University of Calgary

A comparison of comorbidity scores to predict 1 year mortality in critically ill adult patients

by

Susan Quach ©

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF COMMUNITY HEALTH SCIENCES

CALGARY, ALBERTA

NOVEMBER, 2008

UNIVERSITY OF
CALGARY

ABSTRACT

Introduction: This study explored the predictive ability of the Charlson and Elixhauser, as a possible risk adjustment method used in an ICU study sample. The study also examined whether Charlson and Elixhauser contained all comorbidities that were associated with mortality in the ICU population.

Methods: Study participants include adult patients admitted to ICU in the Calgary Health Region between April 2000 and March 2004. Clinical data was collected prospectively and linked to hospital discharge data. Logistic regression analysis was used to compare the performance of Charlson index and Elixhauser with APACHE II. A Delphi process was used to elicit information from ICU physicians about comorbidities

Results: The original Charlson index and Elixhauser had adequate ability to predict one year mortality in ICU patients (C=0.70 to 0.74). The addition of other variables, age and sex, could substantially improve predictive power (C=0.69 vs. C=0.77). Charlson and Elixhauser comorbidity lists are complete, with no other comorbidities missing for risk adjustment in the ICU.

Conclusion: The Charlson index and Elixhauser can be used for risk adjustment of comorbidities in the ICU population.

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my supervisor, Dr. Chip Doig, for giving me this opportunity. I really appreciate his inspiration, guidance, and encouragement during this entire process. You provided me with an opportunity to grow academically and professionally with this experience and most importantly, you believed in me. For this, I'm truly grateful to work with you and value all my accomplishments during my graduate studies. I would also like to acknowledge the intellectual contributions and mentorship that my committee members provided- Dr. Hude Quan, Dr. Peter Faris, and Dr. Brenda Hemmelgarn. With your guidance, I learned new skills and knowledge that will prepare me for challenges in my career. Thank you.

I would also like to thank my friend, Deirdre Hennessy, who introduced me the wonderful world of risk adjustment. I will always admire your knowledge and passion for research, in hopes that one day we can work together again. You are a great research buddy and personal life coach, and I will definitely miss you.

To my best friend, Dr. Andrei Ordine, who spent much time listening about the adventures of the thesis and providing constructive feedback on the thesis. I really appreciate your moral support and your willingness to always help me. Thank you for saving me from bootstrapping.

I would also like to acknowledge all my new friends that I made during the time I lived out in Calgary. I arrived in Calgary without knowing anyone and your companionship made it enjoyable to live in a new city.

Finally, I owe a great deal to my parents, "Mommy and Upba", who provided much encouragement and support to help me achieve my goals. You sacrificed a lot for me and dedicated your lives to make me who I'm today. For this, I'm truly grateful and cherish your love.

DEDICATION

In loving memory of my sister, Tara Quach, who passed away after a difficult battle in the ICU. Despite losing her health, she never gave up mentally and taught me to always fight on. I will always cherish our time together as sisters and best friends. Thank you for always being there for me and giving me the strength and determination to live life at its best.

Tara Quach
(July 24 1983 – July 6 2005)

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREIVATIONS

AIC	Akaike's information criterion
AIDS	Auto-immune deficiency syndrome
APACHE	Acute Physiology and Chronic Health Evaluation
APS	acute physiology score
C stat	C statistic
C.cob	Charlson Comorbidities
D. cob	Delphi Comorbidities
CHP	chronic health points
CHR	Calgary Health Region
CI	Confidence Interval
CV ICU	cardio vascular intensive care unit
E.cob	Elixhauser Comorbidities
Exp	Exponent
GERD	gastroesophageal reflux disease
GI	Gastrointestinal
ICD	International Classification of Disease
ICU	intensive care unit
ICNARC	Intensive Care National Audit & Research Centre
IP	in-patient
IQR	Inter-quartile range
LOS	length of stay (days)
LR	likelihood ratio tests
MPM	Mortality Prediction Model
OR	Odds ratio
P	Probability
Pre-ICU	Pre intensive care unit
RR	relative risk
SAPS	Simplified Acute Physiology Score
SD	standard deviation
TISS	Therapeutic Intervention Scoring System

COMPLETE LIST OF MATHEMATICAL MODELS

Model	Description	Equation
CI1	Charlson index original weights	Charlson Index= $w_1*(c.cob_1)+ w_2*(c.cob_2)+\ldots\ldots +w_n*(c.cob_n)$
CI2	Charlson index original weights (scores greater than 6 were combined into one group)	Charlson Index= $w_1*(c.cob_1)+ w_2*(c.cob_2)+\ldots\ldots + w_n*(c.cob_n)$
CI3	Charlson Index with new regression coefficients used instead of odds ratios	Charlson Index= $\gamma_1*(c.cob_1) + \gamma_2*(c.cob_2) + \ldots\ldots + \gamma_n*(c.cob_n)$
C1	Charlson comorbidities entered as dummy variables	Log (p/1-p)= $\beta 0+ \beta 1*(c.cob_1)+ \beta 2*(c.cob_2)+\ldots \beta n*(c.cob_n)$
E1	Elixhauser comorbidities entered as dummy variables	Log (p/1-p)= $\beta 0+ \beta 1*(e.cob_1)+ \beta 2*(e.cob_2)+\ldots \beta n*(e.cob_n)$
Univariate Models		
CI1-CI3	Comparison between Charlson index variations in univariate models	Log (p/1-p)= $\beta 0+ \beta 1$ *(CI)
A1	APACHE II	Log (p/1-p)= $\beta 0+ \beta 1$ *(APACHE II)
A2	Acute physiology status of APACHE II	Log (p/1-p)= $\beta 0+ \beta 1$ *(APS)
A3	Chronic health points of APACHE II	Log (p/1-p)= $\beta 0+ \beta 1$* (CHP)
D1	Comorbidity identified by Delphi process	Log (p/1-p)= $\beta 0+ \beta 1$* (d.cob)
Multivariate Models		
M1	Baseline model	Log (p/1-p)= $\beta 0+ \beta 1$ *(age) + $\beta 1$*(sex)
M2	CHP is the chronic health points of APACHE II (M1+ CHP)	Log (p/1-p)= $\beta 0+ \beta 1*(age)+ \beta 2*(sex) + \beta 3$ *(CHP)
M3	CI1 Charlson index original weights (M1+ CI1)	Log (p/1-p)= $\beta 0+ \beta 1*(age) + \beta 2*(sex) + \beta 3$ *(CI1)
M4	CI3 Charlson index using new regression coefficients (M1 + CI3)	Log (p/1-p)= $\beta 0+ \beta 1*(age) + \beta 2*(sex) + \beta 3$ *(CI3)
M5	Charlson index using dummy variables coded for comorbidities (M1+ M5)	Log (p/1-p)= $\beta 0+ \beta 1*(age) + \beta 2*(sex)+ \beta 3*(c.cob_1)+ \beta 4*(c.cob_2)+\ldots \beta n*(c.cob_n)$
M6	Interaction of Charlson index and age (M1+ M6)	Log (p/1-p)= $\beta 0+ \beta 1(age) + \beta 1(sex) + \beta 3$ (CI1) + $\beta 4$ (age*CI1)
M7	Elixhauser using dummy variables coded for comorbidities (M1+ M7)	Log (p/1-p)= $\beta 0+ \beta 1*$ (age) + $\beta 1*(sex) + \delta_1*(e.cob_1)+\ldots.+ \delta_n*(e.cob_n)$
D2	Comorbidity identified by Delphi process	Log (p/1-p)= $\beta 0+ \beta 1*(e.cob_1)+ \beta 2*(e.cob_2)+\ldots \beta n*(e.cob_n)+ \beta n*(d.cob_n)$

