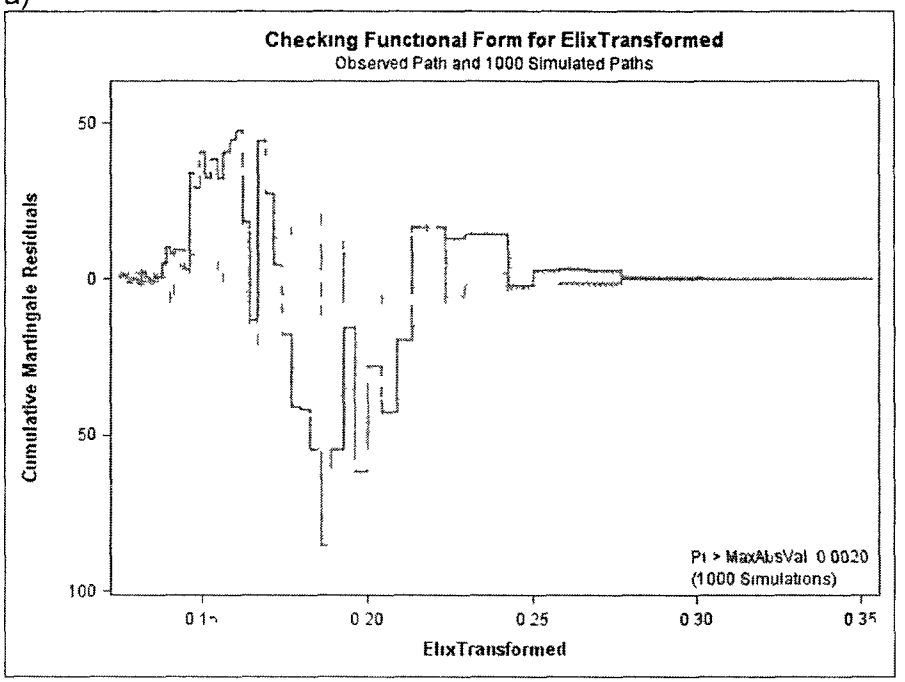
Figure 4.13 Unadjusted parameter estimates for categories of the Elixhauser score



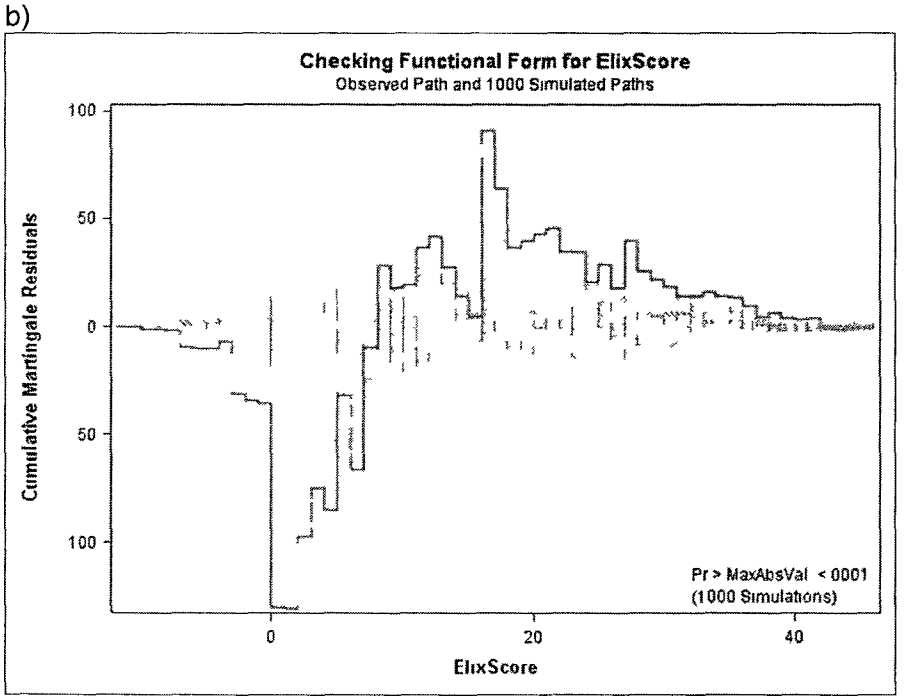
A linear trend in the parameter estimates suggests a linear association between the continuous covariate and the log hazard. The parameter estimates for the Elixhauser score form a fairly straight line, suggesting that the linear form may be appropriate

Figure 4.14 Cumulative Martingale residual plots for the Elixhauser score

a)



b)

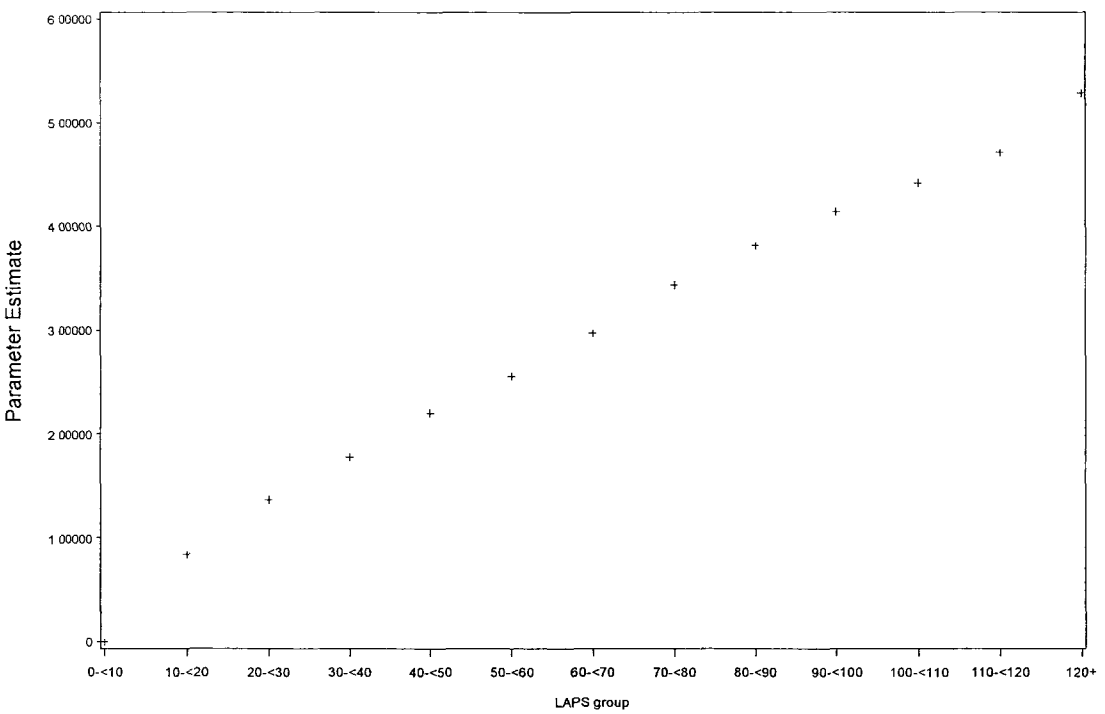


If the functional form of the continuous variable is appropriately specified the observed cumulative Martingale residual path (dark solid line) should remain within the range of the simulated residual paths (light dotted lines) and the *p*-value of the Kolmogorov type supremum test (bottom right) should be non-significant (Elixhauser+20)[05] (graph a) appears more appropriate than the linear form (graph b) because the observed path deviates less from the simulated paths and the *p*-value from the supremum test is larger (though still statistically significant)

4.4.2.3 LAPS

The unadjusted parameter estimates for categories of LAPS (expressed as a time-dependent covariate) strongly suggested that the association was linear (Figure 4.15). The best form from the function selection procedure, however, was $(LAPS+1)^{0.5}$, where a value of 1 was added to each score because the minimum value of LAPS was 0. I compared the fit of these two forms using the AIC because I could not produce cumulative Martingale residuals plots for a time-dependent model in SAS. The model with LAPS expressed as $(LAPS+1)^{0.5}$ had a lower AIC (48475.654) than the model with LAPS expressed as linear term (48546.437), confirming that the transformation was better fit.

Figure 4.15 Unadjusted parameter estimates for categories of LAPS



A linear trend in the parameter estimates suggests a linear association between the continuous covariate and the log hazard. The parameter estimates for LAPS form a very straight line, providing strong evidence that the linear form is appropriate.

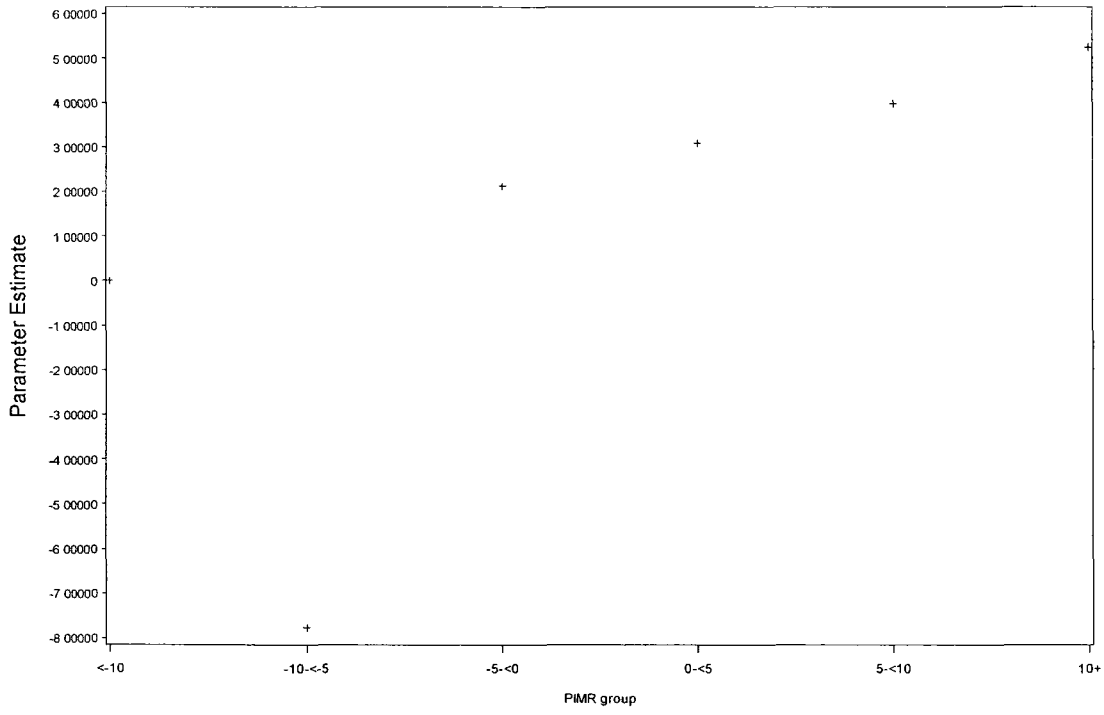
4.4.2.4 PIMR

The unadjusted parameter estimates for categories of PIMR suggested that the association between PIMR and the log hazard was not linear, particularly for scores less

than -5 (Figure 4.16). The best transformation from the function selection procedure was $(PIMR+51)^{-0.5}$, although the AIC for the transformation (54443.548) was only slightly lower than the AIC for the linear form (54450.345). A value of 51 was added to each score in the transformation since the theoretical minimum PIMR score was -50. I also compared the model fit when PIMR was expressed as a categorical variable (using the same cut-points as in Figure 4.16). The AIC for this model (54467.031) was higher than the AIC for the linear model, indicating that a categorical variable was a worse fit.

Next, I used the function selection procedure to determine the best functional form for "number of days since the last PIMR episode," adjusting for $(PIMR+51)^{-0.5}$ in all models. The linear form produced the model with the best fit.

Figure 4.16 Unadjusted parameter estimates for categories of PIMR



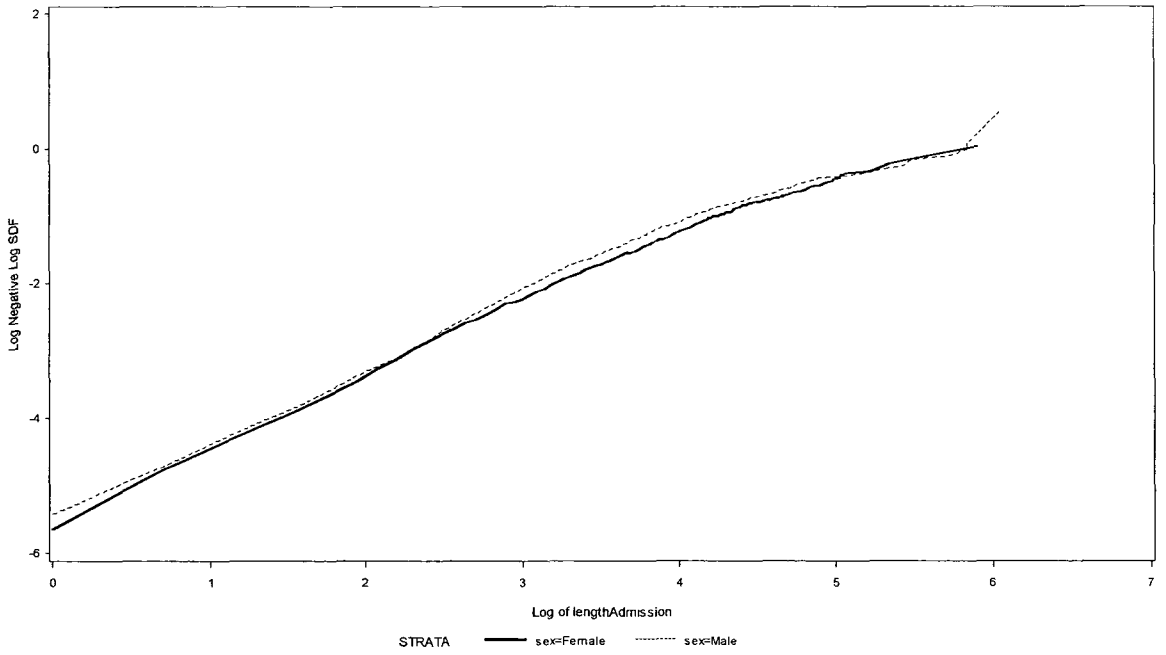
A linear trend in the parameter estimates suggests a linear association between the continuous covariate and the log hazard. The parameter estimates for categories of the PIMR score do not provide support for a linear association, particularly due to the -10-<-5 category.

4.4.3 Tests for non-proportionality

4.4.3.1 Sex

I found no evidence of non-proportionality for sex. There was no significant trend in the weighted Schoenfeld residuals over time ($p = 0.3184$), and the lines in the log-negative-log plot were fairly parallel (Figure 4.17). However, the log-negative-lot plot also suggested that sex had no effect on the risk of in-hospital death because there was little difference between the lines for males and females (the space between the lines represents the hazard difference between the strata).

Figure 4.17 Log-negative-log plot for sex



Length of stay (days)

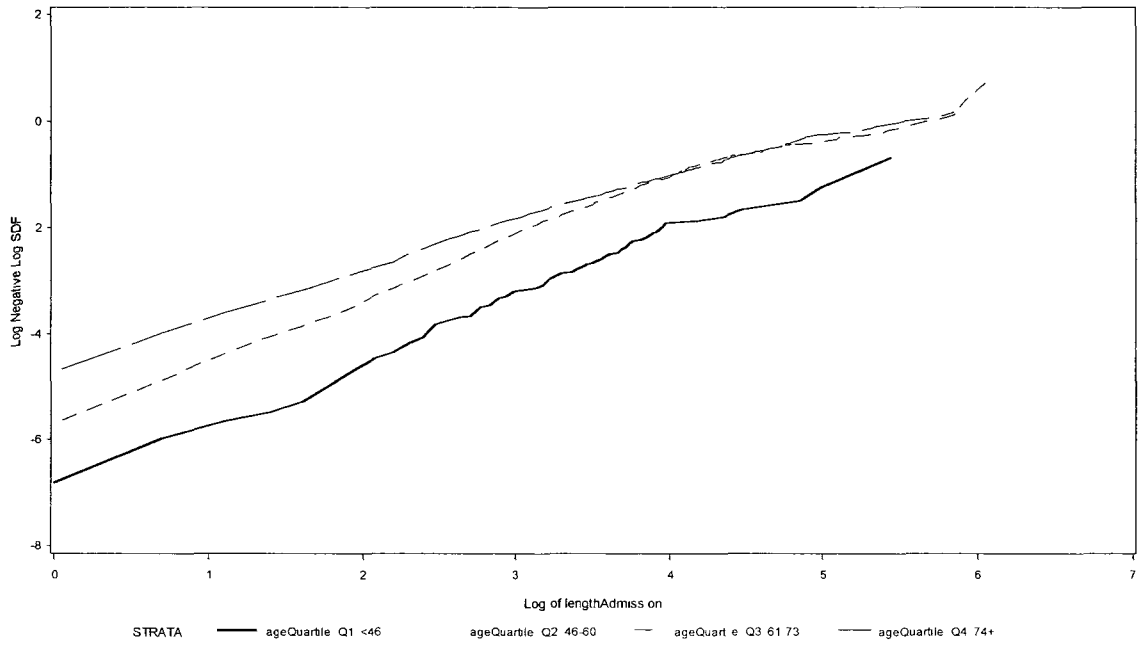
If the hazard difference between strata is proportional over time, the strata-specific log-negative-log plots should be parallel (with the space between the lines representing the hazard difference). Although the log-negative-log plots (above) are fairly parallel, there is little space between the lines, suggesting that the risk of in-hospital death is similar for males and females (i.e. sex is not a significant predictor of death in hospital).

4.4.3.2 Age at admission

I tested the assumption of proportionality for both age and splined age^2. For age, the log-negative-log plot (Figure 4.18) suggested that the hazard difference between patients in quartile three (61 to 73 years) and quartile four (74+ years) decreased over time. The lines for the other age quartiles were fairly parallel over time.

For splined age^2, there was no significant trend in the weighted Schoenfeld residuals over time for each of the spline terms (suggesting proportional hazards). On the contrary, when I plotted the observed standardized score process against 200 simulated paths (Figure 4.19a-c), the observed path for each spline term deviated significantly from the simulated paths and the *p*-value of the Supremum test was highly significant ($p<.0001$) for all terms, thus suggesting non-proportional hazards over time. Adding a time-dependent interaction term between each of the spline terms and either function of time (linear or logarithmic) significantly improved the model fit, also suggesting non-proportional hazards. The interaction with the logarithmic function of time was better because it produced the Cox model with the lowest AIC (Table 4.3).

Figure 4.18 Log-negative-log plot for quartiles of age at admission

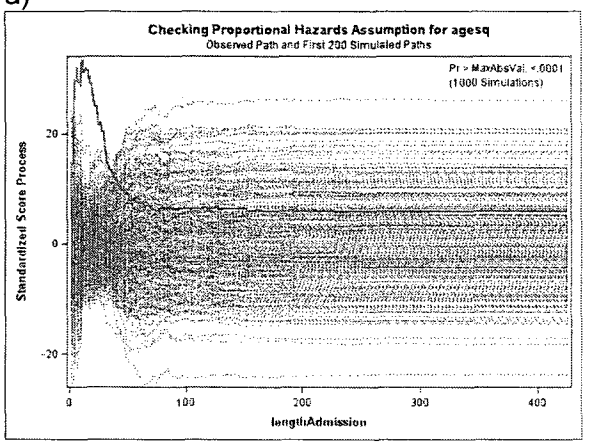


Length of stay (days)

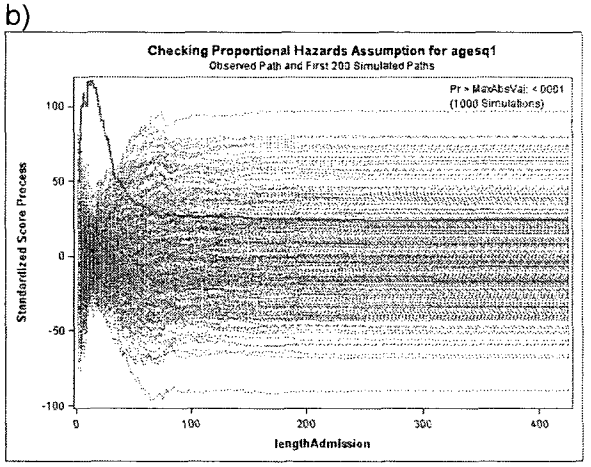
If the hazard difference between strata is proportional over time the strata-specific log negative-log plots should be parallel The log-negative-log plots for Q3 and Q4 converge over time suggesting that the hazard ratio between these age groups is not proportional (decreases) over time

Figure 4.19 Standardized score process plots for splined age^2 terms

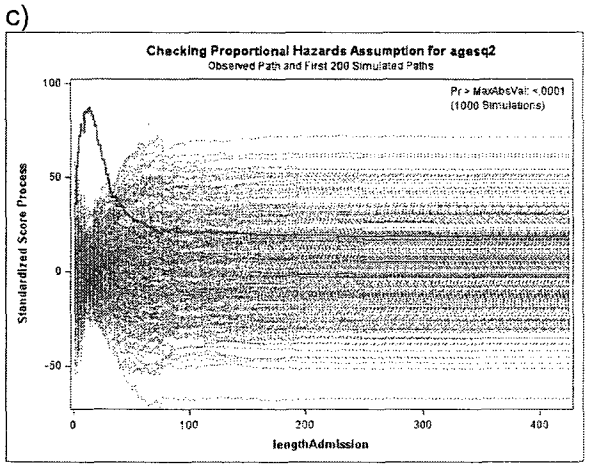
a)



b)



c)



If the proportional hazards assumption is satisified, the observed standardized score process (dark solid line) should remain within the range of the simulated residual paths (light dotted lines) and the *p*-value of the Supremum test (top right) should be non-significant. There is evidence of non-proportionality for all splined age^2 terms (graphs a-c) because the observed score process deviates from the simulated paths (particularly during the first 50 days of admission) and the *p*-value of the Supremum test is highly significant.