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW

Over the past decades, there has been a growing demand for critical care services in Canada. Previous research has shown that utilization of intensive care services increased between 1969 and 1986 in both Canada and the United States(1). Within-hospitals, specialized intensive care units are one of the most costly types of care offered. For example, they accounted for 15.9% of inpatient direct expenses in Ontario acute care hospitals between 1999-2000 and 2003-2004, but only 8.1% of inpatient days(2). Given the increase in demands from an aging population that could escalate the costs of new expensive drugs and technology, critical care services need to continually show that they are effective for achieving optimal long term outcomes.

Comparisons between ICUs is challenging since patient groups may vary due to a number of system level and patient factors. Therefore, it is essential to have risk adjustment methods that compare outcomes and variations in case-mix between intensive care units (ICUs). In the ICU, risk adjustment has been primarily focused on severity of illness models such as Acute Physiology and Chronic Health Evaluation (APACHE) score, the Simplified Acute Physiology Score (SAPS) or the Mortality Prediction Model (MPM)(3-8). These models were developed to predict short term mortality outcomes such as in hospital mortality. Although these systems had excellent discrimination (C stat> 0.80) in their development, subsequent evaluations have suggested that the calibration statistics have deteriorated over time, and vary by case-mix. Because of the change in calibration, either these systems require frequent recalibration for given populations, or health services researchers must accept the changes in calibration. These systems require sophisticated data collection which has limited their uptake across ICUs. Finally, risk adjustment across different scores or different versions of the same score further limit the widespread applicability of these scores as methods of risk adjustment. Therefore, there is a need to seek alternative methods.

The Charlson index and Elixhauser comorbidity scores were developed on non ICU patients but have been used widely to measure and control for comorbidities in a variety of study populations(9-18). These comorbidity scores can be easily obtained in administrative data sources and Elixhauser does not require re-weighting of regression coefficients. Both comorbidity scores were developed to predict 1 year mortality, which

may be more suitable indicator for risk adjustment in studies interested in longer term outcomes than hospital mortality. However, few studies have compared how well the Charlson index or Elixhauser performs for predicting long term mortality in ICU patients(19-23) and whether it can be used in this population.

Since Elixhauser and Charlson method were not developed for the ICU population, the purpose of the study is to explore the predictive ability of the Charlson and Elixhauser in this population. Further investigation will determine whether there are comorbidities missing in Charlson and Elixhauser that can improve risk adjustment in the ICU population.

1.0. The Intensive Care Unit

Over the past decades, there has been a growing concern over health care costs and the quality of health care received in the hospitals. In 2007, Canadian hospitals accounted for about 28% of total health spending (forecast to be 10.6% of the gross domestic product [GDP]) (2)Further, the demand for critical care medicine has been evolving since the 1960s and now remains a major component of hospital services provided in Canada. Previous research shows that utilization of intensive care services increased between 1969 and 1986 in both Canada and the United States(1). As critical care has become more common, a range of specifically designated types of special care units has emerged to care for patients. Medical and surgical intensive care units, for example, remain the most common, accounting for 45% of admissions of critical care units in 2003-2004(2). These specialized intensive care units are one of the most costly types of care offered. For example, they accounted for 15.9% of inpatient direct expenses in Ontario acute care hospitals between 1999-2000 and 2003-2004, but only 8.1% of inpatient days(2). The types of treatments provided in critical care will need to be evaluated carefully to justify spending in a system that has a growing elderly population and escalating drug costs.

This leads to the issue with the appropriateness of prolonging life, in light of the conflict between patient and societal costs. As discussed in the 2002 Brussels Roundtable report: Surviving Intensive Care, there is a growing demand for evidence to examine whether intensive care survivors have optimal long term outcomes and whether ICU care

decisions would change if physicians knew more about the outcomes(24). Critical illness is associated with a wide range of serious and long term events that can affect patient-centered outcomes. Because ICU patients are a heterogeneous group, many illness scores have been developed over the years to evaluate the effectiveness of intervention and outcomes of critical care services in research and clinical practice. Health outcome disparities between different providers have encouraged providers and physicians to continually monitor differences, in order to determine how providers can offer a high quality of care.

1.1. Risk adjustment in the ICU: Background and Theory

Accurate comparisons between critical care services require risk adjustment—a method to account for patient-associated factors—before comparing the outcomes from different patients, treatments, providers, health plans and population (25). This allows health care providers to estimate any variation in outcome explained by patient characteristics, also known as case-mix. Without risk adjustment, the effect of one variable on outcome may be inaccurately estimated because of an inability to account for the effect of case-mix; this is known as confounding. A confounder is defined as an observed association due to mixing of effects between risk factor, the outcome, and a third factor that is associated with both the exposure and is not in the causal pathway between the risk factor and outcome (26). As a consequence, failure to account for confounders can be a problem for interpreting the results of a study and can distort the conclusions of the study, which is a common problem in observational studies. Methods to control for confounding would include matching, stratification, restriction or randomization. The most common method used for adjustment of confounding in non-randomized studies is stratification in the analysis by multivariable adjustment(27). Multivariable adjustment allows multiple confounders to be entered into a model to provide an effect that is “adjusted” for these known confounders. This requires confounders to be easily and accurately measured. This can be a problem in retrospective studies when confounders, such as ICU severity scores, are not collected consistently across all ICUs. Regardless, observational studies are often used first for exploring research questions before engaging in randomized control trials (RCT) to

examine more specific questions. Although RCTs allow researchers to control for confounding by randomization, there are many challenges with this type of study design in ICU research. First, RCTs are costly to conduct since they require personnel to enroll patients into a trial and conduct extensive follow-up. RCTs may be inappropriate or unethical in cases where patients are randomized to receive detrimental interventions. In situations where health care providers are interested in examining ICU's performance, this cannot be done by RCT. In the ICU, it is difficult to conduct randomized studies in patients who are at a high risk of dying. Further, RCTs may be inadequate when patients in a trial represent a very narrow subset of ICU patients. Finally, RCTs lack generalizability because strict inclusion criteria are used to generate the study sample. Because of these limitations associated with RCTs, risk adjustment appears to be suitable option to control case mix so that appropriate comparisons of interventions can be made.