Table 4.3 Likelihood ratio test for interaction between splined age^2 and time

Model	Covariates	# covariates	-2 log L	*p*-value*	AIC
1	agesq + agesq1 + agesq2	3	54265.745	-	54271.745
2	agesq + agesq1 + agesq2 + agesq*time + agesq1*time + agesq2*time	6	54220.834	<.0001	54232.834
3	agesq + agesq1 + agesq2 + agesq*log(time) + agesq1* log(time) + agesq2* log(time)	6	54167.299	<.0001	54179.299

p-value from the likelihood ratio test comparing models 2 and 3 to model 1

4.4.3.3 Admission type

The log-negative-log plot for admission type provided strong evidence that the hazard difference between emergent (surgical/non-surgical) and elective (surgical/non-surgical) admissions was not proportional over time (Figure 4.20). However, within each admission urgency category (i.e. surgical versus non-surgical admissions), the hazard difference did appear fairly proportional over time, although there was little difference between elective surgical and non-surgical admissions (Figure 4.20).

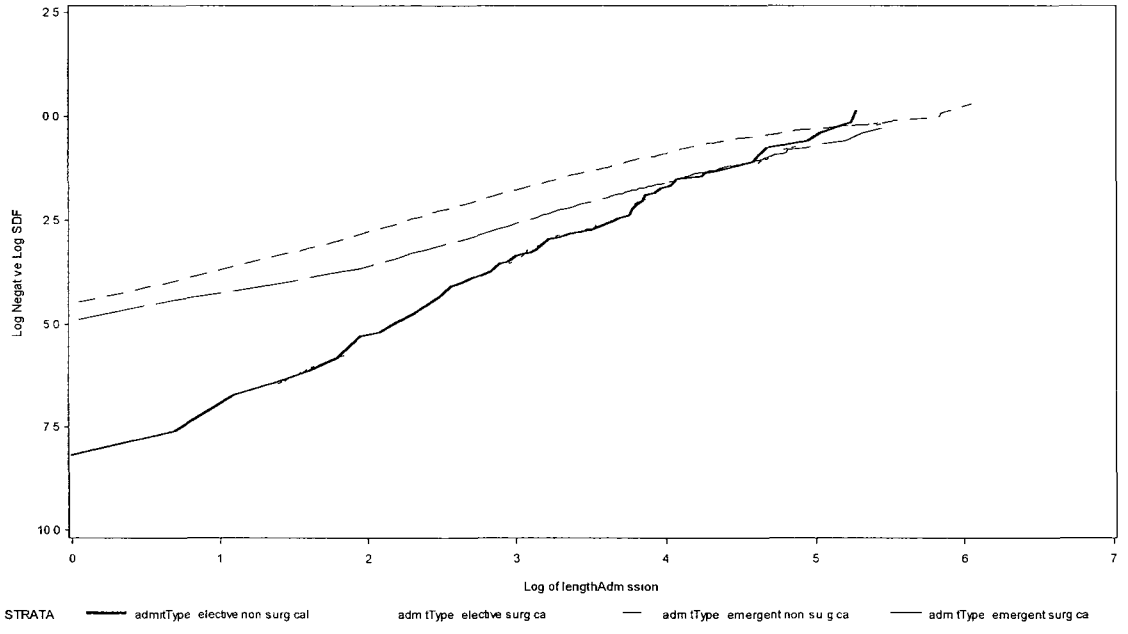
The weighted Schoenfeld residuals also suggested non-proportionality among admission types. Using elective surgical admissions as the reference type, there was a significant decreasing linear trend in the residuals over time for emergent surgical and emergent non-surgical admissions, but not for elective non-surgical admissions (Figure 4.21a-c).

When I added a time-dependent interaction term between each admission type and either function of time to the Cox model, the model fit significantly improved, once again suggesting non-proportional hazards. The interaction with the logarithmic function of time was better because it produced the Cox model with the lowest AIC (Table 4.4).

However, when I plotted the observed standardized score process against 200 simulated paths for three admission types (using elective surgical admissions as the reference in all plots), there was no apparent deviation between the observed and

simulated paths in all plots (suggesting proportional hazards) The *p*-value of the Supremum test was also non-significant for emergent and elective non-surgical admissions (suggesting proportional hazards), although the *p*-value was significant for emergent surgical admissions (*p*=0 0010) (Figure 4 22a-c)

Figure 4.20 Log-negative-log plot for admission type

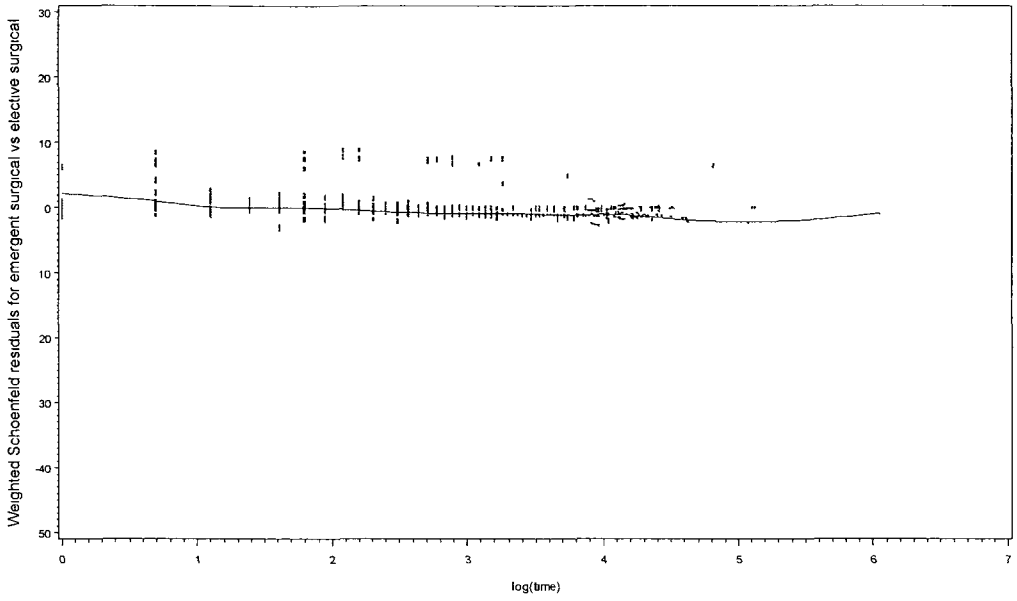


Length of stay (days)

If the hazard difference between strata is proportional over time the strata-specific log-negative-log plots should be parallel The log negative-lot plots for emergent (surgical/non surgical) and elective (surgical/non-surgical) admissions converge over time suggesting that the hazard ratio between emergent and elective admissions decreases over time

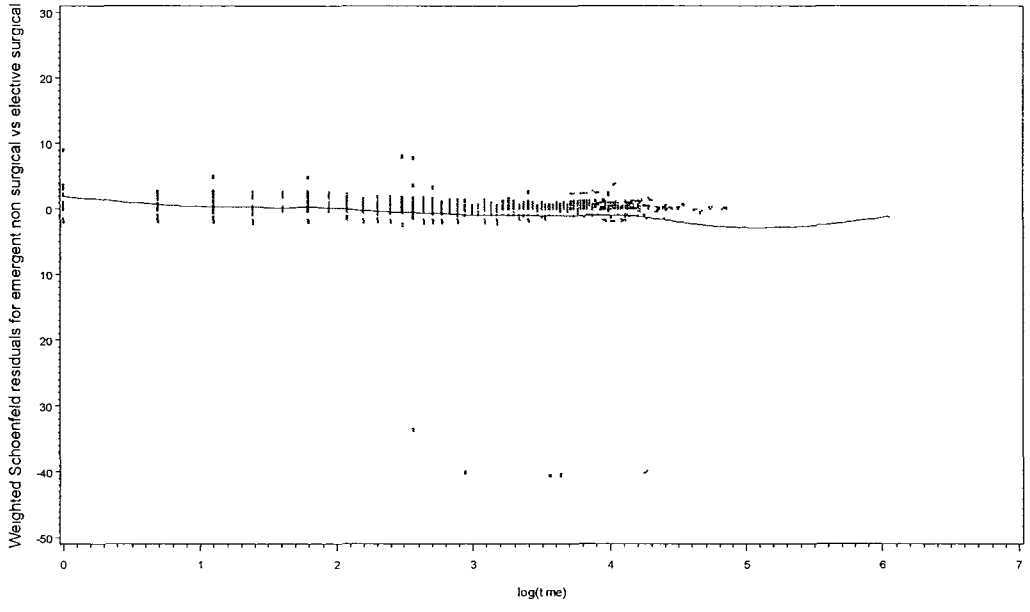
Figure 4.21 Weighted Schoenfeld residuals plots for admission type

a)



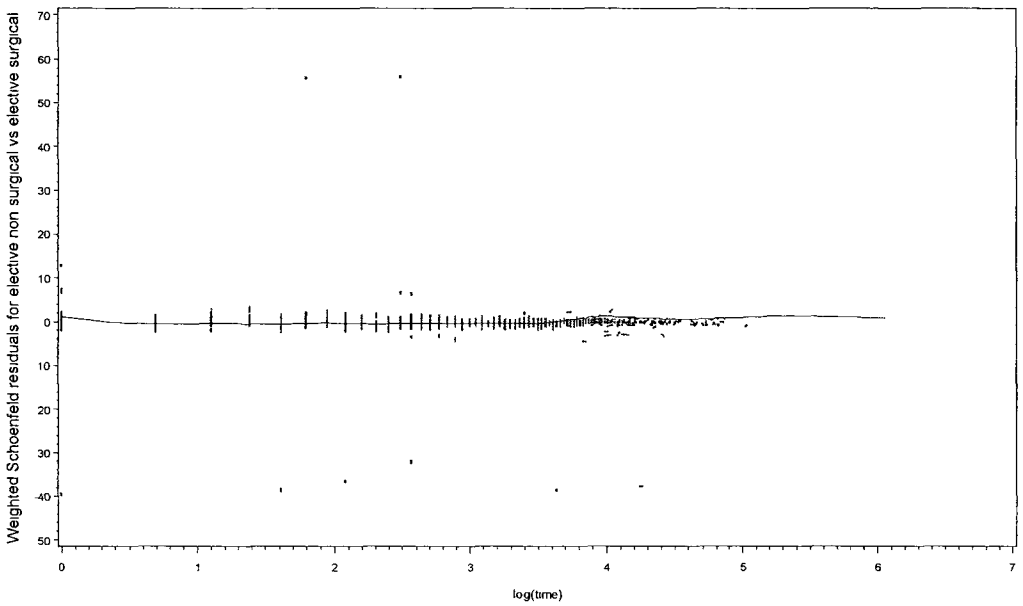
If the proportional hazards assumption is satisfied the value of the weighted Schoenfeld residuals should be random over time (i e no significant trend in the residuals over time) For emergent versus elective surgical admissions (above) there is a significant decreasing trend in the residuals over time (p< 0001) suggesting that the hazard difference between these admission types decreases over time

b)



For emergent non-surgical versus elective surgical admissions (above), there is a significant decreasing trend in the residuals over time (p< 0001) suggesting that the hazard difference between these admission types decreases over time

c)



For elective non-surgical versus surgical admissions (above) there is no significant trend in the residuals over time (p=0 7147) suggesting that the hazard difference between these admission types is proportional over time

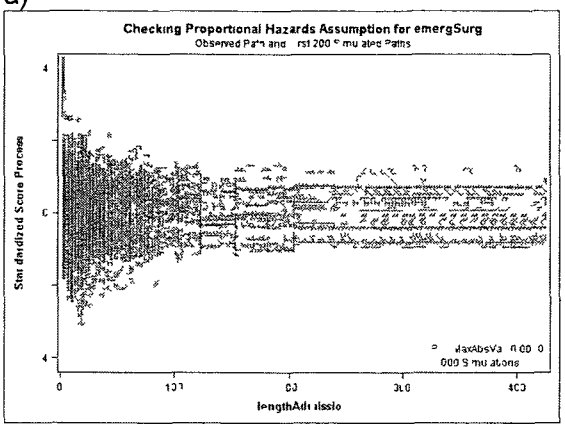
Table 4.4 Likelihood ratio test for interaction between admission type and time

Model	Covariates	# covariates	-2 log L	p-value*	AIC
1	emergSurg + emergNonsurg + electNonsurg	3	53936.384	-	53942 384
2	emergSurg + emergNonsurg + electNonsurg + emergSurg*time + emergNonsurg*time + electNonsurg*time	6	53828 156	<.0001	53840 156
3	emergSurg + emergNonsurg + electNonsurg + emergSurg*log(time) + emergNonsurg*log(time) + electNonsurg*log(time)	6	53754.628	<.0001	53766 628

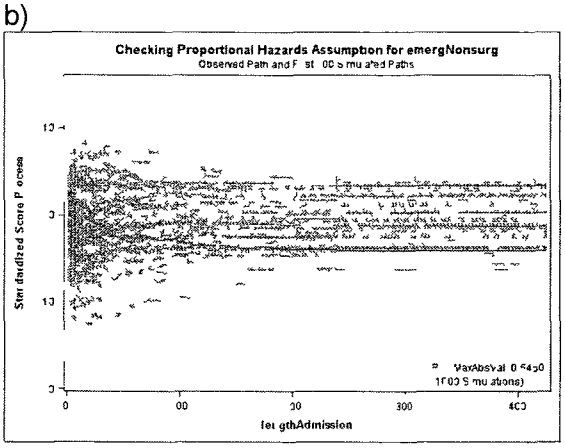
*p-value from the likelihood ratio test comparing models 2 and 3 to model 1

Figure 4.22 Standardized score process plots for admission type

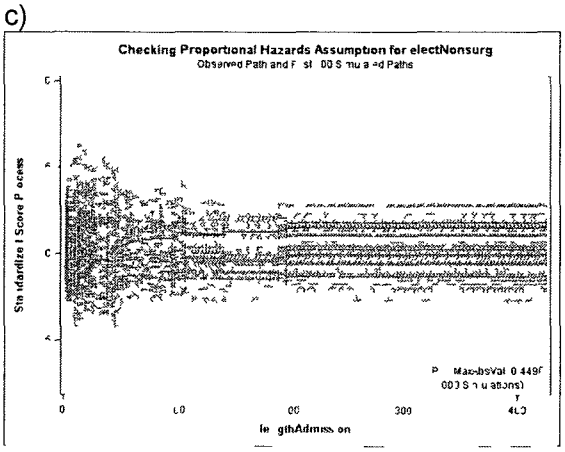
a)



b)



c)



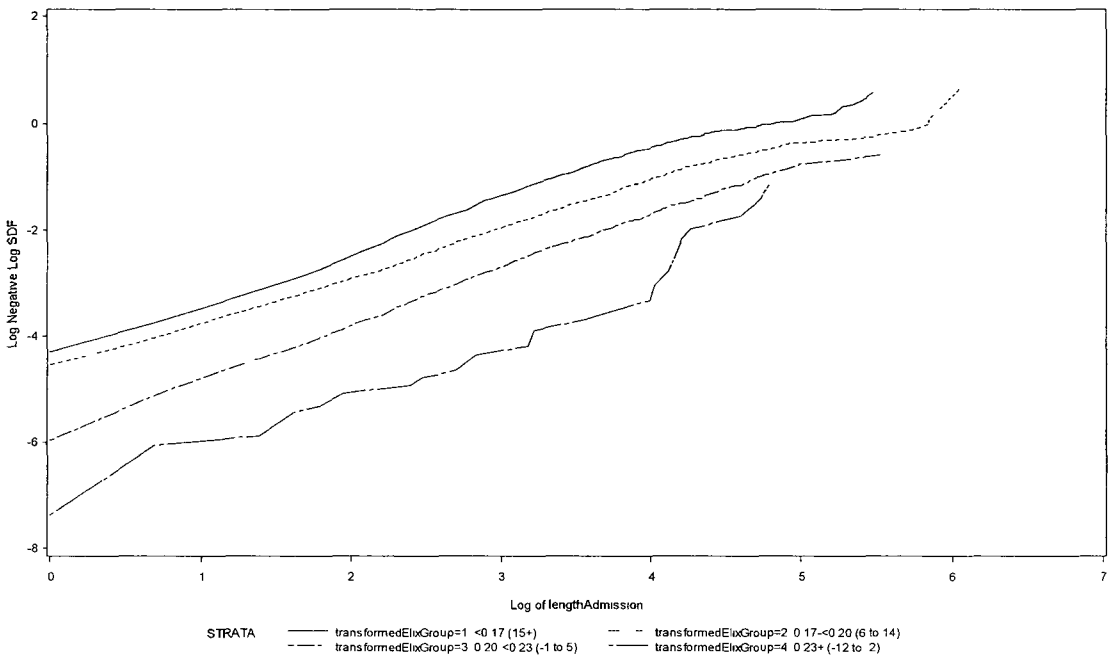
If the proportional hazards assumption is satisfied the observed standardized score process (dark solid line) should remain within the range of the simulated residual paths (light dotted lines) and the *p*-value of the Supremum test (bottom right) should be non-significant The proportional hazards assumption appears satisfied for emergent and elective non surgical admissions (graphs b and c) For emergent surgical admissions (graph a) the observed and simulated residual paths appear similar, but the *p* value of the Supremum test is significant

4.4.3.4 Elixhauser score

I conducted all tests for non-proportionality using the transformation of the Elixhauser score [i.e. $(\text{Elixhauser}+20)^{-0.5}$]. To create the log-negative-log plot, I could not use quartile cut-points because the distribution of scores was highly skewed. I therefore used equidistant cut-points to create four categories (each category was sufficiently large with at least 3000 admissions per category). The lines between each of the Elixhauser categories in the log-negative-log plot were fairly parallel, suggesting that the hazard difference between the categories was proportional over time (Figure 4.23).

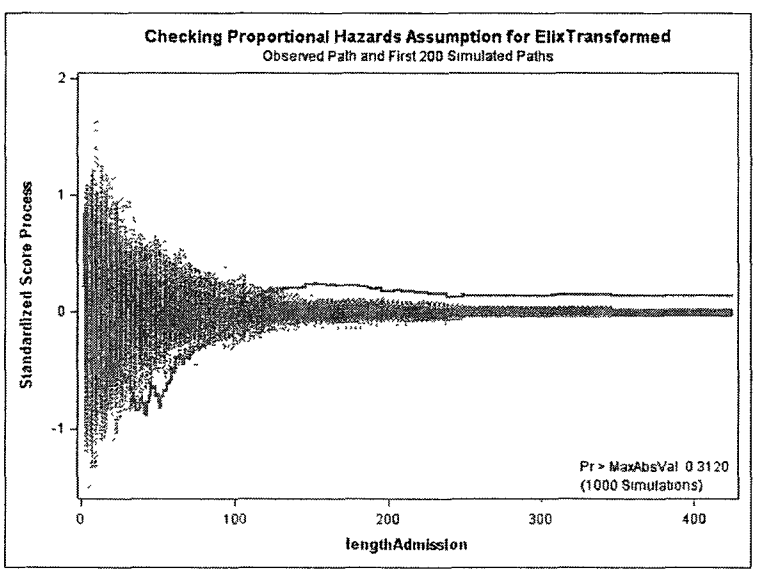
The other analytical methods also suggested proportional hazards. There was no significant linear trend in the weighted Schoenfeld residuals over time (p=0.7643), the observed standardized score process did not deviate significantly from the simulated paths (Figure 4.24), and the p-value from the Supremum test was not statistically significant (p=0.3120). The addition of time-dependent interaction terms with either function of time (linear or logarithmic) also did not significantly improve the Cox model fit.

Figure 4.23 Log-negative-log plot for categories of transformed Elixhauser



In the plot above, the Elixhauser score is expressed as $(\text{Elixhauser}+20)^{-0.5}$ (the best functional form for the Elixhauser score in a univariable Cox model according to the fractional polynomial function selection procedure). The bracketed values in the legend represent the range of original (untransformed) Elixhauser scores in each category. The lines for the different Elixhauser categories are fairly parallel, suggesting that the hazard difference between the categories is proportional over time.