Risk adjustment appeared in the early 1950s when Virginia Apgar published a simple physiological scoring tool to evaluate newborn children(28). Although the tool was only based on two physiological systems: cardiopulmonary and central nervous function, it contributed to the development of other risk adjustment methods. During the late 1980s and early 90s, researchers developed severity of illness scores to adjust for patient prognostic factors such as age, physiological disturbances, principal diagnoses and pre existing conditions(5, 6, 29). These models assigned weights or points to factors based on their observed statistical associations with the outcomes. Using the points, these outcome prediction models were initially thought to provide health care providers and patients with an estimated probability of survival for counseling and decision making. However, the utility for this purpose were no longer accepted due to statistical limitations with the accuracy and calibration of these models (3, 8, 30). For example, there have been problems with "over-fitting" the data by developing models with too many predictor variables based on small sample sizes. Other issues involve the predictive accuracy of using population based models to predict outcomes for individuals. For example, these models perform poorly for the extreme cases of the population (i.e. high risk of death or low risk of death) because of small sample sizes.

Regardless, the primary utility of these models now in ICU research is for risk adjustment in observational and outcome comparison studies.

1.1.1. Severity illness scores in the ICU

The three most widely used taxonomies in the ICU are the Acute Physiology and Chronic Health Evaluation (APACHE) score, the Simplified Acute Physiology Score (SAPS) or the Mortality Prediction Model (MPM) (3, 6, 8). These systems are mostly based on the severity of physiologic derangement during a discreet period of the patient's ICU stay and have been evaluated against short term mortality outcomes such as ICU or hospital mortality. Although the methods for selecting the predictive variables are different, all of the models used logistic regression techniques to derive a model relating the predictor variables to the probability of the outcome. Each system also includes points for underlying chronic health status or comorbidity. Table 1 shows a list of comorbidities in each ICU scoring system.

Table 1. Comparison of comorbidities between different ICU scoring systems

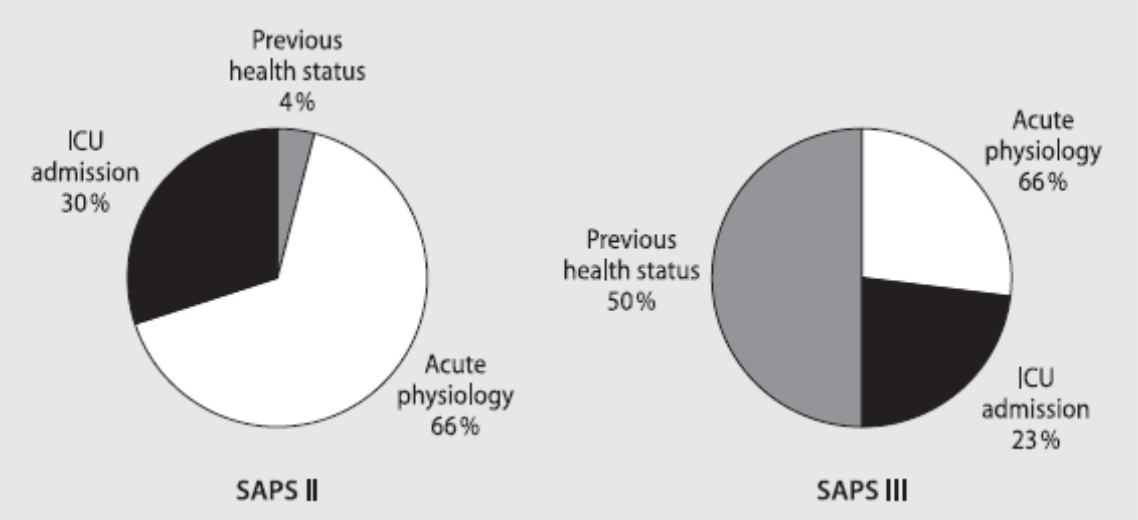
ICU scoring system	Comorbidities	Features/Weight
APACHE II	Organ insufficiency Immuno-compromised state **Liver:** • cirrhosis and portal hypertension • Past upper GI bleeding attributed to portal hypertension • Prior episodes of hepatic failure/encephalopathy/coma **Cardiovascular:** • New York Heart Association Class IV Unable to carry out any physical activity without discomfort. Symptoms of cardiac insufficiency at rest. If any physical activity is undertaken, discomfort is increased **Respiratory:** • Chronic restrictive , obstructive or vascular disease resulting in severe exercise restriction	If patient has a history of severe organ system insufficiency or is immuno-compromised state, assign the points: 5 points- for non operative or emergency postoperative patients 2 points- for elective post operative patients

	• Chronic hypoxia • Hypercapnia • Secondary polycythemia • Severe pulmonary hypertension • Respiratory dependency **Renal:** • Receiving chronic dialysis **Immuno-compromised:** • Received therapy that suppresses resistance to infection • Has disease that is sufficiently advanced to surprises resistance to infection • Leukemia, lymphoma, AIDS	
APACHE III	**Points** AIDS 23 Hepatic Failure 16 Lymphoma 13 Metastatic cancer 11 Leukemia/multiple myeloma 10 Immunosuppression 10 Cirrhosis 4 None/not available 0	If there is more than one comorbidity, then select the one with the highest value. Select for non operative or emergency surgical patients only, otherwise 0.
APACHE IV (2004)	• AIDS • Cirrhosis • Hepatic failure • Immuno-suppression • Lymphoma • Leukemia or myeloma • Metastatic tumor	Not used for elective surgery patients Unique relative contribution of each risk factor to hospital mortality prediction (5% for chronic health status)
SAPS II (1993)	• Metastatic carcinoma • Hematological malignancy • AIDS	**AIDS** HIV positive with clinical complications as *pneumocystis carinii* pneumonia, Kaposi's sarcoma, Lymphoma, tuberculosis or toxoplasma infection. **Hematologic malignancy** lymphoma, acute leukemia, or multiple myeloma.