Figure 4.24 Standardized score process plot for transformed Elixhauser



The plot above is for the Elixhauser score expressed as $(\text{Elixhauser}+20)^{-0.5}$. The proportional hazard assumption appears satisified for the Elixhauser score because the observed standardized scores process (dark solid line) is fairly similar to the simulated paths (lighter dotted lines), and the *p*-value of the Supremum test (bottom right) is not statistically significant.

4.5 Derivation of Escobar[+]

4.5.1 Preliminary main effects model

I entered all candidate main effects into an initial multivariable model. I expressed age as splined age^2, Elixhauser as (Elixhauser+20)$^{-0.5}$, LAPS as (LAPS+1)$^{0.5}$, PIMR as (PIMR+51)$^{-0.5}$, and "number of days since the last PIMR episode" as a linear term. All main effects were highly significant (p<.0001) except AMA (p=0.8161) and sex (p=0.4216).

I first removed AMA from the model because it had the largest p-value in the initial model. The likelihood ratio test comparing the fit of the multivariable model with and without AMA confirmed that AMA was not a significant predictor (p=0.816). AMA was also not an important confounder because there was little change (<2%) in the other coefficients when AMA was removed.

I next removed sex from the model. The likelihood ratio test comparing the fit of the model with and without sex confirmed that it was not a significant predictor (p=0.422). The other coefficients remained virtually unchanged when sex was removed (<0.6% change), indicating that sex was not an important confounder. When I added AMA back into the model with sex removed, AMA was still not a significant predictor or confounder.

The eight remaining main effects were highly significant. I then used the fractional polynomial function selection procedure to determine the best functional form for each continuous variable in a *multivariable* model containing all significant main effects. The only exception was age, which I kept fixed (as splined age^2 expressed as three non-linear terms) in all models during the function selection procedure to preserve the spline. The multivariable model with the lowest deviance was the one that included Elixhauser, LAPS, and PIMR expressed as linear terms, and "days since the last PIMR episode" expressed as ("days"+1)$^{-1}$. Interestingly, these forms were different (and simpler) than the best forms recommended by the function selection procedure in the univariable analysis.

The final preliminary main effects model contained eight main effects expressed as 12 covariates (Table 4.5). Increased values of splined age^2, Elixhauser, LAPS, and PIMR, and being in the ICU were all associated with an increased hazard of death. Emergent (surgical or non-surgical) admissions were also associated with a higher risk of death, as compared to elective surgical admissions. In contrast, being on ALC and an increasing number of "days since the last PIMR episode" were associated with a decreased hazard of death.

Table 4.5 Preliminary main effects model

Main effect	Parameter estimate	Standard error	*p*-value	Hazard ratio
Agesq	0.0003261	0.0001		1.00
Agesq1	-0.0005971	0.0002	<.0001*	1.00
Agesq2	0.0016465	0.0004		1.00
Emergent surgical	1.0415254	0.0937		2.83
Emergent non-surgical	1.3858122	0.0887	<.0001*	4.00
Elective non-surgical	0.0417985	0.1343		1.04
Elixhauser	0.0148616	0.0020	<.0001	1.01
LAPS	0.0304474	0.0005	<.0001	1.03
PIMR	0.0734309	0.0074	<.0001	1.08
(Days since the last PIMR episode+1)$^{-1}$	0.3786332	0.0521	<.0001	1.46
ICU	0.7082700	0.0457	<.0001	2.03
ALC	-1.5243525	0.0945	<.0001	0.22

*p-value from the likelihood ratio test when the covariates were removed collectively from the model

4.5.2 Interactions between main effects

I considered a total of nine interaction terms: splined age^2*LAPS, splined age^2*Elixhauser, Elixhauser*LAPS, Elixhauser*PIMR, Elixhauser*ICU, PIMR*(Days since the last PIMR episode + 1)$^{-1}$, splined age^2*PIMR, LAPS*PIMR, and LAPS*ICU. The first three interaction terms were included in the original Escobar model, and the remaining six were identified through consultation with Drs. van Walraven and Forster. Because each

interaction involving splined age^2 added three covariates to the model, I tested the significance of interactions with splined age^2 using the likelihood ratio test.

All interactions were significant when added separately to the preliminary main effects model. When these interactions were added jointly to the main effects model, three interactions were no longer significant: Elixhauser*PIMR (p=0.3615), splined age^2*LAPS (p=0.0701), and Elixhauser*ICU (p=0.0506). When I removed Elixhauser*PIMR from the model, Elixhauser*ICU became significant (p=0.0118), but splined age^2*LAPS was still not significant (p=0.0736). When I removed splined age^2*LAPS from the model, all seven of the remaining interaction terms were significant (Table 4.6).

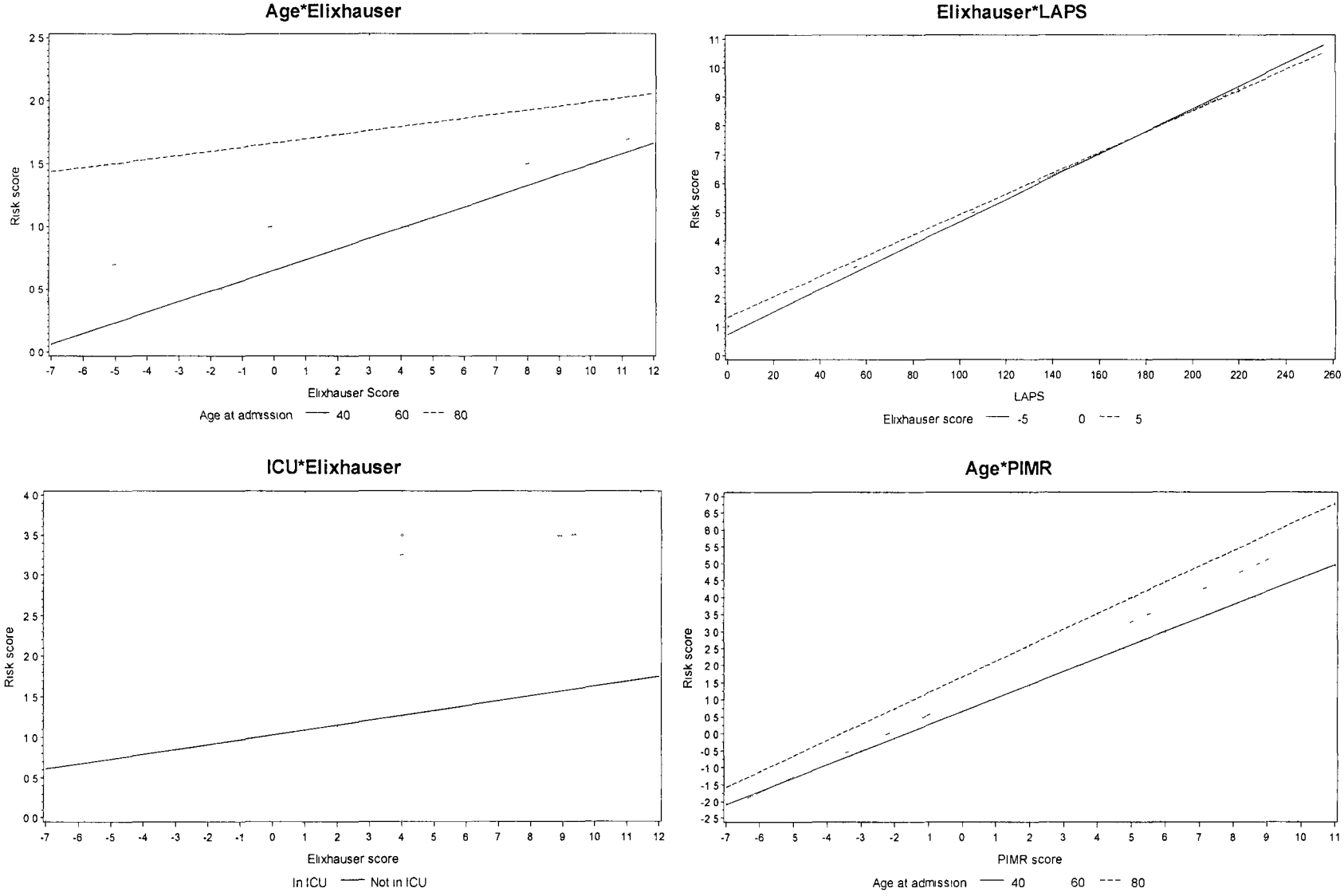
For each of the interaction terms, I determined the nature of the relationship between the two covariates by graphing the model-based predicted risk (i.e. the linear predictor) from the model when the value of one covariate was allowed to vary by different values of the other covariate (as described in Section 3.8.4.4.2). These "risk score plots" (Figure 4.25) revealed the following: the effect of Elixhauser decreased with increasing age; the effect of LAPS decreased slightly with increasing Elixhauser score; the effect of Elixhauser decreased slightly if the patient was in the ICU; the effect of the PIMR score increased with increasing age; the effect of the PIMR score decreased slightly with increasing LAPS; and the effect of LAPS decreased quite drastically if the patient was in the ICU. The interaction between the PIMR score and "days since the last PIMR episode" was particularly interesting. For higher-risk procedures (i.e. PIMR > 0), the risk of in-hospital death decreased as the number of days since the last PIMR episode increased. However, for low-risk procedures (i.e. PIMR < 0), the risk of in-hospital death *increased* as the number of days since the last PIMR episode increased. The plot also revealed that the predicted effect of "days since the last PIMR episode" was most prominent during the first five days after the episode and was minimal thereafter.

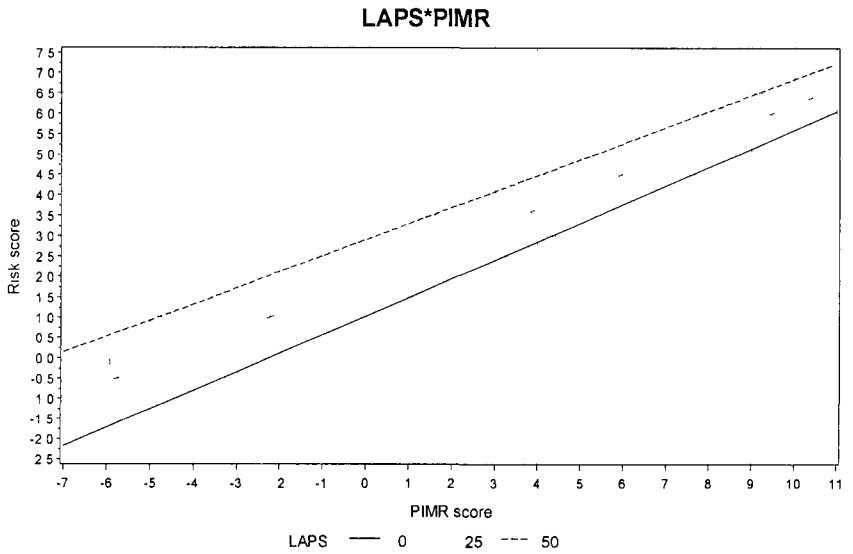
Table 4.6 Model with main effects and effect interactions

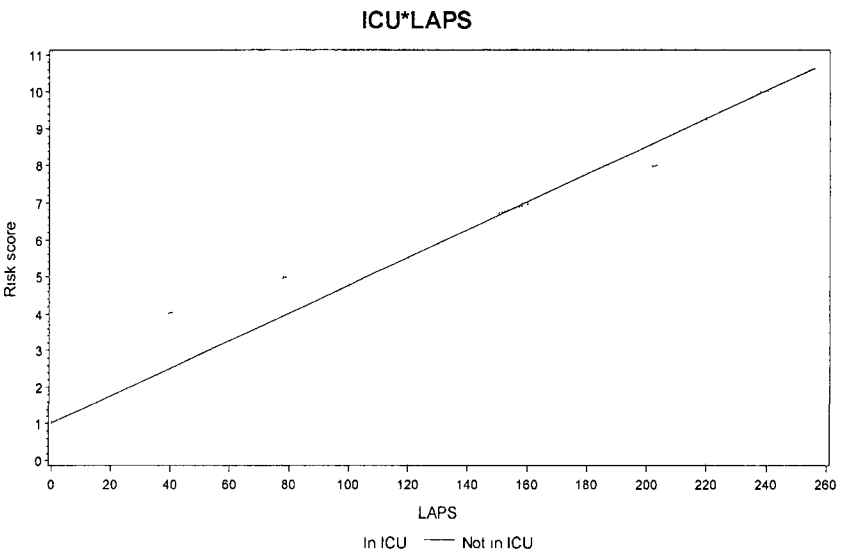
Covariate	Parameter estimate	Standard error	*p*-value	Hazard ratio
Agesq	0.0001992	0.0001	0.0381	-
Agesq1	-0.0000874	0.0003	0.7276	-
Agesq2	0.0004273	0.0006	0.4887	-
Emergent surgical	0.9640671	0.0937		2.62
Emergent non-surgical	1.2944270	0.0889	<.0001*	3.65
Elective non-surgical	0.1296946	0.1338		1.14
Elixhauser	0.0991374	0.0152	<.0001	-
LAPS	0.0375391	0.0008	<.0001	-
PIMR	-0.0408630	0.0579	0.4800	-
(Days since last PIMR episode+1)$^{-1}$	0.3389941	0.0547	<.0001	-
ICU	2.0359055	0.1292	<.0001	-
ALC	-1.4996791	0.0946	<.0001	0.22
Agesq*Elixhauser	-0.0000093	0.0000		-
Agesq1*Elixhauser	-0.0000078	0.0000	<.0001*	-
Agesq2*Elixhauser	0.0000263	0.0000		-
Elixhauser*LAPS	-0.0003365	0.0001	<.0001	-
Elixhauser*ICU	-0.0125269	0.0049	0.0100	-
PIMR*(Days since last PIMR episode+1)$^{-1}$	0.3520850	0.0330	<.0001	-
Agesq*PIMR	0.0000512	0.0000		-
Agesq1*PIMR	-0.0000699	0.0001	0.0033*	-
Agesq2*PIMR	0.0001187	0.0002		-
LAPS*PIMR	-0.0012645	0.0002	<.0001	-
LAPS*ICU	-0.0132014	0.0013	<.0001	-

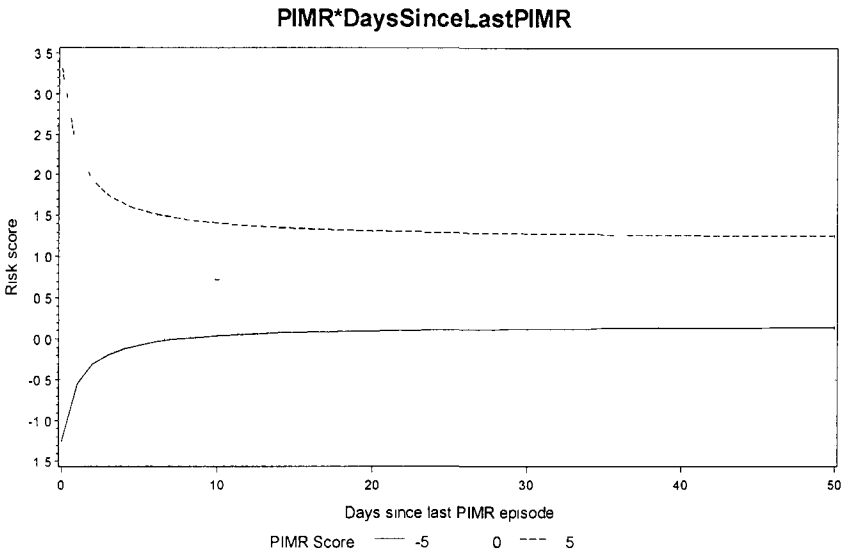
*p-value from the likelihood ratio test when the covariates were removed collectively from the model

Figure 4.25 Risk score plots for interactions between main effects









The "risk score" (or linear predictor) is equal to the sum of each coefficient in the model multiplied by the corresponding covariate value To calculate the risk scores, I used the following covariate values if the covariate was not involved in the interaction age = 61 years, admission type = elective surgical, Elixhauser = 0, LAPS = 0, PIMR = 0, and days since last PIMR episode = 0

4.5.3 Tests for non-proportionality among time-independent covariates

Of the eight main effects and seven interaction terms in the model (Table 4.6), only three main effects (splined age^2, admission type, and Elixhauser score) and one interaction term (splined age^2*Elixhauser) were time-independent covariates.

The weighted Schoenfeld residuals for splined age^2, elective non-surgical admissions (compared to elective surgical admissions) and splined age^2*Elixhauser suggested that the proportional hazards assumption was satisified for these covariates because there was no significant trend over time in the residuals. On the other hand, there was a significant decreasing trend in the residuals for emergent surgical (p<.0001) and emergent non-surgical admissions (p = 0.0381), suggesting that the difference in hazard (compared to elective surgical admissions) decreased over time. For the Elixhauser score, there was a significant increasing trend in the residuals (p = 0.0354), suggesting that the effect of a higher Elixhauser score increased over time.

Next, I tested the significance of adding an interaction term between each time-independent main effect (splined age^2, admission type, and the Elixhauser score) and time to the model. I first added these interactions to the model separately and used the AIC to determine which function of time produced the Cox model with the best fit (i.e. the model with the lower AIC). I found that an interaction with the linear function of time was best for splined age^2, and an interaction with the logarithmic function of time was best for admission type and the Elixhauser score (Table 4.7). I then added all of these interactions jointly to the Cox model (because they were each significant separately) and used the likelihood ratio test to test the significance of each interaction in the "full" model (i.e. the model with all interactions added jointly). I did this by comparing the fit of the full model with the model fit when each interaction was removed in turn (Table 4.7). I found that an interaction with time was statistically significant for all three time-independent covariates.

The final model ("*Escobar*[+]") contained eight main effects, seven interactions between main effects, and three interactions with time, expressed as 30 covariates (Table 4.8).

Table 4.7 Likelihood ratio test for interactions between time-independent main effects and time

Model	Covariates	# covariates	AIC	-2 log L	*p*-value
	INTERACTIONS ADDED SEPARATELY				
1	[Main effects + effect interactions*]	23	46337.514	46291.514	-
2	[Main effects + effect interactions] + agesq*time + agesq1*time + agesq2*time	26	46332.622	46280.622	0.0123[†]
3	[Main effects + effect interactions] + agesq*log(time) + agesq1*log(time) + agesq2*log(time)	26	46336.182	46284.182	-
4	[Main effects + effect interactions] + emergSurg*time + emergNonsurg*time + electNonsurg*time	26	46313.668	46261.668	-
5	[Main effects + effect interactions] + emergSurg*log(time) + emergNonsurg*log(time) + electNonsurg*log(time)	26	46260.981	46208.981	<.0001[†]
6	[Main effects + effect interactions] + Elixhauser*time	24	46333.727	46285.727	-
7	[Main effects + effect interactions] + Elixhauser*log(time)	24	46299.441	46251.441	<.0001[†]
	INTERACTIONS ADDED JOINTLY				
8	[Main effects + effect interactions] + agesq*time + agesq1*time + agesq2*time + emergSurg*log(time) + emergNonsurg*log(time) + electNonsurg*log(time) + Elixhauser*log(time)	30	46223.057	46163.057	-
9	[Main effects + effect interactions] + emergSurg*log(time) + emergNonsurg*log(time) + electNonsurg*log(time) + Elixhauser*log(time)	27	46227.422	46173.422	0.0157[‡]
10	[Main effects + effect interactions] + agesq*time + agesq1*time + agesq2*time + Elixhauser*log(time)	27	46294.867	46240.867	<.0001[‡]
11	[Main effects + effect interactions] + agesq*time + agesq1*time + agesq2*time + emergSurg*log(time) + emergNonsurg*log(time) + electNonsurg*log(time)	29	46256.355	46198.355	<.0001[‡]

*Covariates from the model in Table 4.6
[†]*p*-value from the likelihood ratio test comparing the model to model 1
[‡]*p*-value from the likelihood ratio test comparing the model to model 8
Note: The *p*-value for models 9, 10, and 11 are testing the significance of adding an interaction between time and splined age[2], admission type, and the Elixhauser score to the model, respectively.