<table>
<tr><td></td><td></td><td>Metastatic cancer
Metastasis identified by surgery, C.T. scan or any other method.</td></tr>
<tr><td>SAPS III (2005)</td><td><table>
<tr><td>Cancer therapy</td><td>3</td></tr>
<tr><td>Chronic heart failure(New York Heart Association Class IV)</td><td>3</td></tr>
<tr><td>Hematological cancer</td><td>3</td></tr>
<tr><td>Cirrhosis</td><td>8</td></tr>
<tr><td>AIDS</td><td>8</td></tr>
<tr><td>Cancer</td><td>11</td></tr>
</table></td><td>Cancer therapy is defined as Chemotherapy, Immunosupression other, Radiotherapy, Steroid treatment</td></tr>
<tr><td>MPM II (1993)</td><td>• Metastatic neoplasm
• Cirrhosis</td><td>stage IV carcinomas with distant metastases</td></tr>
<tr><td>MPM III (2005)</td><td>• Chronic renal
• Chronic GI
• Metastatic neoplasm
• Acute renal
• Dysrhythmia
• Cerebrovascular incidence
• GI bleeding</td><td></td></tr>
<tr><td>ICNARC (2007)</td><td>• Chronic renal replacement
• Chronic cardiovascular disease
• Chronic respiratory disease
• Chronic liver disease
• Immune-comprised</td><td></td></tr>
</table>

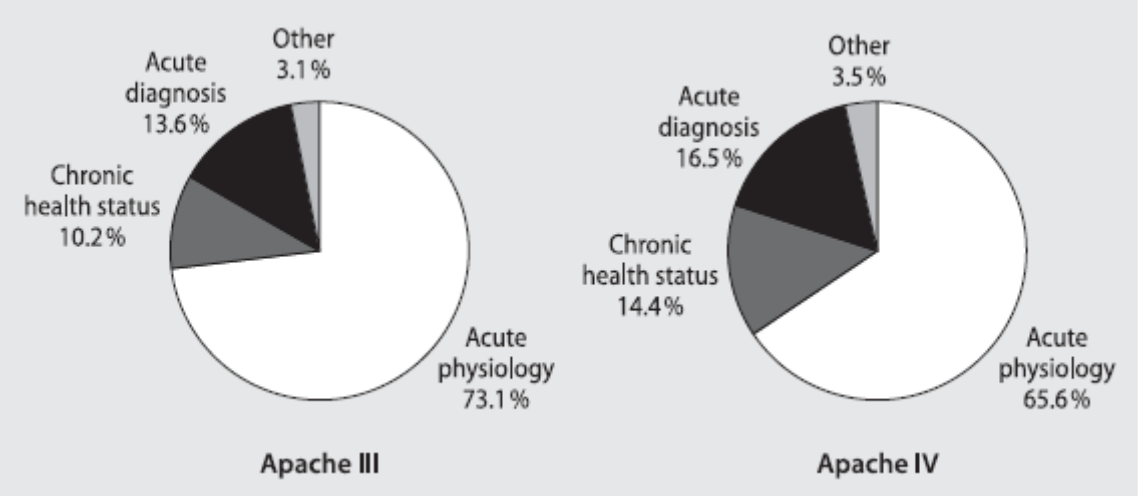
During the evolution of these models, there has been a gradual shift away from physiological derangement variables to a greater emphasis on prior health status and physiological reserves. The explanatory power of chronic health status, circumstances of ICU admission and degree of physiologic derangement at ICU admission is presented for SAPS II & III and APACHE III & IV in Figure 1 and Figure 2. The impact of chronological age was collapsed on the chronic health status.

Figure 1: Explanatory Power for SAPS II & III



Adapted from data published in(8)

Figure 2: Explanatory Power for APACHE III & IV



Adapted from data published in (4, 31)

1.1.2. Acute Physiology and Chronic Health Evaluation (APACHE)

The original Acute Physiology and Chronic Health Evaluation (APACHE)score was one of the first physiology-based scoring systems to predict in-hospital mortality for ICU patients(5). The original APACHE system provided weighting for 34 potential physiologic measures, which sum to an acute physiology score (APS). In addition, the original APACHE incorporated a four letter code to correspond to a range of chronic diseases. Because the original APACHE was complex, APACHE II was developed to refine the model(3). APACHE II is composed of three components: physiological variables, age, and chronic health points (CHP). Further, information about major disease category (reason for ICU admission) allows physicians to estimate the probability of in-hospital death using a multiple logistic regression equation. In APACHE II, the score varies between 0 to 71 points; with 60 points for physiological or laboratory variables, 6 points for age, and up to 5 points for chronic health points. Within the physiological component, the model consisted of the most abnormal value of 12 physiological variables recorded within the first 24 hour after admission into the ICU. Patients with a history of severe organ system insufficiency or who were immuno-compromised received points based on their surgical status. Only severe organ system failures found in liver, cardiovascular, respiratory and renal systems are captured in the chronic health points. Poses et al., and Pittet et al., both found that the chronic health points component of APACHE II were either not statistically significant or did not contribute discriminative ability for predicting mortality in their studies(20, 32).

The initial validation of APACHE II was done in 13 hospitals to predict hospital mortality(3). Since, many studies worldwide have validated the APACHE II model for predicting in-hospital mortality in different patient disease populations such as stroke, cancer, pneumonia, acute MI, congestive heart failure, liver cirrhosis (33-37). Some studies have suggested that APACHE II has moderate performance (C stat<0.80) for predicting mortality outside the ICU setting (36), and may not be suitable for predicting long term outcomes such as 1 year mortality. Since then, new versions of APACHE III & IV have been developed providing slightly better prediction but require more sophisticated data collection and calculation(4, 38). For example, APACHE III includes five new physiological measurements, and the chronic health points are modified to

include seven variables, with up to 23 points assigned. APACHE IV has incorporates the use of spline terms for the acute physiology variable, age and prior length of stay which requires more sophisticated data analysis techniques. In addition, APACHE III & IV includes information on patient's prior location before ICU admission. However, using APACHE III & IV requires intensive data collection and is complex, and these limitations have hindered the uptake of newer APACHE score systems. As a result, APACHE II has remained the most widely used severity illness score system in the ICUs.

Appendix 1: APACHE II calculation formula

1.1.3. Limitations of severity illness scores

Over the years, it was realized that the performance of severity of illness scores began to deteriorate with time in respect to calibration. In the UK, the Intensive Care National Audit and Research Center used a database with 141 106 patients from 164 adult general ICUs in England, Wales, and Northern Ireland to compare the published versions of APACHE II UK, APACHE III, SAPS II, and MPM II(39). For patients admitted during December 1995 to August 2003, they showed that the models had good discrimination but imperfect calibration. These results suggest that there may have been gradual changes in baseline characteristics of admitted patients, circumstances for ICU admission and availability of general and specific therapeutic measures which has contributed to an increasing gap between actual mortality and predicted mortality(40, 41).