Table 4.8 Model with main effects, effect interactions, and interactions with time

Covariate	Parameter estimate	Standard error	*p*-value
Main effects			
Agesq	0.0002153	0.0001	0.0371
Agesq1	-0.0001692	0.0003	0.5274
Agesq2	0.0006240	0.0007	0.3422
Emergent surgical	1.8457704	0.1990	<.0001
Emergent non-surgical	1.5801596	0.1903	<.0001
Elective non-surgical	-0.2160286	0.3496	0.5366
Elixhauser	0.0844500	0.0155	<.0001
LAPS	0.0382106	0.0008	<.0001
PIMR	-0.0369053	0.0576	0.5220
(Days since last PIMR episode+1)$^{-1}$	0.3165349	0.0546	<.0001
ICU	2.0569861	0.1294	<.0001
ALC	-1.5467522	0.0969	<.0001
Effect interactions			
Agesq*Elixhauser	-0.0000099	0.0000	
Agesq1*Elixhauser	-0.0000093	0.0000	<.0001*
Agesq2*Elixhauser	0.0000322	0.0000	
Elixhauser*LAPS	-0.0003561	0.0001	<.0001
Elixhauser*ICU	-0.0122942	0.0049	0.0115
PIMR*(Days since last PIMR episode+1)$^{-1}$	0.3354822	0.0333	<.0001
Agesq*PIMR	0.0000493	0.0000	
Agesq1*PIMR	-0.0000650	0.0001	0.0027*
Agesq2*PIMR	0.0001079	0.0002	
LAPS*PIMR	-0.0012505	0.0002	<.0001
LAPS*ICU	-0.0136417	0.0013	<.0001
Interactions with time			
Agesq*time	-0.0000008	0.0000	
Agesq1*time	0.0000070	0.0000	0.0157*
Agesq2*time	-0.0000190	0.0000	
Emerg surg*log(time)	-0.3840285	0.0752	
Emerg nonsurg*log(time)	-0.1052175	0.0715	<.0001
Elect nonsurg*log(time)	0.1367376	0.1212	
Elixhauser*log(time)	0.0094108	0.0016	<.0001

*p-value from the likelihood ratio test when the covariates were removed collectively from the model

4.5.4 Outliers and influential admissions

The plots of the scaled score residuals for each covariate in the final model (Appendix F) suggested there were no admissions with high leverage for eight covariates, and between one and three admissions with high leverage for each of the other 22 covariates (Table 4.9). After determining the study ID of the outliers, I identified 14 unique admissions with high leverage. When I excluded these 14 admissions from the derivation set, the relative change in the model coefficients was substantial, ranging from 0.39% to 176.42% (Table 4.10). After discussion with Drs. van Walraven and Forster, we decided to

exclude all 14 admissions because they represented a minimal proportion (<0.01%) of all admissions in the derivation cohort, and most admissions were rare or exceptional cases. For example, some patients were admitted due to catastrophic conditions, while others had an unusually high or low Elixhauser score for their age.

Table 4.9 High leverage admissions identified by the scaled score residuals

Covariate	# high leverage admissions*	Study ID(s)
Agesq	2	75860, 159016
Agesq1	3	75860, 159016, 28982
Agesq2	0	-
Emergent surgical	0	-
Emergent non-surgical	0	-
Elective non-surgical	0	-
Elixhauser	1	53414
LAPS	2	28982, 82532
PIMR	2	2825, 42080
(Days since last PIMR episode+1)$^{-1}$	0	-
ICU	1	2825
ALC	0	-
Agesq*Elixhauser	3	75860, 53414, 40061
Agesq1*Elixhauser	2	75860, 53414
Agesq2*Elixhauser	2	75860, 53414
Elixhauser*LAPS	1	2825
Elixhauser*ICU	2	2825, 165117
PIMR*(Days since last PIMR episode+1)$^{-1}$	3	2825, 61377, 10333
Agesq*PIMR	2	2825, 42080
Agesq1*PIMR	2	2825, 42080
Agesq2*PIMR	2	2825, 42080
LAPS*PIMR	1	2825
LAPS*ICU	1	2825
Agesq*time	1	53170
Agesq1*time	1	53170
Agesq2*time	0	-
Emerg surg*log(time)	1	50660
Emerg nonsurg*log(time)	1	50660
Elect nonsurg*log(time)	0	-
Elixhauser*log(time)	1	43639

*outliers in the plot of the scaled score residuals for the corresponding covariate

Table 4.10 Final model coefficients derived with and without influential admissions

Covariate	Parameter estimate		% change*
	Full cohort	Reduced cohort	
Agesq	0.0002153	0.0002391	11.03
Agesq1	-0.0001692	-0.0002572	52.02
Agesq2	0.0006240	0.0008745	40.16
Emergent surgical	1.8457704	1.8832885	2.03
Emergent non-surgical	1.5801596	1.6073867	1.72
Elective non-surgical	-0.2160286	-0.1859008	13.95
Elixhauser	0.0844500	0.0973452	15.27
LAPS	0.0382106	0.0384822	0.71
PIMR	-0.0369053	-0.0636552	72.48
(Days since last PIMR episode+1)$^{-1}$	0.3165349	0.3338004	5.45
ICU	2.0569861	2.0504239	0.32
ALC	-1.5467522	-1.5405372	0.40
Agesq*Elixhauser	-0.0000099	-0.0000171	72.68
Agesq1*Elixhauser	-0.0000093	0.0000071	176.42
Agesq2*Elixhauser	0.0000322	-0.0000071	121.93
Elixhauser*LAPS	-0.0003561	-0.0003457	2.90
Elixhauser*ICU	-0.0122942	-0.0101163	17.71
PIMR*(Days since last PIMR episode+1)$^{-1}$	0.3354822	0.3476764	3.63
Agesq*PIMR	0.0000493	0.0000621	25.96
Agesq1*PIMR	-0.0000650	-0.0000836	28.59
Agesq2*PIMR	0.0001079	0.0001389	28.72
LAPS*PIMR	-0.0012505	-0.0013471	7.72
LAPS*ICU	-0.0136417	-0.0136954	0.39
Agesq*time	-0.0000008	0.0000005	158.49
Agesq1*time	0.0000070	0.0000038	46.75
Agesq2*time	-0.0000190	-0.0000112	41.17
Emerg surg*log(time)	-0.3840285	-0.4085252	6.38
Emerg nonsurg*log(time)	-0.1052175	-0.1237983	17.66
Elect nonsurg*log(time)	0.1367376	0.1104737	19.21
Elixhauser*log(time)	0.0094108	0.0108091	14.86

*% change = $[(|PE_{reduced} - PE_{full}|)/PE_{full}] \times 100$

4.6 Predictive accuracy of the Escobar+ model

4.6.1 Discrimination

The *Escobar*+ model had excellent discrimination in both the derivation and validation sets. The overall *c*-statistic of the model in the derivation and validation cohort was 0.896 (95% CI 0.891-0.901) and 0.895 (95% CI 0.889-0.902), respectively.

I also calculated the *C*-statistic separately for each of the 41 Primary Condition groups (Table 4.11). In the derivation and validation sets, the *c*-statistic was >0.90 for 11 and 16 Primary Condition groups, and between 0.80 and 0.90 for 24 and 18 groups, respectively. The "Pericarditis" group had the lowest *c*-statistic (0.611 in the derivation set and 0.704 in the validation set).

Table 4.11 Discrimination of the Escobar[+] model by Primary Condition group

	Derivation			Validation		
Primary Condition	N[*]	*c*[*]	95% CI[*]	N	*c*	95% CI
Acute myocardial infarction	10945	0.891	0.869-0.913	5407	0.879	0.843-0.914
Appendicitis	8352	0.951	0.916-0.986	4205	0.968	0.927-1.009
Arthropathies	15033	0.962	0.917-1.006	7549	0.936	0.83-1.042
Cancer A	2985	0.849	0.831-0.867	1418	0.851	0.825-0.877
Cancer B	9011	0.883	0.865-0.902	4455	0.864	0.819-0.909
Ovarian and metastatic cancer	2641	0.779	0.748-0.81	1326	0.832	0.795-0.868
Catastrophic conditions	2000	0.729	0.708-0.751	1036	0.712	0.681-0.743
Congestive heart failure	2449	0.818	0.782-0.853	1224	0.785	0.741-0.829
Chronic obstructive pulmonary disorder	2676	0.840	0.804-0.875	1351	0.853	0.815-0.892
Fluid and electrolyte	652	0.882	0.825-0.94	306	0.900	0.812-0.987
Fractures and dislocations	4123	0.888	0.799-0.976	2078	0.858	0.775-0.94
Gastrointestinal bleeding	4249	0.893	0.861-0.925	2071	0.913	0.873-0.953
Gastrointestinal, inflammatory bowel disease, and obstruction	2546	0.913	0.887-0.939	1303	0.874	0.823-0.926
Gynecologic cancers	12794	0.992	0.986-0.997	6375	0.993	0.986-1
Other cardiac conditions	3670	0.879	0.831-0.927	1780	0.835	0.767-0.902
Atherosclerosis and pulmonary vascular disease	3525	0.828	0.781-0.876	1769	0.908	0.874-0.942
Non-malignant hematologic	1614	0.964	0.944-0.984	825	0.791	0.651-0.93
Hip fracture	1758	0.832	0.77-0.895	905	0.826	0.763-0.888
All other infections	4994	0.908	0.876-0.94	2435	0.879	0.831-0.928
Liver disorders	470	0.891	0.849-0.933	249	0.819	0.74-0.899

(cont'd on next page)

(Table 4.11 cont'd)

	Derivation			Validation		
Primary Condition	N	*c*	95% CI	N	*c*	95% CI
Diabetic ketoacidosis and related metabolic	913	0.839	0.706-0.973	481	0.821	0.713-0.929
Other metabolic	2019	0.861	0.765-0.958	1046	0.874	0.781-0.968
Miscellaneous cardiac	671	0.727	0.57-0.884	359	0.945	0.867-1.022
Miscellaneous 1	5216	0.876	0.791-0.961	2606	0.926	0.822-1.031
Miscellaneous 2	8450	0.886	0.871-0.901	4269	0.870	0.849-0.892
Miscellaneous 3	3501	0.963	0.928-0.998	1752	0.986	0.959-1.012
Other neurological	38748	0.971	0.955-0.987	1939 1	0.933	0.88-0.986
Ingestions and benign tumors	6124	0.918	0.854-0.982	3057	0.948	0.922-0.974
Pericarditis	995	0.611	0.465-0.757	491	0.704	0.374-1.035
Pancreatic disorders	653	0.956	0.927-0.985	333	0.939	0.865-1.014
Pneumonia	2462	0.822	0.796-0.849	1209	0.757	0.712-0.803
Acute renal failure	899	0.828	0.778-0.878	427	0.911	0.864-0.958
Chronic renal failure	393	0.887	0.816-0.959	177	0.757	0.637-0.877
Other renal	4353	0.608	0.345-0.871	2218	0.948	0.923-0.973
Acute respiratory	2851	0.928	0.905-0.951	1476	0.920	0.878-0.961
Seizure	632	0.839	0.736-0.942	333	0.903	0.688-1.119
Sepsis	824	0.775	0.749-0.801	410	0.786	0.75-0.822
Skin and autoimmune disorders	1158	0.876	0.793-0.958	593	0.981	0.951-1.012
Stroke	1750	0.800	0.767-0.833	901	0.830	0.779-0.881
All other trauma	2765	0.859	0.832-0.885	1355	0.860	0.808-0.911
Urinary tract infections	1762	0.870	0.779-0.962	869	0.857	0.758-0.957

N = number of admissions; *c*=*c*-statistic; CI=confidence interval

4.6.2 Calibration

The calibration of the *Escobar*[+] model was excellent, as indicated by the close agreement between the observed and expected number of deaths, particularly in the lower risk deciles. In the derivation set, the observed and expected number of deaths was significantly different in only two deciles on day 1, one decile on day 2, three deciles on day 6, and one decile on day 22 (Table 4.12). The calibration was even better in the validation cohort. The observed and expected number of deaths was significantly different in zero deciles on days 2 and 22 and only two deciles on both days 1 and 6 (Table 4.13).

Table 4.12 Calibration of the *Escobar*[+] model in the derivation cohort

	Day 1				Day 2				Day 6				Day 22			
Decile	N*	O*	E*	$p^{\dagger}$	N	O	E	p	N	O	E	p	N	O	E	p
1	18433	0	1.11	0.292	9791	0	0.24	0.625	4677	0	0.36	0 547	968	0	0.11	0.745
2	18084	2	1.39	0.606	9786	2	0.88	0 234	4672	0	1.02	0.313	968	1	0.37	0.297
3	18593	1	1.60	0.633	9787	1	1.79	0 553	4681	1	1.80	0.552	968	0	0.80	0.371
4	18065	0	2.02	0.155	9789	2	2.73	0.660	4676	0	3.11	0.078	968	2	1 39	0.607
5	18940	5	2.57	0.131	9787	6	4.45	0.462	4676	9	5.57	0 145	968	1	2 17	0.427
6	18061	6	3.93	0.295	9789	15	8 08	0.015	4677	8	9.43	0 642	968	1	3 47	0 185
7	18330	14	15.94	0.628	9789	15	16.63	0.689	4677	7	15.10	0.037	968	2	4 99	0 181
8	18396	56	35.33	0.001	9788	40	31 62	0.136	4676	14	25.20	0 026	968	6	8 23	0.437
9	18360	133	109.95	0 028	9788	71	60.61	0 182	4677	58	43.80	0 032	968	21	12 38	0 014
10	18364	505	548 15	0.065	9788	306	330 98	0.170	4676	137	128 63	0.460	968	31	31 10	0.986
TOTAL	183626	722	722	1.000	97882	458	458	1.000	46765	234	234	1 000	9680	65	65	1 000

*N=number of admissions; O=observed number of deaths; E=predicted (or expected) number of deaths from the model
†p-value of z-statistic, where z = (observed-expected)/(√expected)

Table 4.13 Calibration of the *Escobar*[+] model in the validation cohort

	Day 1				Day 2				Day 6				Day 22			
Decile	N*	O*	E*	$p^{\dagger}$	N	O	E	p	N	O	E	p	N	O	E	p
1	9153	0	0 54	0.461	4819	0	0 13	0 719	2323	0	0 18	0 667	474	0	0.05	0.823
2	9207	0	0.69	0.405	4819	1	0.44	0.403	2323	1	0.52	0.503	475	0	0.18	0.671
3	9328	0	0.84	0 361	4821	0	0 91	0 340	2324	1	0.88	0 898	475	0	0 35	0.552
4	9058	2	1.04	0.344	4822	1	1.40	0.737	2324	1	1 59	0.638	475	1	0.66	0 680
5	9372	1	1.26	0.818	4814	3	2.24	0 609	2323	4	2 72	0.437	474	0	1.06	0.303
6	8975	3	1.88	0.414	4819	3	4 13	0.579	2324	0	4.88	0 027	475	1	1.68	0 599
7	9175	7	7.81	0 771	4820	4	8 34	0 133	2324	4	7 48	0.203	475	2	2 74	0 654
8	9188	25	17.07	0.055	4819	21	15 82	0.193	2323	11	12 50	0.671	475	3	3 84	0 669
9	9182	76	54.61	0.004	4819	28	30.01	0.714	2324	25	22.04	0 529	475	5	6 06	0.666
10	9182	226	267 63	0 011	4819	149	172 00	0.079	2323	83	65 59	0 032	474	18	15 92	0.602
TOTAL	91820	340	353.37	0 477	48191	210	235 42	0 098	23235	130	118 39	0 286	4747	30	32 55	0 655

*N=number of admissions, O=observed number of deaths; E=predicted (or expected) number of deaths from the model
†p-value of z-statistic, where z = (observed-expected)/(√expected)

4.7 Predictive accuracy of Escobar$^+$ versus Escobar

With the same cohort I used to derive the *Escobar*$^+$ model (i.e. the derivation set excluding the 14 high leverage admissions), I fit a Cox regression model containing the Escobar model covariates. I assessed the model's predictive accuracy (discrimination and calibration) in the validation cohort only.

Compared to the Escobar model, the discrimination of the *Escobar*$^+$ model was better. The overall *c*-statistic of the Escobar model was 0.838 (0.830-0.847), whereas the overall *c*-statistic of the *Escobar*$^+$ model was 0.895 (0.889-0.902) in the validation cohort. When I compared the discrimination by primary condition group (Table 4.14), the *c*-statistic was higher for the *Escobar*$^+$ model in all but three primary condition groups ("Fractures and dislocations," "Ingestions and benign tumors," and "Urinary tract infections"). The *c*-statistic for the *Escobar*$^+$ model was significantly higher than the Escobar model (i.e. the 95% confidence intervals did not overlap) in seven primary condition groups ("Cancer A," "Catastrophic conditions," "Atherosclerosis and pulmonary vascular disease," "Acute renal failure," "Other renal," "Seizure," and "Stroke").

The numbers of *overall* deaths predicted by the Escobar model and the *Escobar*$^+$ model were very similar (Table 4.15). However, the calibration of the *Escobar*$^+$ model was better when the number of predicted deaths was broken down by risk deciles. The *Escobar*$^+$ model predicted closer to the observed number of deaths (i.e. the *p*-value associated with the *z*-statistic was larger) in all but three deciles on day 1, all deciles on day 2, and all but one decile on both days 6 and 22 (Table 4.15). Even in the five deciles where the Escobar model predicted closer, the number of deaths predicted by the *Escobar*$^+$ model was not significantly different from the observed number of deaths in three of the deciles: decile 8 on day 1 ($p = 0.055$), decile 9 on day 6 ($p = 0.529$), and decile 9 on day 22 ($p = 0.666$).

Table 4.14 Discrimination of *Escobar*$^{+}$ versus Escobar by Primary Condition group

Primary Condition	# of admissions	*Escobar*$^{+}$		Escobar	
		c*	95% CI*	*c*	95% CI
Acute myocardial infarction	5407	0.879	0.843-0.914	0.820	0.782-0.858
Appendicitis	4205	0.968	0.927-1.009	0.827	0.627-1.026
Arthropathies	7549	0.936	0.830-1.042	0.913	0.780-1.045
Cancer A	1418	0.851	0.825-0.877	0.769	0.736-0.803
Cancer B	4455	0.864	0.819-0.909	0.771	0.720-0.821
Ovarian and metastatic cancer	1326	0.832	0.795-0.868	0.751	0.706-0.796
Catastrophic conditions	1036	0.712	0.681-0.743	0.578	0.544-0.612
Congestive heart failure	1224	0.785	0.741-0.829	0.695	0.644-0.746
Chronic obstructive pulmonary disorder	1351	0.853	0.815-0.892	0.809	0.767-0.851
Fluid and electrolyte	306	0.900	0.812-0.987	0.860	0.769-0.951
Fractures and dislocations	2078	0.858	0.775-0.940	0.879	0.798-0.961
Gastrointestinal bleeding	2071	0.913	0.873-0.953	0.871	0.819-0.923
Gastrointestinal, inflammatory bowel disease, and obstruction	1303	0.874	0.823-0.926	0.808	0.745-0.870
Gynecologic cancers	6375	0.993	0.986-1.000	0.985	0.971-0.999
Other cardiac conditions	1780	0.835	0.767-0.902	0.791	0.707-0.875
Atherosclerosis and pulmonary vascular disease	1769	0.908	0.874-0.942	0.810	0.749-0.870
Non-malignant hematologic	825	0.791	0.651-0.930	0.669	0.541-0.796
Hip fracture	905	0.826	0.763-0.888	0.762	0.691-0.833
All other infections	2435	0.879	0.831-0.928	0.803	0.753-0.852
Liver disorders	249	0.819	0.740-0.899	0.698	0.598-0.798
Diabetic ketoacidosis and related metabolic	481	0.821	0.713-0.929	0.794	0.677-0.911
Other metabolic	1046	0.874	0.781-0.968	0.683	0.543-0.824
Miscellaneous cardiac	359	0.945	0.867-1.022	0.814	0.598-1.031
Miscellaneous 1	2606	0.926	0.822-1.031	0.875	0.798-0.951
Miscellaneous 2	4269	0.870	0.849-0.892	0.853	0.828-0.878
Miscellaneous 3	1752	0.986	0.959-1.012	0.983	0.954-1.013
Other neurological	19391	0.933	0.880-0.986	0.841	0.714-0.967
Ingestions and benign tumors	3057	0.948	0.922-0.974	0.826	0.727-0.925
Pericarditis	491	0.704	0.374-1.035	0.713	0.493-0.934
Pancreatic disorders	333	0.939	0.865-1.014	0.858	0.695-1.022
Pneumonia	1209	0.757	0.712-0.803	0.708	0.661-0.754
Acute renal failure	427	0.911	0.864-0.958	0.757	0.676-0.838
Chronic renal failure	177	0.757	0.637-0.877	0.687	0.499-0.876
Other renal	2218	0.948	0.923-0.973	0.884	0.858-0.909
Acute respiratory	1476	0.920	0.878-0.961	0.916	0.840-0.992
Seizure	333	0.903	0.688-1.119	0.364	0.264-0.463
Sepsis	410	0.786	0.750-0.822	0.712	0.670-0.754
Skin and autoimmune disorders	593	0.981	0.951-1.012	0.859	0.670-1.047
Stroke	901	0.830	0.779-0.881	0.699	0.638-0.761
All other trauma	1355	0.860	0.808-0.911	0.780	0.720-0.840
Urinary tract infections	869	0.857	0.758-0.957	0.876	0.795-0.957

**c*=*c*-statistic; CI=confidence interval

Table 4.15 Calibration of *Escobar*[+] versus Escobar

Decile*	Day 1			Day 2			Day 6			Day 22		
	Deaths	*Escobar*[+]	Escobar	Deaths	*Escobar*[+]	Escobar	Deaths	*Escobar*[+]	Escobar	Deaths	*Escobar*[+]	Escobar
1	0	0.54	1.58	0	0.13	2.14	0	0.18	1.11	0	0.05	0.43
2	0	0.69	2.41	1	0.44	2.34	1	0.52	1.99	0	0.18	0.94
3	0	0.84	3.28	0	0.91	4.57	1	0.88	2.74	0	0.35	1.48
4	2	1.04	4.37	1	1.40	5.38	1	1.59	4.52	1	0.66	2.18
5	1	1.26	5.32	3	2.24	7.99	4	2.72	6.70	0	1.06	2.66
6	3	1.88	7.72	3	4.13	13.31	0	4.88	10.65	1	1.68	3.82
7	7	7.81	14.33	4	8.34	22.85	4	7.48	13.97	2	2.74	4.29
8	25	17.07	33.12[†]	21	15.82	34.50	11	12.50	18.78	3	3.84	4.56
9	76	54.61	79.62[†]	28	30.01	46.89	25	22.04	24.24[†]	5	6.06	5.47[†]
10	226	267.63	203.52[†]	149	172.00	87.07	83	65.59	32.62	18	15.92	7.27
TOTAL	340	353.37	355.28	210	235.42	227.05	130	118.39	117.32	30	32.55	33.10

*Deciles based on predicted risk of in-hospital death from the *Escobar*[+] model

[†]Deciles where the *p*-value from the *z*-statistic was larger for the original Escobar model than the new *Escobar*+ model

4.8 Sensitivity analyses

4.8.1 Informative censoring

When I re-fit the *Escobar*$^{+}$ model changing the censoring time of all censored patients to one day greater than the largest event time (425 days), the parameter estimates were substantially different than those obtained using the observed censoring times (Table 4.16). However, I found there was virtually no change in the model discrimination when I used these parameter estimates to calculate the *c*-statistic in the validation set (0.892, 95% CI 0.885-0.899).