Using severity of illness scores require much work to continually update the model in order to maintain good calibration. Poor uniformity of fit within ICU subgroup has also been reported as a common problem in severity of illness scores(41). Another limitation in using severity scores relate to difficulty in comparing ICUs when different scores and versions are collected. For example, APACHE has been updated four times since its development, while there are several versions of SAPS and MPM(8, 42). A study conducted by Wunsch et al,. on 127 adult general ICUs (N=120 503) found that there were differences in-hospital mortality before and after exclusion criteria was applied to APACHE II (-3.1% to 9.5%); similarly, other severity illness scores also had differences in-hospital mortality(43). When they did a review on studies that used APACHE II, they found wide variation in the exclusion criteria reported. These studies

show that the exclusion criteria were inconsistently applied and may introduce biases into risk adjustment studies. As a result of these limitations in severity illness scores, there is need for alternative methods to risk adjust in the ICU population.

1.2. Overview of comorbidity scores

One of the earliest comorbidity measurement tools was developed by Kaplan and Feinstein in 1974. This score classified diabetic patients from no comorbidity to severe comorbidity (grade 0 to grade 3) and was developed based on clinical agreement of the importance of comorbidities(44). Following this, Charlson et al., developed a method in 1987 that uses a single summary score based on the presence of 17 comorbidities and takes into account the severity of the comorbidity(10). Since then, researchers have modified the original Charlson index by changing the weights(45, 46), adding other clinical variables (11, 22, 23), excluding non relevant comorbidities(45, 46) to suit a specific study population and outcome, in hopes of improving the performance. Other methods have relied on the total count of the comorbidities or the assessment of each of the comorbidity in a model (22, 47, 48). Specifically, Elixhauser incorporates 32 comorbidities in a model and has been validated on many study samples(17, 47, 49, 50). Some studies have used outpatient pharmacy dispensing data to assign patients to chronic disease groups. In these scores, an integer weight was given to each comorbidity category represented by medication classes, which were summed to produce a total score(51). Compared to comorbidity scores derived from administrative data (52-54), prescription medication based scores were consistently worse at predicting 1 year mortality(55).

1.3. Factors that influence a comorbidity score's predictive performance

The predictive performance of comorbidities and its use as a risk adjustment method will depend on several factors. First, it will be important to consider the comorbidities included in the score and their relative weights which can affect the predictive performance of the score. Next, it is important to consider the distribution and comorbidity profile of the source population. For instance, a comorbidity score that predicts well in one disease population may perform poorly in a different disease

population if it is missing important comorbidities that are relevant in the second population.

In addition, comorbidity scores will vary according to the outcome predicted. While most comorbidity scores were developed to observe their association with mortality, outcomes such as length of stay, hospital readmissions, complications, or resource use are also correlated with mortality and their performance will vary depending on how well the measures relate to the original outcome. A review on comorbidity studies(56) showed that comorbidities were a significant predictor of mortality in studies with a long follow up period (1 month or longer) compared to short time period, even after adjustment for other covariates(10, 45, 56). Evidence from this review showed that comorbidities may have a smaller influence on mortality when the index disease was lethal or when the explanatory model included a number of clinical variables that were associated with the index disease(56). Because researchers assume that comorbidities should correlate with poor health outcomes, the validity of comorbidity score is assessed by how well the scores predict related outcomes. In turn, this determines how well the measure can control for confounding. This is measured by examining discrimination and calibration for the predicted outcome. In earlier research, the validity of prediction was determined by the strength of association between the comorbidity score and the outcome(10, 33, 46, 57).

Finally, the predictive performance of a comorbidity score will depend on accuracy of the data source for comorbidities(58). For example, studies have consistently shown that comorbidities derived from chart data provide better predictive performance than comorbidities collected through administrative data(14, 20, 21, 59). Minor comorbidities are seriously under-coded in administrative data and this phenomenon has been well documented as a source of coding bias. The effect of this phenomenon may cause some comorbidities to be viewed as protective(60).

1.4. Charlson index

Originally, the Charlson index was developed to predict 1 year survival in medical patients admitted into a teaching hospital and further validated on a cohort of breast cancer patients(10). This index contained 17 comorbidities which were weighted and

summed to produce a score. Many studies have validated the Charlson index on different patient populations to predict short and long term mortality (12, 15-18, 20, 53, 61). Various adaptations of the index were developed to accommodate different study populations. To derive the comorbidity index, the number and severity of comorbidities at the time of admission were collected from patient charts. In her study, Charlson *et al*, was able to show that the weighted index of comorbidity was a significant predicator ($p<0.0001$) of 1 year survival(10).

Appendix 2: Charlson index score calculation

1.4.1. Modifications to the Charlson index

Many studies have evaluated and validated the Charlson index in different patient populations, and other than 1 year outcome. To improve the performance, studies have readjusted weights, deleted comorbidities, and recoded the Charlson weights into fewer categories. For instance, D'Hoore determined that the Charlson index had a non-linear relationship with the log odds of death and the Charlson Index was based on highly skewed scores. The scores were recoded in the less categories (0, 1-2,3-4,5-6,>6)) for medical inpatients with ischemic heart disease and this was found to improve discrimination(52).

Some research has suggested that deriving empirical weights for a specific study population and outcome measure can improve discrimination and model performance (9, 12, 45, 46). For example, comorbidities such as HIV may be present infrequently in a study population or may have a small impact on outcome and consequently, may be over or under weighted in the Charlson index. As a result, weights can vary according to time, geographic areas, study population, and outcomes(12, 45, 46). To improve the weights, Romano *et al*, has suggested that researches working with large administrative databases should use multivariate models to derive their own study specific weights. Further, this can lead to better control for confounding in studies than using the original Charlson weights.

Ghali *et al*, derived his own empirical weights for CABG surgical patients to predict in mortality(45). Only comorbidities such as myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, chronic pulmonary

disease, diabetes and renal disease were significant in bi-variate analysis. Further, when a multivariate model of the comorbidities was considered, all comorbidities that were significant in bi-variate analysis except chronic pulmonary disease, received a revised study weight from one to four. Using new study weights, Ghali identified a 10% subset of patients with 15% mortality, whereas the 5% highest-risk patients according to the Charlson index had only 8% mortality ($p = 0.01$). This resulted in a model that performed slightly better discrimination than the original Charlson weights ($c= 0.74$ vs. 0.70). Likewise, Romano *et al*, showed that the weights were quite different according to which study group was considered and varied slightly when they compared coding algorithms (Deyo vs. Dartmouth-Manitoba)(54). In particular, when the Manitoba CABG data set was used to predict 1 year mortality, 7/12 Charlson comorbidities were still significant. When Cleaves *et al*, compared hospitalized patients for one of the six medical or surgical procedures (back pain, stroke, pneumonia, hip replacement, transurethral radical prostatectomy, or lysis of peritoneal adhesion), different weights were assigned to comorbidities across the procedures (12).