4.8.2 Clustering of admissions

When I re-fit the *Escobar*$^{+}$ model adjusting for the effect of multiple admissions per patient [i.e. using the COV(AGGREGATE) option in the PROC PHREG statement] the parameter estimates did not change and the standard errors and *p*-values were very similar (Table 4.17).

4.8.3 Irregular measurement intervals for LAPS

When I added the time-dependent covariate "time elapsed since the last change in LAPS" to the *Escobar*$^{+}$ model (as a proxy to adjust for the effect of irregularly collected laboratory test results), it was a highly significant predictor of in-hospital mortality ($p<.0001$).

Table 4.16 *Escobar*[+] coefficients derived using observed versus extreme censoring times

Covariate	Parameter estimate: Observed censoring time	Parameter estimate: Extreme censoring time (426 days)	% change*
Agesq	0.0002391	-0.0003005	225.67
Agesq1	-0.0002572	0.0008939	447.59
Agesq2	0.0008745	-0.0016923	293.50
Emergent surgical	1.8832885	0.7052783	62.55
Emergent non-surgical	1.6073867	0.3761402	76.60
Elective non-surgical	-0.1859008	-1.3975979	651.80
Elixhauser	0.0973452	0.1549543	59.18
LAPS	0.0384822	0.0419536	9.02
PIMR	-0.0636552	0.0307982	148.38
(Days since last PIMR episode+1)$^{-1}$	0.3338004	0.2554856	23.46
ICU	2.0504239	2.2567932	10.06
ALC	-1.5405372	-1.4330272	6.98
Agesq*Elixhauser	-0.0000171	-0.0000392	129.87
Agesq1*Elixhauser	0.0000071	0.0000499	600.37
Agesq2*Elixhauser	-0.0000071	-0.0000980	1286.02
Elixhauser*LAPS	-0.0003457	-0.0004615	33.48
Elixhauser*ICU	-0.0101163	-0.0081573	19.37
PIMR*(Days since last PIMR episode+1)$^{-1}$	0.3476764	0.3472349	0.13
Agesq*PIMR	0.0000621	0.0000261	58.02
Agesq1*PIMR	-0.0000836	-0.0000133	84.08
Agesq2*PIMR	0.0001389	-0.0000117	108.45
LAPS*PIMR	-0.0013471	-0.0014672	8.92
LAPS*ICU	-0.0136954	-0.0152774	11.55
Agesq*time	0.0000005	0.0001171	24015.67
Agesq1*time	0.0000038	-0.0002307	6249.25
Agesq2*time	-0.0000112	0.0005008	4589.65
Emerg surg*log(time)	-0.4085252	0.2862418	170.07
Emerg nonsurg*log(time)	-0.1237983	0.5889118	575.70
Elect nonsurg*log(time)	0.1104737	0.8446387	664.56
Elixhauser*log(time)	0.0108091	0.0139609	29.16

*% change = $[(|PE_{observed} - PE_{extreme}|)/PE_{observed}] \times 100$

Table 4.17 *Escobar*[+] coefficients unadjusted and adjusted for clustering of admissions

	Unadjusted			Adjusted*		
Covariate	Parameter estimate	Standard error	*p*-value	Parameter estimate	Standard error	*p*-value
Agesq	0.0002391	0.0001	0.0230	0.0002391	0.0001	0.0221
Agesq1	-0.0002572	0.0003	0.3438	-0.0002572	0.0003	0.3366
Agesq2	0.0008745	0.0007	0.1894	0.0008745	0.0007	0.1818
Emerg surg	1.8832885	0.2006	<.0001	1.8832885	0.1970	<.0001
Emerg nonsurg	1.6073867	0.1919	<.0001	1.6073867	0.1855	<.0001
Elect nonsurg	-0.1859008	0.3509	0.5963	-0.1859008	0.3791	0.6238
Elixhauser	0.0973452	0.0160	<.0001	0.0973452	0.0161	<.0001
LAPS	0.0384822	0.0008	<.0001	0.0384822	0.0008	<.0001
PIMR	-0.0636552	0.0600	0.2885	-0.0636552	0.0583	0.2752
(Days since last PIMR episode+1)$^{-1}$	0.3338004	0.0549	<.0001	0.3338004	0.0550	<.0001
ICU	2.0504239	0.1299	<.0001	2.0504239	0.1481	<.0001
ALC	-1.5405372	0.0967	<.0001	-1.5405372	0.1028	<.0001
Agesq*Elixhauser	-0.0000171	0.0000	0.0228	-0.0000171	0.0000	0.0236
Agesq1*Elixhauser	0.0000071	0.0000	0.7079	0.0000071	0.0000	0.7054
Agesq2*Elixhauser	-0.0000071	0.0000	0.8797	-0.0000071	0.0000	0.8782
Elixhauser*LAPS	-0.0003457	0.0001	<.0001	-0.0003457	0.0001	<.0001
Elixhauser*ICU	-0.0101163	0.0049	0.0392	-0.0101163	0.0049	0.0387
PIMR*(Days since last PIMR episode+1)$^{-1}$	0.3476764	0.0338	<.0001	0.3476764	0.0385	<.0001
Agesq*PIMR	0.0000621	0.0000	0.0221	0.0000621	0.0000	0.0208
Agesq1*PIMR	-0.0000836	0.0001	0.2070	-0.0000836	0.0001	0.2056
Agesq2*PIMR	0.0001389	0.0002	0.3894	0.0001389	0.0002	0.3904
LAPS*PIMR	-0.0013471	0.0002	<.0001	-0.0013471	0.0003	<.0001
LAPS*ICU	-0.0136954	0.0013	<.0001	-0.0136954	0.0014	<.0001
Agesq*time	0.0000005	0.0000	0.8880	0.0000005	0.0000	0.8725
Agesq1*time	0.0000038	0.0000	0.6561	0.0000038	0.0000	0.6061
Agesq2*time	-0.0000112	0.0000	0.5837	-0.0000112	0.0000	0.5236
Emerg surg*log(time)	-0.4085252	0.0757	<.0001	-0.4085252	0.0727	<.0001
Emerg nonsurg*log(time)	-0.1237983	0.0718	0.0848	-0.1237983	0.0676	0.0671
Elect nonsurg*log(time)	0.1104737	0.1215	0.3630	0.1104737	0.1290	0.3916
Elixhauser*log(time)	0.0108091	0.0016	<.0001	0.0108091	0.0016	<.0001

*Adjusted for clustering using the COV(AGGREGATE) option in the PROC PHREG statement

5. Discussion

5.1 Study findings

5.1.1 Primary findings

In this study, I derived and internally validated a time-dependent risk prediction model for in-hospital mortality that could be used on a generic (i.e. disease non-specific) patient population. I found that the following factors were significant predictors of in-

hospital mortality: age at admission, admission type (emergent/elective and surgical/non-surgical), Elixhauser score, LAPS, PIMR, number of days elapsed since the last PIMR episode, admission to ICU, and change to ALC status. The first three factors were time-independent covariates whose values could be determined at admission and stayed constant throughout the hospitalization. The other five factors were time-dependent covariates whose values could change throughout the hospitalization.

The final risk prediction model (the *Escobar*$^{+}$ model) contained eight main effects and seven interactions between the main effects. The model also included interactions between each of the three time-independent main effects (age, admission type, and the Elixhauser score) and a function of time (linear or logarithmic).

The discrimination of the *Escobar*$^{+}$ model was excellent. In the derivation and validation set, the model predictions were concordant approximately 90% of the time (i.e. the model predicted a higher risk for admissions in which the patient died sooner). The discrimination was also good when I determined the *c*-statistic separately for each of the 41 Primary Condition groups: the predictions were concordant more than 80% of the time in 35 Primary Condition groups in the derivation set and 34 groups in the validation set.

The calibration of the *Escobar*$^{+}$ model was also excellent, evidenced by the close agreement between the expected and observed number of deaths in most risk deciles at four different time-points during the admission (day 1, 2, 6, and 22). I found that the calibration was better on day 2 than day 1, possibly because the LAPS was more accurate on day 2 (49% of admissions had at least one LAPS laboratory test performed prior to day 2, compared to only 31% of admissions prior to day 1). I also found that, compared to the derivation cohort, the model calibration was better in the validation cohort since the number of expected and observed deaths was significantly different in fewer risk deciles. I was surprised to find that the calibration was still excellent on day 22 (the 95th percentile of length of stay). I expected the model calibration to worsen with time since I calculated the

daily estimated number of events (i.e. the daily hazard) using the daily baseline hazard estimate, which becomes increasingly unstable over time due to lack of observations.

5.1.2 Secondary findings

5.1.2.1 Predictive accuracy of the Escobar model fit using logistic versus Cox regression

In order to fairly compare the predictive accuracy of the *Escobar*$^{+}$ model (a Cox regression model) to the original Escobar model (a logistic regression model), I first had to determine the predictive accuracy of the Escobar model fit using Cox regression methods. This was necessary to ensure that the observed differences between the *Escobar*$^{+}$ model and the original Escobar model were not due to the different regression methods used to derive them, or the different methods used to obtain the *c*-statistic (to measure discrimination) and the model-predicted number of deaths (to measure calibration).

I found that the discrimination of the Escobar model appeared worse when it was derived (and assessed) as a Cox rather than a logistic regression model. The *c*-statistic for the logistic model was 93%, whereas the *c*-statistic for the Cox model was 84%. The lower *c*-statistic for the Cox model was not surprising given that the criteria for a concordant prediction is more stringent for a Cox model compared to a logistic model. A concordant pair for a Cox model requires that the event occurs *earlier* for the admission with the higher risk score. In contrast, a concordant pair for a logistic model requires only that the event occurs in the admission with the higher predicted risk. The additional consideration of the event time for measuring the concordance probability (or *c*-statistic) from a Cox model makes it more difficult for the model to achieve excellent discrimination.

It was unclear whether the calibration of the Escobar model was better when it was fit using logistic or Cox regression. The calibration of the Cox model appeared slightly better when admissions were grouped into risk deciles based on the predicted risk from its respective model. However, the calibration of the logistic model appeared better when

admissions were grouped into the same risk deciles (based on the predicted risk from the logistic model).

5.1.2.2 Predictive accuracy of Escobar$^+$ versus Escobar

When I compared the predictive accuracy of the *Escobar*$^+$ model to the Escobar model (fit using Cox regression), I found that the *Escobar*$^+$ model was superior. The proportion of concordant predictions (discrimination) was 90% for the *Escobar*$^+$ model versus 84% for the Escobar model. As well, compared to the Escobar model, the daily number of deaths (calibration) predicted by the *Escobar*$^+$ model on days 1, 2, 6, and 22 was much closer to the observed number of deaths on each day.

The superiority of the *Escobar*$^+$ model over the original Escobar model was expected given that the *Escobar*$^+$ model used much more information about the patient to predict risk of in-hospital death. While the Escobar model only used patient information available at the time of admission, the *Escobar*$^+$ model used information generated both pre- and post-admission to predict risk of death.

5.1.3 Other findings

5.1.3.1 Use of the fractional polynomial function selection procedure to determine the best functional form for continuous variables

In the exploratory analysis for the *Escobar*$^+$ model, I used the fractional polynomial function selection procedure to determine the best functional form for each continuous covariate (i.e. the functional form that made each covariate as linear as possible in the log hazard) in a *univariable* model. Then during the model-building stage, I entered all continuous covariates (except age since it was expressed as a spline) into the initial multivariable model in these functional forms to reduce the chance that a truly significant continuous covariate was eliminated because its form was improperly specified. Next, I

used the function selection procedure again to determine the best functional form for all continuous covariates (except age) in a *multivariable* model containing all significant covariates. Contrary to my expectations, I found that the best functional form for most continuous covariates was different and simpler in the multivariable model than in the univariable models. In fact, the best functional form in the multivariable model was linear for all continuous covariates except "days since the last PIMR episode."

In future studies, I would recommend using the function selection procedure in the pre-model building stage only with covariates that are non-significant in a univariable model (to determine if any transformation will make the covariate significant). For covariates that are highly significant in their linear form in a univariable model, the function selection procedure may not be necessary in the pre-model building stage because these covariates (in their linear state) will likely still be significant in the multivariable model, and the best functional form for the covariate in the multivariable model will likely differ from the best form in a univariable model.

5.1.3.2 Disparity between the results of different methods for assessing the proportional hazards assumption

In this study, I used the following methods to assess if the proportional hazards assumption was satisfied for time-independent covariates: log-negative-log plots, graphs comparing the observed standardized score process with simulated residual paths, the Supremum test for proportional hazards, the weighted Schoenfeld residuals, and the addition of interaction terms between covariates and a function of time to the Cox model. Evidence to support the proportional hazards assumption included: parallel lines for the different strata in the log-negative-log plot; an observed standardized score process that closely resembled the simulated residual paths; a non-significant *p*-value for the

Supremum test; no significant trend over time in the weighted Schoenfeld residuals; and a non-significant interaction between the covariate and a function of time in the Cox model.

In the exploratory analysis, I used all five methods to assess the proportional hazards assumption. In the *Escobar*+ model, however, I could only use the latter two methods to assess proportionality because the former three methods could not be used with a time-dependent Cox model.

I found that, for some covariates, there was a discrepancy in the results from these different methods. For example, in the *Escobar*+ model, there was no significant linear trend over time in the weighted Schoenfeld residuals for splined age^2 (suggesting proportional hazards), but the addition of an interaction between splined age^2 and time to the Cox model was significant (suggesting non-proportional hazards). In the exploratory analysis, when I assessed the proportional hazards assumption for admission type, the log-negative-log plot, weighted Schoenfeld residuals, and interaction with time all suggested non-proportionality. On the other hand, the *p*-value of the Supremum test and the graphs comparing the observed standardized score process with the simulated residuals paths suggested that the hazards were proportional over time. These examples highlight the importance of using multiple methods to assess the proportional hazards assumption rather than forming conclusions based on the results of one method alone.

To adjust for non-proportionality in this study, I included all interaction terms with time in the final model if they were statistically significant (regardless of the result from other methods) because there were a large number of events in the cohort (i.e. the model could handle a large number of covariates without risk of overfitting).

5.2 Rationale for using time-dependent Cox regression methods to derive the Escobar$^+$ model

In this study, I used time-dependent Cox regression methods to derive the *Escobar*$^+$ model. Alternatively, I could have added post-admission events to the Escobar model using the original methodology (logistic regression) and moving the study "baseline" from admission (as in the Escobar model) forward to a new time to avoid time-dependent bias. I chose the former approach in this study because there were a number of advantages to using this statistical methodology.

First, the *Escobar*$^+$ model (derived using time-dependent Cox regression methods) was highly parsimonious. The model had only one set of parameter estimates that could be applied to patients on all days of admission. These parameter estimates could also be applied to patients in all primary condition groups (because I adjusted for primary conditions in the Cox model using the STRATA statement). On the other hand, using logistic regression would have required deriving a *different* model (and set of parameter estimates) on each day for each primary condition group (since the original Escobar model used the BY statement to adjust for the primary condition groups). This would have made the resulting model extremely cumbersome and difficult to use.

Second, because the *Escobar*$^+$ model used all the data to estimate the model coefficients, the coefficients were extremely stable (as indicated by the small standard errors) and the model had excellent predictive accuracy at various time points during the admission. With logistic regression, the model for each day would have been derived using only admissions where the patient was still in hospital on that day. Since most admissions were short (less than six days for 75% of admissions), the model for longer lengths of stay would have had unstable coefficients (with large standard errors) or would have had problems converging (due to lack of sufficient numbers of admissions).

Consequently, the predictive accuracy of the model on these days would have likely been poor.

Third, because the *Escobar*$^{+}$ model was a Cox (i.e. time-to-event) model, it could produce time-specific estimates of death risk (i.e. risk of in-hospital death in the next day, week, month, etc). The logistic models, on the other hand, would have only been able to produce one estimate of death risk for each patient (the probability that the patient died in hospital at some unspecified time after admission).

5.2 Study strengths

To my knowledge, this is the first study to derive a disease non-specific risk prediction model for in-hospital mortality that includes both pre- and post-admission events. When I searched the scientific literature, I identified only six previously published risk prediction models for in-hospital mortality that could be applied to a disease non-specific patient population. These six models were derived using logistic regression, which means they could include only pre-admission covariates to safely avoid time-dependent bias. I was able to include post-admission covariates in the *Escobar*$^{+}$ model and avoid time-dependent bias because I derived the model using time-dependent Cox regression methods.

This study employs a very large cohort of hospitalizations. As a result, the study had ample power to detect associations, and the cohort could produce a risk prediction model with stable parameter estimates.

I assessed the predictive accuracy of the *Escobar*$^{+}$ model on the training sample (the derivation set) and internally validated the model by assessing its performance on separate sample of admissions (the validation set). Assessing a model's predictive accuracy is important to determine its utility for making predictions and to identify any lack of fit (62). Assessing a model's performance on a separate sample is especially important

because it can reveal overfitting and lack of fit better than with the training sample (63). The *Escobar*$^{+}$ model had excellent predictive accuracy when tested on both the derivation and validation set. I also found that the number of deaths predicted by the model was extremely accurate at various time points during the admission (day 1, 2, 6, and 22).

Finally, as mentioned in Section 5.2, the *Escobar*$^{+}$ model can produce time-specific risk predictions (i.e. risk of in-hospital death in the next day, week, month, etc.) that can be updated on a daily basis depending on the events that occur during the hospitalization. In comparison, the original Escobar model (derived using logistic regression) can only predict the risk of in-hospital death at some unspecified time point after admission, and this estimate remains constant for the entire admission (regardless of the events that occur post-admission). The predictions from the *Escobar*$^{+}$ model have greater utility because they can allow physicians to identify patients whose health condition is deteriorating (indicated by an increasing trend over time in the patient's daily death risk) and help hospitals plan for bed and resource needs (by predicting the number of patients who will die or be discharge in the next day).

5.3 Study weaknesses

This study has a number of limitations. First and foremost, the study violates a key assumption in survival analysis: the assumption of non-informative censoring, which assumes that censored subjects have the same risk of experiencing the event as uncensored subjects. In this study, patients who were censored (i.e. discharged) were almost certainly at a lower risk of in-hospital death compared to uncensored patients. Although there are no standard methods to adjust for informative censoring (56), I performed a sensitivity analysis under the extreme assumption that all censored patients, had they not been discharged from hospital, would have survived past the largest event time (425 days). Since the discrimination of the *Escobar*$^{+}$ model was virtually unchanged

under this extreme assumption, the model's predictive accuracy was likely unaffected by informative censoring. Although the model coefficients were substantially different, this does not affect the study results since I did not interpret the model coefficients in this study.

I did not assess the assumption of proportionality for the time-dependent covariates in the *Escobar*$^{+}$ model. Currently, there is no established methodology for doing this since time-dependent covariates are usually the means used to adjust for non-proportionality in time-independent covariates. If the proportional hazard assumption was violated for a time-dependent covariate in the *Escobar*$^{+}$ model, however, the consequences would not be detrimental to the validity of the model. The parameter estimate would represent the average effect of the covariate over the observed range of event times (64), and it is unlikely that the hazard ratio would change dramatically over time because the observation period for most admissions in this study was very short (six days or less for 75% of admissions). Violation of the proportional hazards assumption is bigger concern in studies where the follow-up period is long (i.e. years) and the magnitude of change in the hazard ratio over time is likely to be greater (65). Furthermore, since only a small proportion of admissions experienced at least one post-admission event, violation of the proportional hazard assumption for the time-dependent covariates would unlikely affect the predictive accuracy of the model.

The study cohort included multiple admissions per patient, which violates the assumption of independent observations in a Cox regression model. When I performed a sensitivity analysis to adjust for clustering of admissions, however, the model coefficients and standard errors were similar, suggesting that the effect of clustering was negligible.

Most of the data used in this study came from the hospital's administrative databases. Consequently, some information may have been inaccurate or incomplete since the data was not collected for research purposes. While data on patient demographics and medical procedures was likely accurate, the data on diagnoses may

have been incomplete and/or inaccurate. Studies comparing information from Canadian administrative databases to re-abstracted data from other sources (i.e. patient medical charts, clinical databases) have found high levels of agreement for demographic information and procedure codes, but have found variation in the accuracy and completeness of diagnosis codes (66). As well, because I was limited to administrative (and laboratory) data in this study, I was unable to consider clinical factors (i.e. daily vital signs) since such data is not captured in administrative databases. Future models will need to determine the added value of including clinical factors in the *Escobar*$^{+}$ model.

The value of LAPS may have been inaccurate at times due to lack of available or updated test results. Escobar *et al.* (10), however, found that the discrimination of the original Escobar model was unchanged when they introduced random changes in LAPS and COPS in as many as 25% of their data. I did not adjust for the effect of irregular measurement intervals for LAPS using the method recommended by De Bruijne *et al.* (67) because I could not accurately quantify the time elapsed since the last observation. Since all 14 LAPS laboratory tests were rarely ever performed at the same time, the time elapsed since the last measure was different for each of the component tests. As a sensitivity analysis, however, I tested the significance of a proxy measure ("time elapsed since the last change in LAPS") under the assumption that, for some patients, LAPS would change if enough tests results were updated. Although I found that this covariate was significant, I did not include it in the final model because it was a very crude proxy measure. Since the component scores for LAPS were assigned based on test result *ranges*, it was highly probable that a patient's LAPS could have stayed the same even though some of the test results were updated.

Finally, because this study was limited to admissions from only one health care institution, it is unknown well the *Escobar*$^{+}$ model predicts risk of in-hospital death among patients from other health care centres.

5.4 Impact of study findings

The findings from this study have the potential to greatly improve clinical research and the delivery of quality patient care.

In studies where the start of observation occurs after admission to the hospital, the *Escobar*$^{+}$ model can be used to improve risk adjustment by more accurately adjusting for baseline differences in patient risk. Such examples include studies of outcomes associated with nosocomial infections or certain in-hospital procedures, or studies comparing mortality rates between different nursing stations or hospital units. Whereas the Escobar model could only adjust for the effect of factors known at the time of admission, *Escobar*$^{+}$ can additionally adjust for the effect of post-admission factors that occur prior to the start of observation.

The *Escobar*$^{+}$ model can enable physicians and hospital administrators to improve patient care and monitor hospital performance. Physicians can use the model to obtain accurate, up-to-date estimates of their patients' death risk to inform medical decisions regarding treatment and care. Hospital administrators can use the *Escobar*$^{+}$ model to determine the severity of illness among their current patient population and better meet the needs of their patients by forecasting the hospital's resource needs. By comparing the number of expected to observed deaths, hospital administrators can also use the *Escobar*$^{+}$ model to monitor the performance of the entire hospital or specific units on a yearly, monthly, weekly, or even daily basis.

Finally, the findings from this study can be used to fuel further research on in-hospital death and prediction of post-discharge outcomes. For example, by plotting the predicted risk of in-hospital death on each day of admission from the *Escobar*$^{+}$ model, these "hazard profiles" can be used to study how the risk of death typically changes during a hospitalization and identify factors associated with different trends (i.e. admissions where the risk of death decreases, increases, or stays constant during the hospitalization). The

Escobar$^+$ model can also predict patients' probability of survival at discharge using Equation 5.1:

$$S(t_d) = \exp[-H(t_d)] \quad (5.1)$$

where t_d represents the time of discharge and H represents the estimated cumulative hazard. Since this estimate incorporates the effect of all events that occurred during the admission, it is likely a good predictor of post-discharge outcomes (i.e. unplanned re-admission or death after discharge). One could compare the predictive accuracy of this estimate with other predictors of post-discharge outcomes, such as the LACE index (68).

5.5 Future work

In this study, I used administrative data to derive and validate the *Escobar*$^+$ model. However, health care professionals who have access to primary data may want to use primary data to predict death risk. Because primary data (especially data on procedures and patient diagnoses) are usually more accurate and complete than administrative data, future studies are required to determine how well the *Escobar*$^+$ model performs using primary data.

Although the *Escobar*$^+$ model had excellent predictive accuracy in the validation set, the model should be prospectively validated to ensure it is not overfitted to the data. Prospective validation will also determine how well the model performs on administrative data from admissions that occur during a different time period, since coding practices used by health records abstractors can change over time. Prospective validation can easily be done because The Ottawa Hospital collects the required data for the *Escobar*$^+$ model on all admissions on an automatic and ongoing basis.

Finally, the *Escobar*$^+$ model should be externally validated to determine its predictive performance in a different patient population. There are few external sites that can validate the *Escobar*$^+$ model, however, because few hospitals maintain automated

laboratory databases. In addition, the hospital must record the date and time of all post-admission events included in the *Escobar*$^{+}$ model (admission to ICU, change to ALC status, and performance of in-hospital procedures).

6. Conclusions

In this study, I have derived and validated the *Escobar*$^{+}$ model, a highly accurate risk prediction model for in-hospital mortality that can be applied to a generic patient population. The model expands on a previous risk prediction model that uses only information available at the time of admission. The *Escobar*$^{+}$ model achieves excellent predictive accuracy by making use of information available prior to *and* after admission.

The *Escobar*$^{+}$ model can be used as a powerful risk-adjustment methodology in research studies because it is able to predict patients' risk of death at any post-admission time point. The predictions from the *Escobar*$^{+}$ model also have the potential to improve patient care by assisting treatment decisions by physicians and enabling hospitals to monitor their performance on a daily basis.

Finally, this study is proof of the utility of administrative and clinical databases, and maintaining a data warehouse. By storing administrative and clinical information from the hospital's various databases in one location, the data warehouse is a valuable tool that can be used by academics for clinical research, or by hospital administrators to monitor and improve hospitals' performance and delivery of quality health care.

7. References

(1) Canadian Institute for Health Information. HSMR: A New Approach for Measuring Hospital Mortality Trends in Canada. 2007. Ottawa, C.

(2) Lied TR, Kazandjian VA, Hohman SF. Impact of risk adjusted clinical outcomes methodology--quality measures on hospital mortality data: a statistical and case study approach. Am J Med Qual 1999; 14(6):255-261.

(3) Iezzoni LI. Risk Adjustment for Measuring Health Care Outcomes (Third Edition). Chicago, IL: Health Administration Press, 2003.

(4) Smith DW. Evaluating risk adjustment by partitioning variation in hospital mortality rates. Stat Med 1994; 13(10):1001-1013.

(5) Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996; 15(4):361-387.

(6) Kremers WK. Concordance for Survival Time Data: Fixed and Time-Dependent Covariates and Possible Ties in Predictor and Time. 2007. Mayo Foundation.

(7) van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. Med Care 2009; 47(6):626-633.

(8) Bottle A, Aylin P. Intelligent Information: A National System for Monitoring Clinical Performance. Health Serv Res 2008; 43(1, Part I):10-31.

(9) Miyata H, Hashimoto H, Horiguchi H, Matsuda S, Motomura N, Takamoto S. Performance of in-hospital mortality prediction models for acute hospitalization: hospital standardized mortality ratio in Japan. BMC Health Serv Res 2008; 8:229.

(10) Escobar GJ, Greene JD, Scheirer P, Gardner MN, Draper D, Kipnis P. Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases. Med Care 2008; 46(3):232-239.

(11) Ramiarina RA, Ramiarina BL, Almeida RM, Pereira WC. Comorbidity adjustment index for the international classification of diseases, 10th revision. Rev Saude Publica 2008; 42(4):590-597.

(12) John-Baptiste A, Naglie G, Tomlinson G, Alibhai SM, Etchells E, Cheung A et al. The effect of English language proficiency on length of stay and in-hospital mortality. J Gen Intern Med 2004; 19(3):221-228.

(13) Glenn LL, Jijon CR. Risk-adjusted in-hospital death rates for peer hospitals in rural and urban regions. J Rural Health 1999; 15(1):94-107.

(14) van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. Med Care 2009; 47(6):626-633.

(15) Aylin P, Bottle A, Majeed A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. BMJ 2007; 334(7602):1044.

(16) van Walraven C, Davis D, Forster AJ, Wells GA. Time-dependent bias was common in survival analyses published in leading clinical journals. J Clin Epidemiol 2004; 57(7):672-682.

(17) van Walraven C, Escobar GJ, Greene JD, Forster AJ. The Kaiser Permanente inpatient risk adjustment methodology was valid in an external patient population. J Clin Epidemiol. In press.

(18) Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 2001; 54(8):774-781.

(19) Bleeker SE, Moll HA, Steyerberg EW, Donders AR, Derksen-Lubsen G, Grobbee DE et al. External validation is necessary in prediction research: a clinical example. J Clin Epidemiol 2003; 56(9):826-832.

(20) Escobar GJ, Greene JD, Scheirer P, Gardner MN, Draper D, Kipnis P. Appendix to: Risk-adjusting hospital inpatient mortality using automated inpatient, outpatient, and laboratory databases. 2008.

(21) Fisher LD, Lin DY. Time-dependent covariates in the Cox proportional-hazards regression model. Annu Rev Public Health 1999; 20:145-157.

(22) Ontario Hospital Association, Ontario Ministry of Health, Hospital Medical Records Institute. Report of the Ontario Data Quality Reabstracting Study. 1991. Toronto, Ontario Hospital Association.

(23) Harrell FE, Jr. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. NY, USA: Springer-Verlag New York, Inc., 2001.

(24) Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987; 40(5):373-383.

(25) van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. Med Care 2009; 47(6):626-633.

(26) Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care 1998; 36(1):8-27.

(27) Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care 2005; 43(11):1130-1139.

(28) Southern DA, Quan H, Ghali WA. Comparison of the Elixhauser and Charlson/Deyo methods of comorbidity measurement in administrative data. Med Care 2004; 42(4):355-360.

(29) Stukenborg GJ, Wagner DP, Connors AF, Jr. Comparison of the performance of two comorbidity measures, with and without information from prior hospitalizations. Med Care 2001; 39(7):727-739.

(30) Dominick KL, Dudley TK, Coffman CJ, Bosworth HB. Comparison of three comorbidity measures for predicting health service use in patients with osteoarthritis. Arthritis Rheum 2005; 53(5):666-672.

(31) Canadian Institute for Health Information. Alternate Level of Care in Canada. 1-14-2009.

(32) Tekkis PP, Prytherch DR, Kocher HM, Senapati A, Poloniecki JD, Stamatakis JD et al. Development of a dedicated risk-adjustment scoring system for colorectal surgery (colorectal POSSUM). Br J Surg 2004; 91(9):1174-1182.

(33) Tran Ba LP, du Montcel ST, Duron JJ, Levard H, Suc B, Descottes B et al. Elderly POSSUM, a dedicated score for prediction of mortality and morbidity after major colorectal surgery in older patients. Br J Surg 2010; 97(3):396-403.

(34) Lloyd H, Ahmed I, Taylor S, Blake JR. Index for predicting mortality in elderly surgical patients. Br J Surg 2005; 92(4):487-492.

(35) Alves A, Panis Y, Mantion G, Slim K, Kwiatkowski F, Vicaut E. The AFC score: validation of a 4-item predicting score of postoperative mortality after colorectal resection for cancer or diverticulitis: results of a prospective multicenter study in 1049 patients. Ann Surg 2007; 246(1):91-96.

(36) Sutton R, Bann S, Brooks M, Sarin S. The Surgical Risk Scale as an improved tool for risk-adjusted analysis in comparative surgical audit. Br J Surg 2002; 89(6):763-768.

(37) Prytherch DR, Whiteley MS, Higgins B, Weaver PC, Prout WG, Powell SJ. POSSUM and Portsmouth POSSUM for predicting mortality. Physiological and Operative Severity Score for the enUmeration of Mortality and morbidity. Br J Surg 1998; 85(9):1217-1220.

(38) Tekkis PP, Prytherch DR, Kocher HM, Senapati A, Poloniecki JD, Stamatakis JD et al. Development of a dedicated risk-adjustment scoring system for colorectal surgery (colorectal POSSUM). Br J Surg 2004; 91(9):1174-1182.

(39) Tran Ba LP, du Montcel ST, Duron JJ, Levard H, Suc B, Descottes B et al. Elderly POSSUM, a dedicated score for prediction of mortality and morbidity after major colorectal surgery in older patients. Br J Surg 2010; 97(3):396-403.

(40) Lloyd H, Ahmed I, Taylor S, Blake JR. Index for predicting mortality in elderly surgical patients. Br J Surg 2005; 92(4):487-492.

(41) Alves A, Panis Y, Mantion G, Slim K, Kwiatkowski F, Vicaut E. The AFC score: validation of a 4-item predicting score of postoperative mortality after colorectal resection for cancer or diverticulitis: results of a prospective multicenter study in 1049 patients. Ann Surg 2007; 246(1):91-96.

(42) van Walraven C, Wong J, Bennett C, Forster AJ. Derivation and internal validation of the Procedure Independent Mortality Risk (PIMR) Score: an index to quantify independent risk of death after in-hospital procedures using administrative data. 2010. Unpublished work.

(43) Receiver Operating Characteristic (ROC) Curves.: SUGI 31, 2007.

(44) Hosmer EWJr, Lemeshow S. Applied Survival Analysis: Regression Modeling of Time to Event Data. John Wiley and Sons, Inc., 1999.

(45) Daly L. Simple SAS macros for the calculation of exact binomial and Poisson confidence limits. Comput Biol Med 1992; 22(5):351-361.

(46) Sauerbrei W, Meier-Hirmer C, Benner A, Royston P. Multivariable regression model building by using fractional polynomials: Description of SAS, STATA and R programs. Computational Statistics & Data Analysis 2006; 50(12):3464-3485.

(47) A Step-by-Step Guide to Survival Analysis (SUGI): 2008.

(48) Bellera CA, MacGrogan G, Debled M, de Lara CT, Brouste V, Mathoulin-Pelissier S. Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer. BMC Med Res Methodol 2010; 10:20.:20.

(49) Bellera CA, MacGrogan G, Debled M, de Lara CT, Brouste V, Mathoulin-Pelissier S. Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer. BMC Med Res Methodol 2010; 10:20.:20.

(50) Bellera CA, MacGrogan G, Debled M, de Lara CT, Brouste V, Mathoulin-Pelissier S. Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer. BMC Med Res Methodol 2010; 10:20.:20.

(51) Sauerbrei W, Royston P. Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. Journal of the Royal Statistical Society: Series A (Statistics in Society) 1999; 162:71.

(52) Sauerbrei W, Meier-Hirmer C, Benner A, Royston P. Multivariable regression model building by using fractional polynomials: Description of SAS, STATA and R programs. Computational Statistics & Data Analysis 2006; 50(12):3464-3485.

(53) Division of Biomedical Statistics and Informatics. Locally Written SAS Macros. Mayo Clinic . 2010. (Electronic)

(54) May S, Hosmer DW. A simplified method of calculating an overall goodness-of-fit test for the Cox proportional hazards model. Lifetime Data Anal 1998; 4(2):109-120.

(55) Wang Y. 4-29-2010. Personal Communications.

(56) Allison PD. Survival Analysis Using SAS: A Practical Guide. Cary, NC: SAS Institute Inc., 1995.

(57) Analysis of Survival Data with Clustered Events.: SAS Global Forum 2009, 2009.

(58) de Bruijne MH, le Cessie S, Kluin-Nelemans HC, van Houwelingen HC. On the use of Cox regression in the presence of an irregularly observed time-dependent covariate. Stat Med 2001; 20(24):3817-3829.

(59) Daly L. Simple SAS macros for the calculation of exact binomial and Poisson confidence limits. Comput Biol Med 1992; 22(5):351-361.

(60) Daly L. Simple SAS macros for the calculation of exact binomial and Poisson confidence limits. Comput Biol Med 1992; 22(5):351-361.

(61) Daly L. Simple SAS macros for the calculation of exact binomial and Poisson confidence limits. Comput Biol Med 1992; 22(5):351-361.

(62) Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996; 15(4):361-387.

(63) Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996; 15(4):361-387.

(64) Bellera CA, MacGrogan G, Debled M, de Lara CT, Brouste V, Mathoulin-Pelissier S. Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer. BMC Med Res Methodol 2010; 10:20.:20.

(65) Bellera CA, MacGrogan G, Debled M, de Lara CT, Brouste V, Mathoulin-Pelissier S. Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer. BMC Med Res Methodol 2010; 10:20.:20.

(66) Young W, Williams JI. A summary of studies on the quality of health care administrative databases in Canada. In: Canadian Medical Association, editor. The ICES Practice Atlas - Patterns of Health Care in Ontario. Toronto: 1997: 339-345.

(67) de Bruijne MH, le Cessie S, Kluin-Nelemans HC, van Houwelingen HC. On the use of Cox regression in the presence of an irregularly observed time-dependent covariate. Stat Med 2001; 20(24):3817-3829.

(68) van Walraven C, Dhalla IA, Bell C, Etchells E, Stiell IG, Zarnke K et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. CMAJ 2010; 182(6):551-557.

(69) Canadian Institute for Health Information. DAD Abstracting Manual, 2008-2009 Edition. 2008. Ottawa, CIHI.

(70) Escobar GJ. 2009. Personal Communication.

(71) Division of Biochemistry TOH. Pathology and Laboratory Medicine Division of Biochemistry Policy and Procedure Manual: Establishing/Validating Reference Intervals. 1-11-2005.

(72) Aust JB, Henderson W, Khuri S, Page CP. The impact of operative complexity on patient risk factors. Ann Surg 2005; 241(6):1024-1027.

(73) Shah KB, Kleinman BS, Rao TL, Jacobs HK, Mestan K, Schaafsma M. Angina and other risk factors in patients with cardiac diseases undergoing noncardiac operations. Anesth Analg 1990; 70(3):240-247.

(74) Detsky AS, Abrams HB, McLaughlin JR, Drucker DJ, Sasson Z, Johnston N et al. Predicting cardiac complications in patients undergoing non-cardiac surgery. J Gen Intern Med 1986; 1(4):211-219.

(75) Goldman L, Caldera DL, Nussbaum SR, Southwick FS, Krogstad D, Murray B et al. Multifactorial index of cardiac risk in noncardiac surgical procedures. N Engl J Med 1977; %20;297(16):845-850.

(76) Kumar R, McKinney WP, Raj G, Heudebert GR, Heller HJ, Koetting M et al. Adverse cardiac events after surgery: assessing risk in a veteran population. J Gen Intern Med 2001; 16(8):507-518.

(77) Larsen SF, Olesen KH, Jacobsen E, Nielsen H, Nielsen AL, Pietersen A et al. Prediction of cardiac risk in non-cardiac surgery. Eur Heart J 1987; 8(2):179-185.

(78) Sullivan LM, Massaro JM, D'Agostino RB, Sr. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. Stat Med 2004; 23(10):1631-1660.