1.4.2. Deriving empirical weights in the ICU

Charlson identified 30 clinically important comorbidities, of which 19 had conditions with statistically significant hazard ratios for 1 year mortality. In survival analysis, the hazard ratio is the effect of the comorbidity, or explanatory variables, on the hazard or risk of an event, in this case 1 year mortality. In this process, all comorbidities were coded as the present or absent and entered into a Cox's regression model to predict 1 year mortality. Next, these adjusted hazards ratios were calculated by taking the exponential function of the coefficients. These adjusted hazard ratios are known as the weights for different comorbidities. To simplify the score, hazard ratios less than 1.2 were dropped and the rest of the ratios were rounded to the nearest digit. Finally, to calculate the index, the weights were added together for someone with a specific comorbidity profile.

Although many studies have followed this process to derive their own weights using the Charlson index (12, 45, 54, 62, 63), this method has been criticized as incorrect as the method using addition of weights to calculate the index (46, 64, 65). Like odds

ratios, hazard ratios should be multiplied because of the logarithmic properties. Harrell suggests that if the Charlson index were calculated correctly, then the index could have performed better for risk adjustment and prediction. Further, Moon *et al*, has suggested that using original regression coefficients rather than odds ratios to assign overall scores can lead to very different scores or possibly a decrease in calibration and discrimination. This was also supported by Ghali's study where he noted a decrease in discrimination and R-squared values when simple integer weights were used instead of regression coefficients(45). Moon also points out an issue when an odds ratio has a protective or negative effect on the outcome, such as any OR<1.0. In this situation, regression coefficients are negative but would not be accounted for by the given method since there are no weights assigned when odds ratios are smaller than one.

For risk adjustment, Moon and Romano suggest that adding exact regression coefficients would be the most appropriate way to derive comorbidity scores and in turn this would increase statistical accuracy and model performance(46, 65).

1.4.3. Charlson index adaptation in administrative data

Several adaptations of the Charlson index were developed for use in administrative databases. Although the original Charlson index was based on medical chart reviews, most of the diagnoses and procedures could be matched with similar ICD-9 and ICD- 10 codes. (International Classification of Disease, 9^{th} revision & 10^{th} revision)
Deyo *et al*, Romano *et al*, and D'Hoore *et al*, independently defined their own coding algorithms for the Charlson comorbidities, and there are some small differences between the coding schemes(52-54). For example, Deyo and Dartmouth-Manitoba's coding scheme are both based on ICD-9-CM codes and are similar in predictive performance for multivariate models and prevalence of comorbidities(45, 54). The authors noted that Deyo's coding scheme has more limited interpretation than the Charlson comorbidity definitions, while Dartmouth-Manitoba included conceptually similar conditions that were not stated in the original Charlson paper. In contrast, D'Hoore developed their own coding scheme based on hospitalization data and used the first three digits of the ICD-9 diagnosis codes. This limited their ability to distinguish between subgroups of certain clinical conditions (i.e. diabetes with complications vs. without) and to identify

procedures codes(52). Although each of the three coding schemes were developed on different study population to predict different outcomes, the results are similar for measuring the prevalence of comorbidities(45, 54)

When ICD-10 coding system was implemented in 1992 by the World Health Organization, Quan *et al,* defined ICD-10 coding algorithms for Charlson and Elixhauser comorbidities and back-translated the ICD-10 codes to improve Deyo's Charlson and Elixhauser ICD-10-CM coding algorithms(50). In their study, they validated the new coding algorithms to predict in-hospital mortality, with C stats well above 0.80, and showed the frequency of Charlson and Elixhauser comorbidities were consistent across algorithms.

1.4.4. A review of ICU research studies exploring the Charlson index

Few studies have suggested that the Charlson index may perform comparably or better than APACHE II for mortality prediction in an ICU population (20, 66). Poses *et al.*, (20) compared APACHE II's chronic health points (CHP) to the Charlson, and found that the CHP did not have significant discriminative ability for predicting in-hospital death (ROC=0.57, SE=0.05, p=0.19) compared to the Charlson index (ROC=0.67, SE=0.05, p=0.03). Further, when they evaluated the Charlson index added to APACHE II with chronic health points, the improvement in Chi square was significant (chi square= 4.5, p value= 0.03 and borderline significant when the full APACHE II model was considered (p=0.09). Although the authors suggested that the Charlson Index could improve prognostic predictions in ICU patients, the study was conducted in one hospital and the data was collected over a short interval.

A more recent study (23) showed that the Charlson index(OR= 1.167, 1.09 to 1.23) was significantly associated with hospital mortality and length of stay, even after adjusting for age, sex, major clinical category, source of admission and SES (socio-economic status). Likewise, when APACHE II was added into a multivariate model without Charlson index, it was also a significant predictor of in-hospital mortality (OR=1.127, 1.108 to 1.466). When both APACHE II and Charlson index were added to the same multivariate model, both predictors remained statistically significant (OR= 1.125, OR=1.115, respectively). For this study, APACHE II (0.77) had better

discrimination than Charlson (ROC=0.69) for in-hospital mortality. Regardless, this study showed that the Charlson index was useful as a risk adjustment variable in the ICU population. To summarize, few studies have validated the utility of the Charlson index, collected by administrative data, as a risk adjustment method for an ICU population and more studies are warranted(23).

1.5. Elixhauser comorbidity risk adjustment method

A more recent method for assessing comorbidities, developed by Elixhauser (47), used 30 categories of comorbidities to predict in-hospital deaths, length of stay, and hospital charges(47). This risk adjustment method was developed based on adult, non maternal inpatients from 438 acute care hospitals in California in 1992. Unlike the Charlson index, which used medical record review, the Elixhauser was derived based on comorbidities obtained from administrative data. Instead of summarizing the weights into a single index, Elixhauser proposed that the weights for individual comorbidities should be quantified separately to take into account different populations and outcomes. In fact, some studies have found that the Elixhauser model outperformed the Charlson model slightly in predicting mortality(9, 17, 18). Some studies have attempted to compare the performance of Elixhauser to APACHE-weighted comorbidity score (22), Charlson/Deyo method (17, 18), Charlson/Romano (67) to predict in-hospital mortality or time to death, yet studies on 1 year mortality outcomes in ICU patients are lacking. However, the majority of studies on Elixhauser have been conducted in non ICU patients.

1.5.1. A review of ICU research studies exploring Elixhauser

Elixhauser has additional comorbidities not accounted for in Charlson that may be applicable to the ICU population which may improve predictive performance. However, despite Elixhauser being developed approximately 10 years ago, ICUs have been slow to adapt this comorbidity score and there are very few studies that have explored the use of this score in the ICU population for risk adjustment.