8. Abbreviations

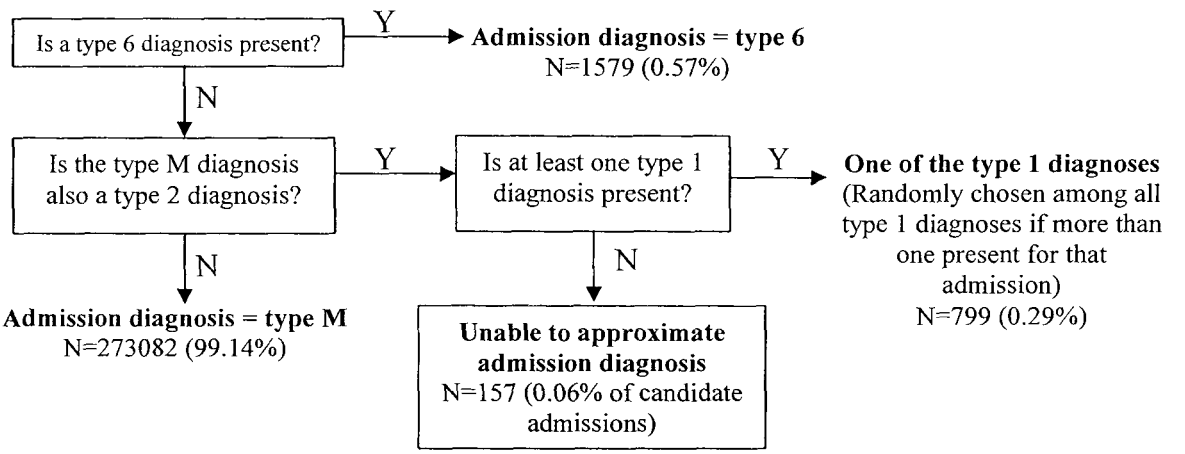
Abbreviation	Full Description
AIC	Akaike's Information Criterion
ALC	Alternative level of care
AMA	Acute monitoring area
CCI	Canadian Classification of Interventions
CI	Confidence interval
COPS	*CO*morbidity *P*oint *S*core
DAD	Discharge Abstract Database
HR	Health Records
ICU	Intensive care unit
IQR	Interquartile range
KPMCP	Kaiser Permanente Medical Care Program
LAPS	*L*aboratory-based *A*cute *P*hysiology *S*core
MIS	Management Information Systems
NACRS	National Ambulatory Care Reporting System
PIMR (score)	*P*rocedure *I*ndependent *M*ortality *R*isk (score)
SAS	Statistical Analysis Software
TOH	The Ottawa Hospital
TOHDW	The Ottawa Hospital Data Warehouse
WID	Warehouse identifier

9. Appendices

Appendix A. Algorithm used to approximate the admission diagnosis

To approximate the admitting diagnosis, I used the algorithm presented in Figure A1. This algorithm compared the diagnosis types recorded for each admission (Table A1).

Figure A1. Algorithm to approximate the admitting diagnosis

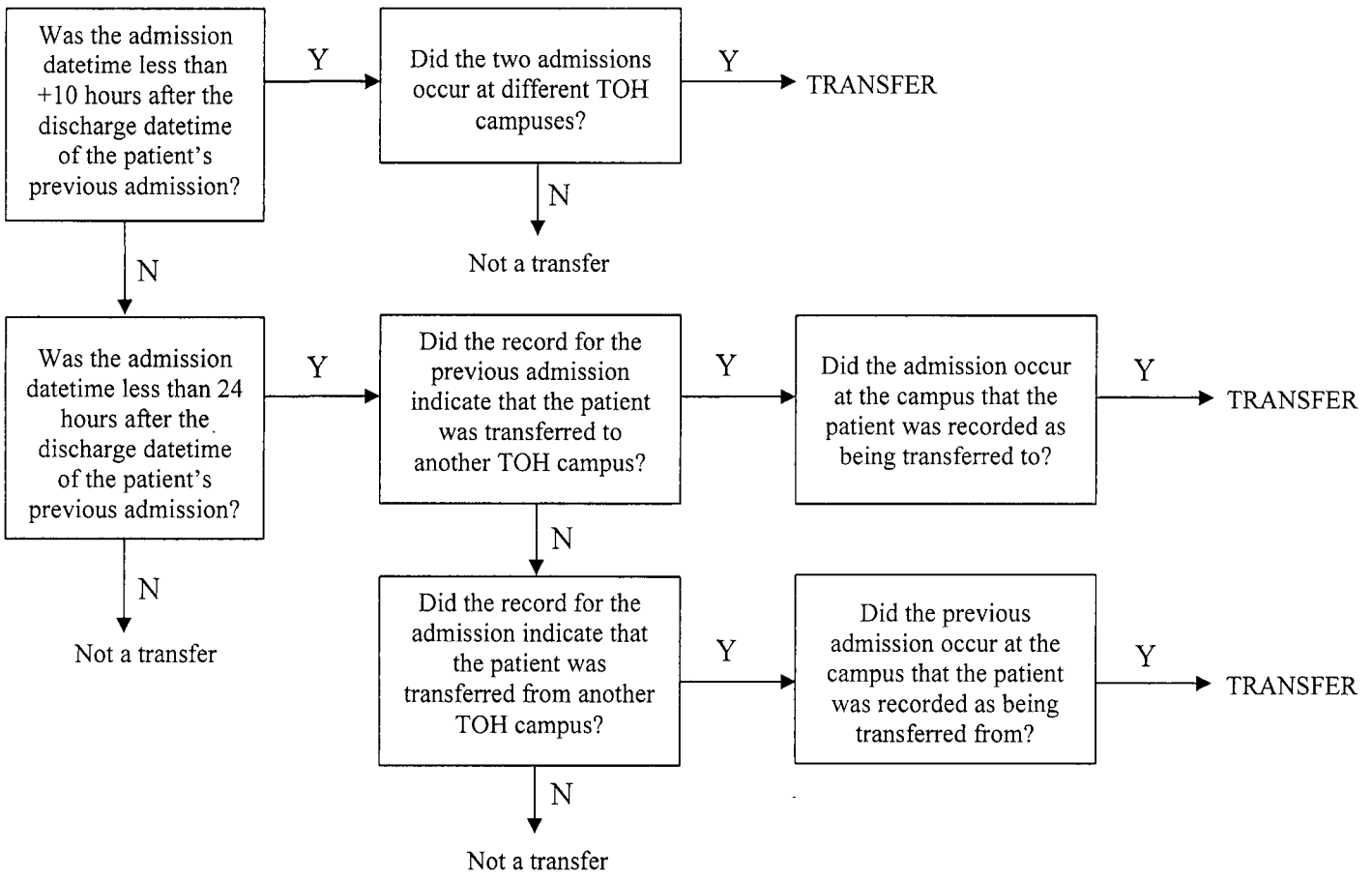


N = number of admissions; % = proportion of study admissions (unless otherwise noted)

Table A1. Definition of diagnosis types used in the algorithm

Diagnosis type	Definition (69)	Frequency of diagnosis type
"M"	"Most responsible diagnosis": the diagnosis most responsible for the patient's stay in hospital, or if more than one, the diagnosis responsible for the greatest use of resources	One and only one recorded per admission
"6"	"Proxy most responsible diagnosis": used when the *manifestation* of an underlying condition is most responsible for an admission. The underlying condition is coded as type M. For example, consider a patient admitted for type 2 diabetes mellitus (coded as "M") whose hospital stay was primarily due to its manifestation diabetic angiopathy (coded as "6").	Rarely used
"1"	"Pre-admit comorbidity": a condition that was present pre-admission	0 to multiple may be recorded per admission
"2"	"Post-admit comorbidity": a condition that arises post-admission (i.e. a complication)	0 to multiple may be recorded per admission

Appendix B. Algorithm used to identify interhospital transfer admissions



Appendix C. Covariates in the Laboratory Score Preliminary Model

Covariate	Categories (70)
Age (years)	< 45
	45-59
	60-64
	65-69
	70-74
	75-79
	80-83
	84-88
	89-92
	93+
Admission type	Emergent surgical
	Emergent non-surgical
	Elective surgical
	Elective non-surgical
Ratio of blood urea nitrogen to creatinine	0-<6
	6-<10
	10-<13
	13-<25[†]
	25+
Ratio of anion gap[*] to serum bicarbonate	0-<575
	575-<625[†]
	625+
Square of deviation from normal sodium range[**]	0-4[†]
	5-9
	10-15
	16-48
	49-120
	121+

[*] Anion gap serum sodium – (serum chloride + serum bicarbonate)
[**] Normal sodium range 135-145 mEq/L
[†] Normal range

Appendix D. Two-step method to standardize laboratory test results

1. Standardization of lab test results between TOH campuses and over time

In the "NlabService" table, a "normal range" is reported with every test result unless the test is a blood gas test (pH, $PaCO_2$, and PaO_2 tests) or was performed on a paediatric patient (<18 years old). The normal range represents the 2.5^{th} to 97.5^{th} percentile for readings produced by the particular machine (or assay kit) on patient laboratory tests. The range is equivalent to either the published reference range for the assay kit or the reported normal range from studies done by the manufacturer and is internally validated by the hospital lab. If no published reference range or manufacturer data exist, then the normal range is established in house by a formal reference interval study (71).

To standardize lab results between the hospital's campuses and over time, I did the following:

1. *Determined a standard TOH normal range for each laboratory test type*

   For laboratory tests in LAPS, I ran a query of all test results in the "NlabService" table obtained between 1 January 2009 and 21 October 2009 and used the most commonly reported normal range as the hospital's standard normal range.

2. *Determined the midpoint of the standard TOH normal range and the midpoint of the normal range for each individual test*

3. *Multiplied each test result by the ratio of the standard midpoint to the individual test's midpoint*

   For example, test "X" for serum albumin would be standardized by the following equation:

$$\text{TOH-standardized result} = \frac{\text{test result x midpoint of standard TOH normal range for albumin}}{\text{midpoint of normal range for test "X"}}$$

Note1: pH, $PaCO_2$, and PaO_2 tests could not be standardized because a normal range was not reported for these tests.

2. Standardization of lab results to the Kaiser Permanente Medical Care Program (KPMCP)

Since the LAPS scoring scheme was derived using laboratory results from KPMCP hospitals, I attempted to standardize TOH lab results to KPMCP lab results by the following:

1. *Determined a standard KPMCP normal value for each laboratory test in LAPS*

   The standard KPMCP normal value was the midpoint of the range that assigned 0 points to LAPS. For tests where the range assigned to 0 points did not have an upper and lower bound, the upper or lower limit was used.

2. *Determined the corresponding standard TOH normal value.*

   If the standard KPMCP normal value was a midpoint, the midpoint of the standard TOH normal range was used as the standard TOH normal value. If the standard KPMCP normal value was an upper or lower limit, the upper or lower limit of the standard TOH normal range was used as the standard TOH normal value, respectively.

3. *Multiplied each TOH-standardized result by the ratio of the standard KPMCP normal value to the standard TOH normal value for that test type.*

Note2: I did not KPMCP-standardize albumin tests because the lower bound of the range assigned to 0 points was lower than the conventionally-recognized normal range for albumin. Similarly, I did not KPMCP-standardize bilirubin tests because the upper bound of the range assigned to 0 points was higher than the conventionally-recognized normal range for bilirubin. I also could not KPMCP-standardize bicarbonate and chloride tests because these tests were only used in the Laboratory Score Preliminary Model and were not assigned points to LAPS.

Appendix E. The Procedure Independent Mortality (PIMR) Index

Derivation of the PIMR Index

To identify candidate medical procedures for the PIMR Index, I retrieved the procedure codes recorded by health records analysts for all admissions in the study cohort. During the study period, TOH coded procedures using the Canadian Classification of Interventions (CCI) coding system. I limited candidate procedures to therapeutic procedures (CCI section 1), and categorized procedures into 1992 different groups using the first five alpha-numerics of the code (which identifies the anatomical area and intervention type). I determined the urgency of each procedure using the admission status of the hospitalization (i.e. elective or emergent) because for certain procedures (i.e. surgeries), urgency status has been found to be an important, independent predictor of post-procedure outcomes (72-77). For procedures that could not be performed electively (including cardiac resuscitation, implantation of an internal device in the thoracic descending aorta, and control of bleeding in the thoracic cavity), I defaulted the urgency of all procedures to "emergent" regardless of the admission urgency.

I used multivariable binomial logistic regression to derive the index with in-hospital death as the model outcome. I derived the index on a randomly selected 50% of admissions from this study. In this sample, there were 4013 deaths. The model could therefore only test a maximum of 400 procedures (10 deaths per procedure) to avoid problems with over-fitting and model instability.

Since the total number of observed procedure-urgency combinations exceeded 400, I applied three filters to reduce the number of candidate procedure-urgency combinations offered to the multivariable logistic model. First, I only considered procedures performed on the day of the principal procedure (defined as the procedure considered by the health records analyst to be most significant during the admission).

Second, I excluded rare procedures, which I defined as any procedure performed less than once per month during the study period regardless of urgency status. Third, I excluded procedure-urgency combinations where the *p*-value of the univariable association with in-hospital mortality was less than 0.5 after adjusting for the Escobar covariates (using LAPS on the day of the procedure for admissions with the procedure-urgency combination of interest). After applying these filters, 212 procedure-urgency combinations (comprising 168 unique procedures) remained.

The initial multivariable model included 212 binary variables (one for each procedure-urgency combination) and the Escobar-predicted risk of death as the adjusting covariate. I used stepwise variable selection and a significance level of 0.05 to retain procedure-urgency combinations significantly and independently associated with in-hospital death. The final multivariable model included 56 procedure-urgency combinations, comprising 52 unique procedures (Table E1). Of these, 37 emergent and eight elective procedures were significantly and independently associated with an increased risk of in-hospital death, while four emergent and seven elective surgeries were associated with a decreased risk.

Using the methods described by Sullivan *et al.* (78), I modified the parameter estimates from the final multivariable model into a scoring index. I determined the number of points assigned to each procedure-urgency combination by dividing its coefficient by the coefficient in the model with the smallest absolute value, and rounding the quotient to the nearest integer. The resulting score for individual procedure-urgency combinations ranged from -7 to +11 (Table E1).

Table E1. Procedure-urgency combinations included in the PIMR Index

Variable	5-digit CCI code	Parameter Estimate	Adjusted Odds Ratio	(95% CI)	Points
Escobar-predicted risk of death	-	1.03	2.79	2 73 - 2.86	-
Emergent Procedures					
Resuscitation, heart NEC	1HZ30	4.26	70.72	41.04 - 121.84	11
Excision partial, ventricle	1HP87	3.71	40.91	5 03 - 333.05	10
Repair, aortic valve	1HV80	2.91	18 27	4.90 - 68 19	8
Immobilization, shoulder joint	1TA03	2.95	19 18	1.55 - 236.57	8
Repair, patella	1VP80	2.81	16.60	1.61 - 171.01	7
Repair, tricuspid valve	1HS80	2.38	10.85	1 77 - 66.34	6
Implantation of internal device, thoracic [descending] aorta	1IC53	2.28	9.82	5 73 - 16 84	6
Occlusion, abdominal arteries NEC	1KE51	2.44	11.53	4.67 - 28.49	6
Implantation of internal device, abdominal cavity	1OT53	2.43	11.33	2 21 - 58.17	6
Excision partial, soft tissue of the chest and abdomen	1SZ87	2 46	11.76	2.38 - 58 01	6
Drainage, ventricles of brain	1AC52	1.88	6.54	3.26 - 13.11	5
Drainage, bronchus NEC	1GM52	1.77	5.89	1 33 - 26 20	5
Ventilation, respiratory system NEC	1GZ31	1.89	6.65	5.73 - 7.71	5
Installation of external appliance, heart NEC	1HZ37	2.00	7.38	2.72 - 20.01	5
Stimulation, heart NEC	1HZ09	1.54	4 68	2.89 - 7.56	4
Extraction, arteries of leg NEC	1KG57	1.47	4.34	2 38 - 7.93	4
Bypass, small intestine	1NK76	1.49	4.43	1.95 - 10.04	4
Repair, small intestine	1NK80	1 35	3 86	1.91 - 7 80	4
Drainage, meninges and dura mater of brain	1AA52	1.23	3.43	2.15 - 5.48	3
Excision partial, brain	1AN87	1.33	3.77	1 98 - 7.20	3
Control of bleeding, thoracic cavity NEC	1GY13	1.06	2.90	1 37 - 6.12	3
Drainage, pericardium	1HA52	1.01	2.74	1.18 - 6.37	3
Occlusion, vena cava (superior and inferior)	1IS51	1.25	3.50	1.60 - 7 65	3
Control of bleeding, esophagus	1NA13	1.14	3 12	1.46 - 6 71	3
Dilation, esophagus	1NA50	1.09	2.98	1 18 - 7.54	3
Control of bleeding, small and large intestine	1NP13	1.10	3.00	1.22 - 7 38	3
Amputation, tibia and fibula	1VQ93	1.03	2.80	1.27 - 6.21	3
Bypass with exteriorization, trachea	1GJ77	0.58	1.78	1.12 - 2.84	2
Implantation of internal device, stomach	1NF53	0.61	1.84	1.25 - 2.70	2
Excision partial, small intestine	1NK87	0.78	2 18	1.33 - 3.58	2
Excision partial, large intestine	1NM87	0.58	1.79	1.20 - 2 66	2
Drainage, abdominal cavity	1OT52	0.67	1.96	1.43 - 2.69	2
Implantation of internal device, hip joint	1VA53	0.82	2.26	1.59 - 3.22	2
Fixation, femur	1VC74	0.76	2 13	1 52 - 2.99	2
Amputation, femur	1VC93	0.92	2.51	1.20 - 5.24	2
Drainage, pleura	1GV52	0.56	1.74	1.36 - 2.24	1
Implantation of internal device, vena cava (superior and inferior)	1IS53	0.38	1.46	1.20 - 1 78	1
Pharmacotherapy (local), vessels of heart	1IL35	-0 68	0.51	0.31 - 0 85	-2
Excision total, appendix	1NV89	-1.13	0 32	0 12 - 0 88	-3
Installation of external appliance, circulatory system NEC	1LZ37	-1.39	0.25	0.17 - 0.36	-4

(Table E1 cont'd)

Variable	5-digit CCI code	Parameter Estimate	Adjusted Odds Ratio	(95% CI)	Points
Excision partial, abdominal cavity	1OT87	-2 35	0.10	0 01 - 0.72	-6
Elective Procedures					
Drainage, pericardium	1HA52	3.37	29.16	6.28 - 135.28	9
Radiation, pelvis	1SQ27	3.35	28.47	2 58 - 314.61	9
Destruction, skin of abdomen and trunk	1YS59	2 75	15.61	1.89 - 128.67	7
Excision partial, cerebellum	1AJ87	2 17	8.75	2.01 - 38.11	6
Pharmacotherapy (local), circulatory system NEC	1LZ35	2 29	9 92	2 26 - 43.51	6
Repair, abdominal arteries NEC	1KE80	1.75	5 77	1.36 - 24.54	5
Amputation, femur	1VC93	1.92	6.81	1.93 - 24.01	5
Ventilation, respiratory system NEC	1GZ31	0 83	2.28	1.66 - 3.14	2
Dilation, coronary arteries	1IJ50	-1.61	0.20	0 06 - 0.69	-4
Implantation of internal device, hip joint	1VA53	-1.57	0 21	0 07 - 0 65	-4
Implantation of internal device, knee joint	1VG53	-1.58	0 21	0 08 - 0 55	-4
Excision total, ovary with fallopian tube	1RD89	-2.07	0.13	0.02 - 0 91	-5
Repair, muscles of the chest and abdomen	1SY80	-2.00	0.14	0 05 - 0.36	-5
Excision partial, prostate	1QT87	-2.24	0 11	0.01 - 0 76	-6
Excision total, uterus and surrounding structures	1RM89	-2.51	0 08	0 01 - 0.59	-7

*The adjusted odds ratio represents the increase in risk for every 10% increase in Escobar-predicted death risk

Assessment and validation of the PIMR Index

I calculated the total PIMR score on the day of the principal procedure by summing the points associated with all procedure-urgency combinations for which it had been coded. Most admissions had a total score of 0 since the majority (84%) of admissions had none of the PIMR procedures performed (Figure E1).

I determined the additional predictive ability of the total PIMR score by comparing the discrimination and calibration of two logistic models: one including the Escobar covariates only, and one including the Escobar covariates plus the total PIMR score. For both models, I applied the coefficients obtained from the derivation cohort to the validation cohort. I assessed discrimination and calibration of the models in the validation cohort using the *c*-statistic and Hosmer-Lemeshow statistic, respectively.

The model including the total PIMR score had better predictive ability than the model including the Escobar covariates only (Table E2). The *c*-statistic of the model increased significantly, and the model calibration improved.