Johnston *et al*, conducted a study on more than 17 000 ICU patients across VA hospitals from 1996 to 1997 to examine the performance of APACHE III and Elixhauser for predicting in-hospital mortality. This study demonstrated that models using

Elixhauser comorbidities discriminated better than APACHE III comorbid conditions or a count of Elixhauser comorbidities (C stat= 0.70 vs. C stat= 0.57,C= stat= 0.63 respectively), when considering index and prior hospitalizations. When they combined other clinical predictors such as age, laboratory values, principal diagnosis, and admission source with the Elixhauser comorbidities, this provided additional significant predictive ability (C=0.88).

Recently Ho *et al*, conducted a study on 24 404 ICU admissions in Western Australian, and found that 3615 (14.9%), 10 223 (42.1%), and 11 597 (47.7%) patients had at least one comorbidity as defined in the APACHE II score, Charlson comorbidity index, and Elixhauser comorbidities, respectively(19). Further, they found that the discrimination of different measures of comorbidity alone was poor, with C Stats all less than 0.61. When the chronic health points component of APACHE II was replaced with other measures of comorbidities, it did not improve discrimination further (C stat=0.83) for predicting in-hospital mortality. In this study, the baseline model with APACHE II score without comorbidities or admitting diagnosis had fairly high discrimination on its own (C stat= 0.83).

1.6. Mortality Outcomes

Currently, mortality is the most common outcome assessed in studies comparing ICU performance because it is easy to measure and define. Measuring mortality rates are important for patients, health care providers and the community. Further, it provides an understandable measure of the effectiveness of critical care services. Different endpoints of mortality exist but the correct mortality endpoint would depend on the specific research question, mechanisms and timing of disease or treatment under study(68). Mortality at ICU discharge may be a good indicator for audit within an individual ICU, but it can be misleading since many patients may soon die after discharge or at home or on the ward(68). Although ICU mortality outcome is independent of effects of subsequent hospital care, it is affected by organizational factors such as discharge policies, transfer patterns, and family and physician's expectation which can determine whether the death occurs in the hospital ward or ICU. Therefore, ICU mortality may be

difficult to interpret when making comparisons of the effectiveness of care between units since it is affected by operational factors and case-mix.

Most severity of illness scores focus on short term mortality such as in-hospital mortality but there are also limitations associated with this outcome. Hospital mortality reflects care given after the ICU such as in the wards but it is also affected by the organization and delivery of ICU factors. While hospital mortality may be important during the ICU stay to the attending physician, it is less useful as the patient leaves the ICU and returns back home. During the ICU stay, patients are subjected to risky procedures and interventions that may prolong their life by a few months, but what is more important in the long run is how these interventions affect patient centered outcomes and long term mortality(24). Patient centered outcomes include quality of life, functional status, symptoms, and satisfaction with medical care (69). In the past, there have been interventions that appeared to benefit shot term outcome, but subsequently shown to have no effect or even harmful effects on long term outcome(70). As a result of this, the 2002 Brussels Roundtable report highlighted the need for future ICU clinical trials to include follow up for survival to at least six months, which is consistent with the International Working Party in clinical trials in sepsis (24). One year mortality would be an important outcome to assess since it would be less influenced by ICU practice patterns and also it would be accessible in administrative data. In the past, long term mortality outcome data was difficult and expensive to collect but recently, administrative data has been a valuable source for this type of data.

To understand how ICU care impacts long term mortality, it is important to consider pre-ICU care factors and intra ICU care factors (Table 2). Most ICU studies to date have explored these factors with short term mortality using association analysis. For studies examining long term outcomes, there is some evidence that suggest comorbidities may play a greater role on long term mortality than severity of illness scores. Therefore, exploring the role of comorbidities and how it can be measured in the ICU population may overcome challenges associated with traditional severity illness scores.

Table 2: Pre ICU and Intra ICU factors that impact long term outcomes after critical illness.

Time	Variable
Pre ICU	- *Underlying illness* - Reasons for ICU admission - Pre-ICU management - Access to the ICU
Intra ICU	- Patient course and events - Treatments - Organization - Iatrogenesis and environment - Patient health care interactions - Sleep disturbance and delirium

Adapted in data published in (24)

1.7. **Coding and accuracy in administrative data sources**

In health service research, the chart is the gold standard for comorbidity assessment. Many studies have explored the prevalence of comorbidities in administrative data and chart review to determine how well each method can predict a given set of outcomes(14, 18, 59, 71, 72).

Administrative data is broadly defined as any data collected for reasons other than research. Administrative data is collected from running the health care system, such as enrolling patients in health plans, paying claims, determining reimbursement amounts, certifying coverage, approving expenditures, tracking service utilization and monitoring costs and performance (25). Over the years, administrative databases such as hospital discharge registries or insurance claim data have been increasingly popular data sources for research. For example, administrative data are highly generalizable because they can provide information for all members of the population they represent. The data represents care practiced throughout the community rather than in specialized settings. The data can be linked which can help track individuals over time and across different health care providers. Since the data has been already collected for billing purposes, this source is relatively inexpensive to acquire and are computer accessible (73). However, there may be some disadvantages with administrative data such as the completeness and accuracy of data, such as missing data elements that can identify timing of comorbidities or

complications. Selection bias can also be a problem when the data is related to insurance status or coding bias when it is linked to reimbursement, which is not the case in Canada.

A number of problems can occur in the coding process from when the patient is admitted to when the diagnosis assignment is confirmed. For example, the quality of information at admission, communication among patients and providers, the clinician's knowledge and experience with the illness, and the clinician's attention to detail are all potential sources of errors(71). In addition, many "paper trail" errors can occur especially when written records are transcribed to electronic record. Coder training and experience are variable and coding errors such as misspecification, unbundling and upcoding are common. Misspecification, also known as creep errors, refers to diagnostic assignments that deviate from the governing rules of coding. For example, misspecification occurs when the primary diagnosis or order for tests and procedures is incorrect compared to the evidence found in the medical record. It can occur when generic codes are assigned in light of the fact that there are more specific codes that would be appropriate for the diagnosis. Upcoding occurs when coders assign codes of higher reimbursement value over codes with lesser values. When coders assign codes for all the separate parts of a diagnosis instead of assigning a code for the overall diagnosis, this is known as unbundling(71). In a study done by Quan *et al*, they compared comorbidities derived from ICD-9-CM administrative data with chart data, and they found that 10/19 Charlson comorbidities were under reported in administrative hospital discharge records from the Calgary Health Region, with kappa values ranging from 0.34 to 0.878. Specifically, those comorbidities that are non-specific, non-life threatening, asymptomatic are more often under-coded than serious comorbidities in administrative data (59, 72, 74, 75). A more recent study to determine the presence or absence of 32 conditions and to assess the agreement between ICD-10 data and chart data was conducted on in-hospital patients in four teaching hospitals in Alberta .The validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions was generally the same, though there were slight differences in validity for some conditions(76). Another study that compared the Charlson index derived from medical record data compared to administrative data found that the medical record data was superior with higher C stats for predicting mortality, complications, and length of stay. The mean