Figure E1. Distribution of the total PIMR score

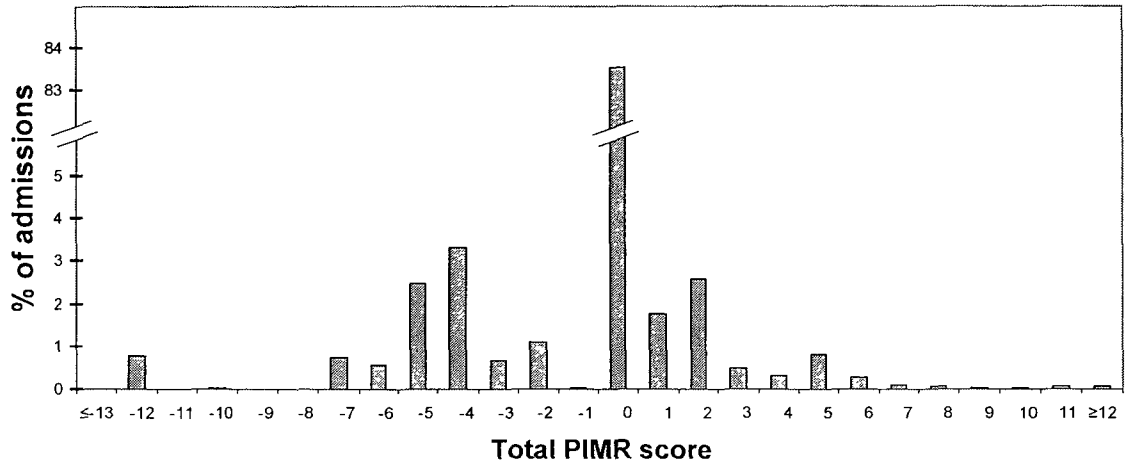
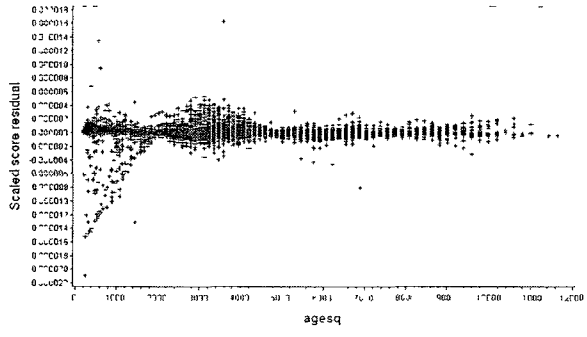
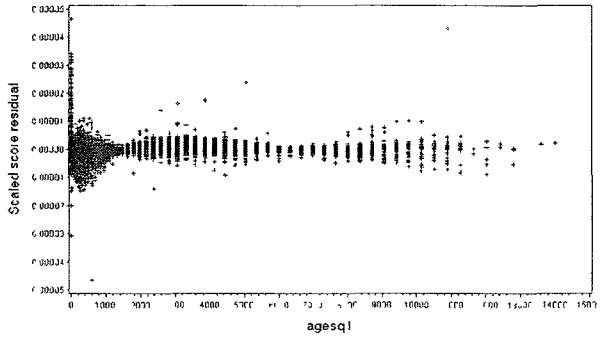


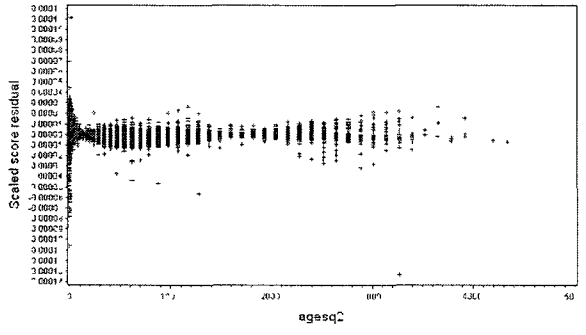
Table E2. Performance of the Escobar model with and without the PIMR score

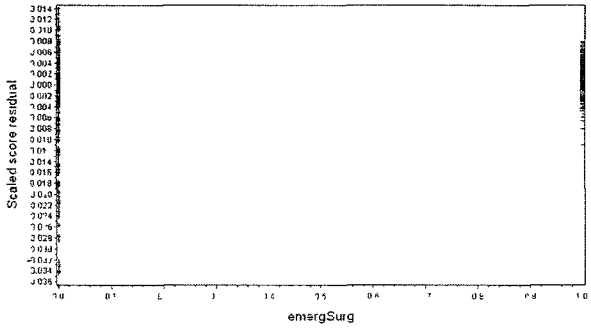
	Escobar covariates only	Escobar covariates + total PIMR score
c-statistic (95% CI)	0.929 (0.926-0.932)	0.938 (0.935-0.941)
Hosmer-Lemeshow statistic (*p* value)	37.56 (<.0001)	36.51 (<.0001)

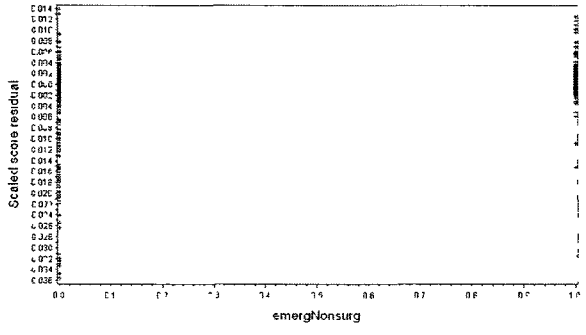
Appendix F. Scaled score residual plots for covariates in the Escobar^{+} model

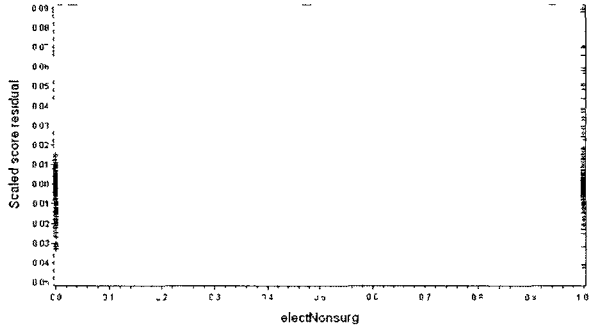


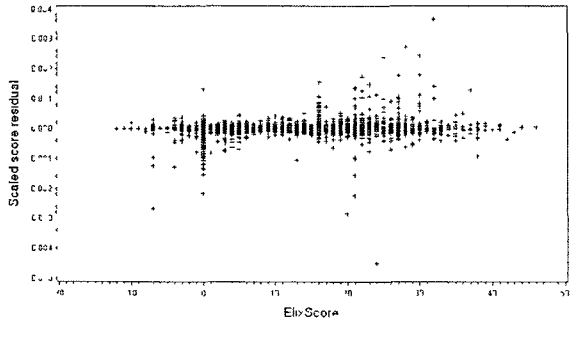


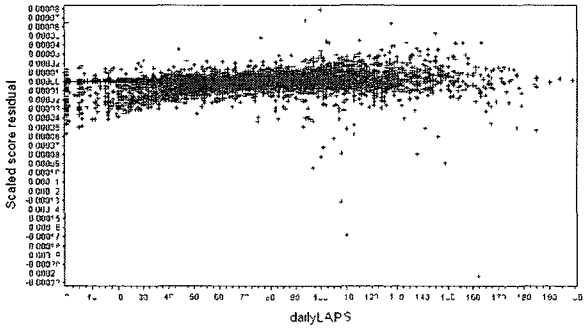


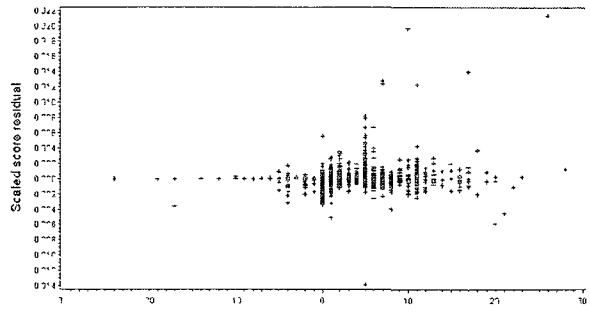


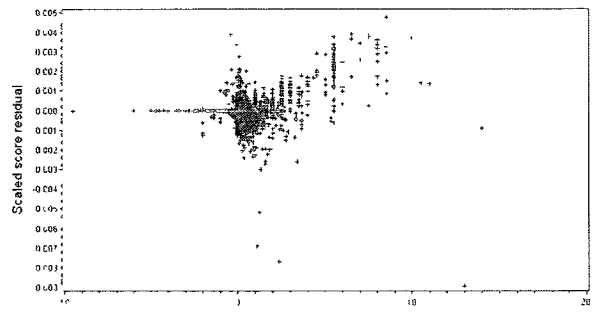


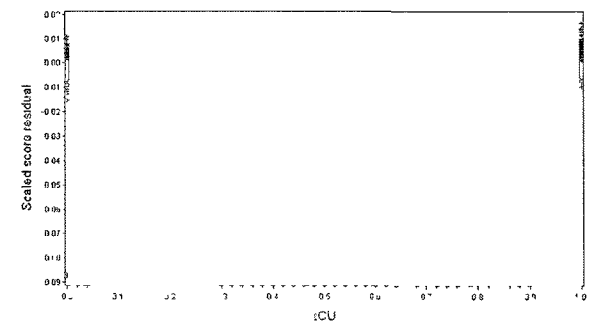


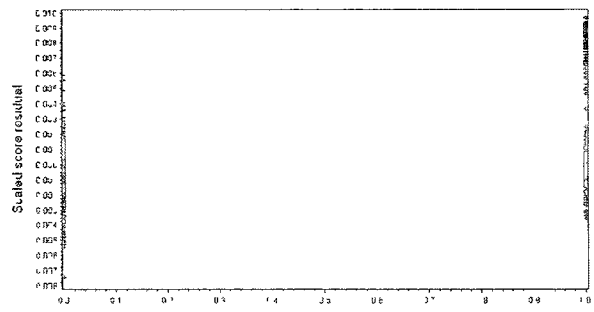


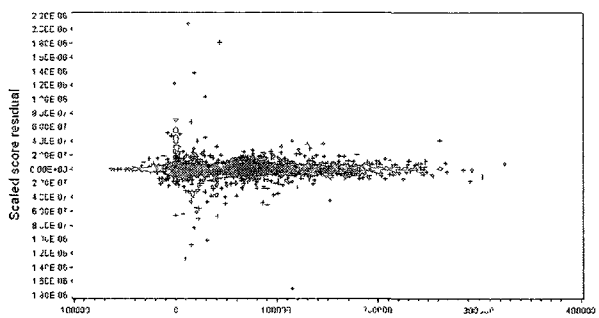


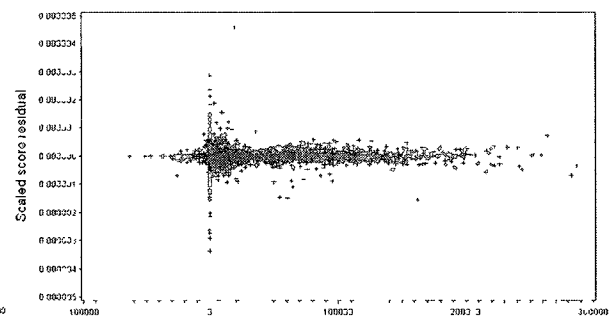

Scaled score residual
PIMR

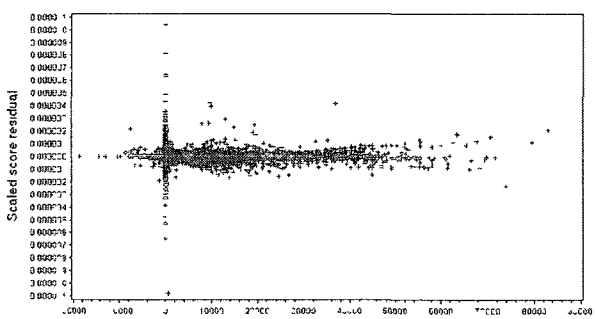

Scaled score residual
PIMR*InvDaysSincePIMR

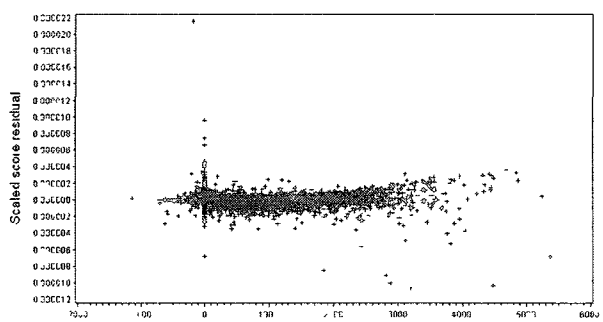

Scaled score residual
tCU

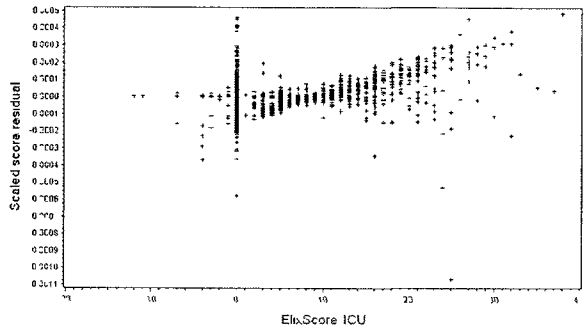

Scaled score residual
ALC

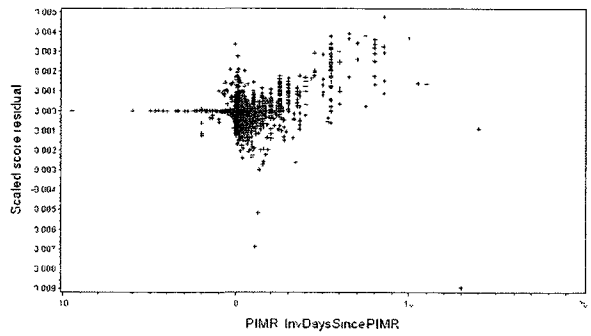

Scaled score residual
agesq*ElixScore

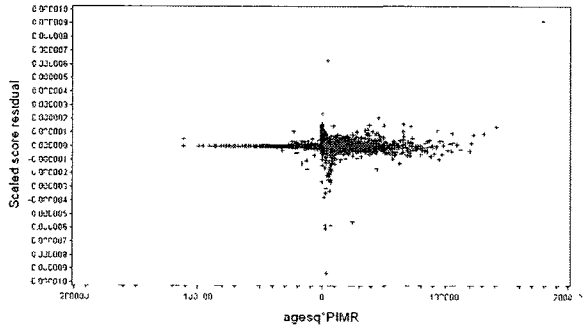

Scaled score residual
agesq1*ElixScore

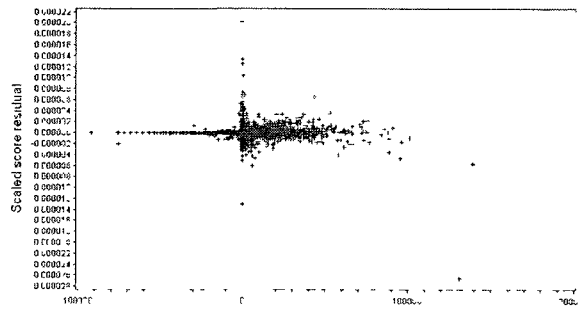

Scaled score residual
agesq2*ElixScore

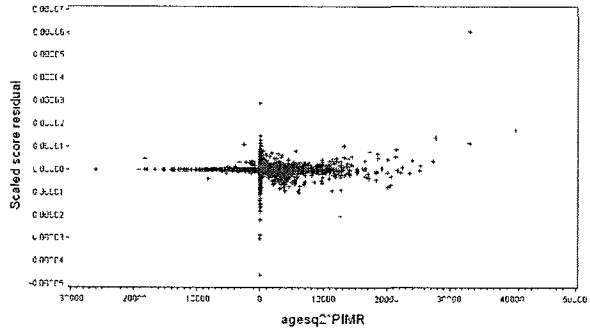

Scaled score residual
ElixScore*LAPS

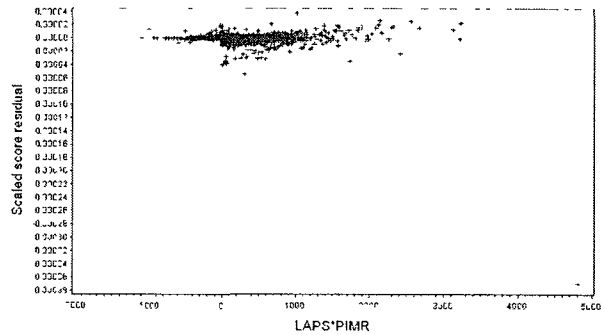

Scaled score residual
ElixScore ICU

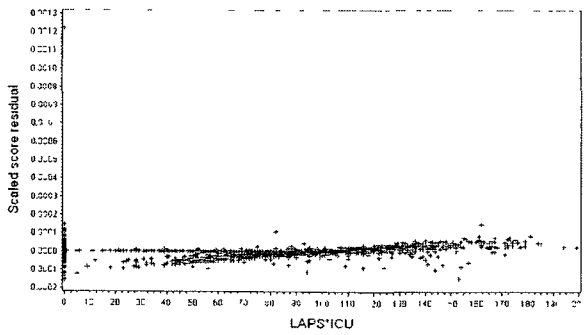

Scaled score residual
PIMR InvDaysSincePIMR

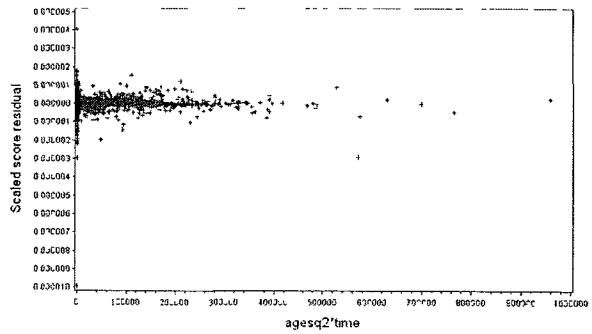

Scaled score residual
agesq*PIMR

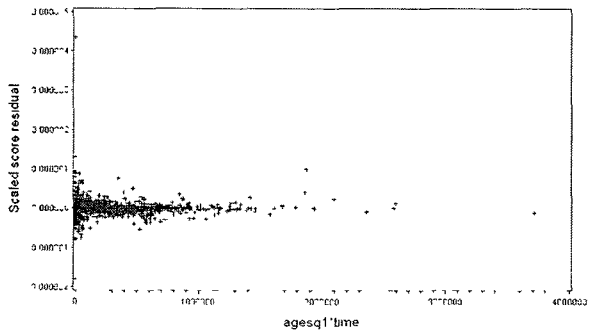

Scaled score residual
agesq1 PIMR

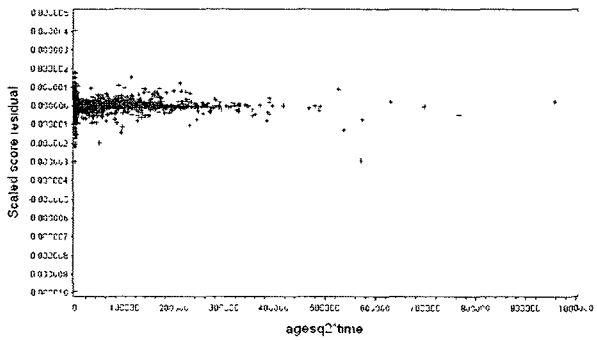

Scaled score residual
agesq2*PIMR

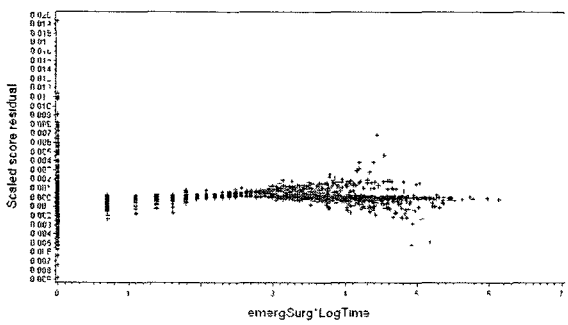

Scaled score residual
LAPS*PIMR

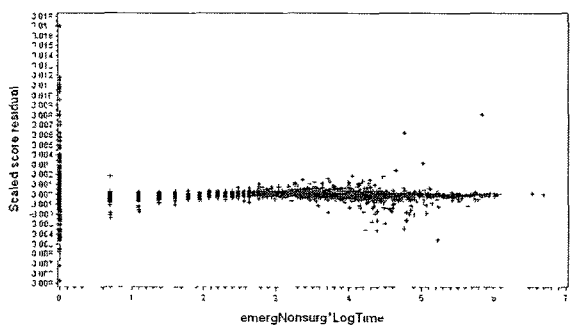

Scaled score residual
LAPS*ICU

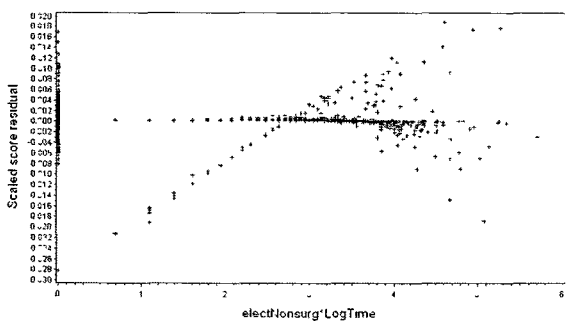

Scaled score residual
agesq2*time

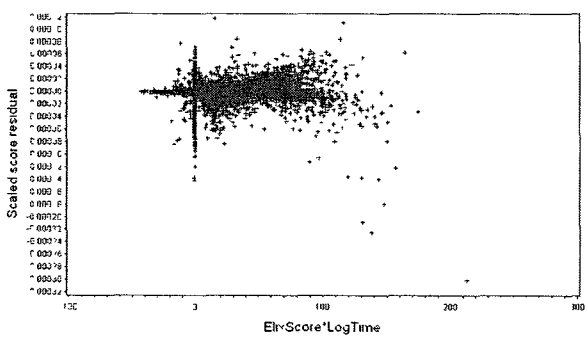

Scaled score residual
agesq1*time


Scaled score residual
agesq2*time


Scaled score residual
emergSurg*LogTime


Scaled score residual
emergNonsurg*LogTime


Scaled score residual
electNonsurg*LogTime


Scaled score residual
ElixScore*LogTime