number of comorbidities abstracted from chart review was higher (1.31) than the mean for administrative data (0.55) in Medicare beneficiaries that underwent carotid endartectomy in Georgia hospitals(14).Similarly, when comorbidities are underreported in administrative data, this can severely distort the results, either leading to erroneous relationships or very different results for risk adjustment(14, 59). For example, Roos *et al*, using an administrative data source found that in patients with benign prostatic hyperplasia, suprapubic prostatectomy compared to transurethral prostatectomy was associated with an increased 6 year mortality rate after comorbidity adjustment (RR=1.45, 95% CI :1.15- 1.83)(77). However, when Concato *et al*, repeated the same study, except with medical record review, the increase in mortality disappeared (RR=1.03, 95% CI:0.51-2.07) (57, 78). Therefore health service researchers should be careful when drawing conclusions from administrative data and understand the limitations.

Canadian hospital discharge databases are less susceptible to coding biases than American databases because of the single payer government infrastructure. To identify complications, Canadian discharge data contains a single-digit "diagnosis type indicator" with every diagnosis code that flags whether the diagnosis was present at time of hospitalization or after admission. The six indicators are as follows:

- M (most responsible diagnosis)
- Type 1 (primary comorbid condition present at time of admission)
- Type 2 (complications arising after admission)
- Type 3 (secondary comorbid conditions present at time of admission)
- Type 4 (morphology code)
- Type 9 (E – code diagnosis from ICD-9)

However, when Quan *et al*, evaluated the accuracy of the diagnosis-type with a chart review of 1200 inpatient separations, he found that the agreement between the two sources varied greatly for the 12 conditions studied, with kappa values ranging from 0 to 0.72. Most concerning, he found that complications were being miscoded as comorbidities based on low sensitivity for conditions (0 to 57.1%), in other words, the agreement between complication status was poor between chart and administrative data(79).

Finally, other studies have explored the possibility of adding more diagnosis fields or using previous hospital records to capture a more complete profile of comorbidities in a study population (49, 60). Iezzoni found that when the number of available coding spaces for diagnosis was increased, this did not enhance comorbidity assessment(80). Further, using previous hospital record data to collect comorbidities could only yield small improvements in model performance using either Charlson or Elixhauser (22, 49).

1.8. Pilot Study & Rationale

In the pilot study, we compared the ability of the APACHE II, and Charlson index to predict in-hospital mortality for critically ill patients. The study population consisted of adult patients admitted to ICU in the Calgary Health Region between April 1st 2002 and March 31st 2004. In the modelling process, we started with a baseline model containing age, sex, and acute physiology score component from APACHE II (C=0.74). There was minimal difference between a baseline model of acute physiology score (C=0.74) with either CHP (C=0.76) or Charlson index variations added (C= 0.75, 0.76, 0.77). The Charlson index was not a good predictor of mortality on its own(C= 0.63). No improvement occurred when the Charlson index was added to the full APACHE II model (C= 0.808 to C=0.8135). Although the addition of the Charlson index did not improve discrimination substantially, there are still several issues that remain to be explored. First, the Charlson index was developed using 1 year survival data on medical in-patients and its performance may be improved if long term mortality outcomes were considered. Secondly, as Moon and Romano suggested, using the original weights do not accurately reflect the sample and the calculation used to calculate the index was in correct. Therefore, further analysis will compare the original weights with empirical weights to determine if mortality prediction improves. It remains to be explored how Elixhauser performs compared to Charlson index and the differences between the two methods. Finally, both comorbidity tools were developed on non ICU patient populations, and no study has yet to explore whether these comorbidity tools contain all relevant comorbidities for the ICU population. A better understanding of how Elixhauser and

Charlson index performs in the ICU will provide information on which method is more appropriate for risk adjustment.

Appendix 3: Pilot Study Results

1.9. **Summary**

To date, much of the work done in ICU risk adjustment has involved using severity illness scores such as APACHE, SAPS, and MPM. Although ICU physicians are familiar with these scores, there are many issues that exist which remain a challenge for risk adjustment. First, these scores require complicated data collection and calculation which can be labor intensive and expensive. There is a constant need to re-estimate the coefficients because the performance of the models deteriorates with time. This leads to a variety of severity scores in different versions collected throughout ICUs around the world, which can make comparisons between ICUs difficult. Further, these scores have strict exclusion criteria which if applied inappropriately, can affect the outcome results. As recent evidence suggested, there has been gradual shift away from short term outcomes and it has become more important to evaluate how critical care services impact a patient's long term survival. Despite this movement, the most up to date ICU severity score's are still only able to provide prognostic information on in-hospital mortality; an outcome that becomes less important with time for the majority of ICU survivors.

Currently, there are several comorbidity assessment scores that predict long term mortality outcomes. The Charlson index and Elixhauser can be derived easily from administrative data, do not have exclusion groups, and are used widely across different patient populations. In addition, Elixhauser does not require coefficients to be re-estimated which can make it more adaptable for use. In spite of this, ICU research on these scores has been scarce and the only existing studies focus on in hospital mortality(19, 20, 22, 66). Further, in these studies it is unknown whether the comorbidity lists are complete for the ICU population. Therefore, this is the first study to validate the Charlson index and Elixhauser in an ICU sample, in hopes of providing researchers and physicians a tool to gain a better understanding of the impact of critical care services on long term survival.

CHAPTER 2: RESEARCH QUESTIONS

Objective 1: To examine the impact of the Charlson index on 1 year predicted mortality in ICU patients.

Primary Research Questions:

- How well does the Charlson index predict 1 year mortality in ICU patients?

Secondary Research Questions:

- How much is gained for mortality prediction when you add Charlson or comorbidities with age and sex?
- How does the effect of comorbidity on mortality depend on age?
- How do the new ICU specific weights differ from the original Charlson index weights?

Objective 2: To examine the impact of Elixhauser on 1 year predicted mortality in ICU patients.

Primary Research Questions:

- How well does the Elixhauser predict 1 year mortality in ICU patients?

Secondary Research Questions:

- How much is gained for mortality prediction when you add Elixhauser comorbidities with age and sex?
- What is the impact of a specific comorbidity on mortality?
- Does the predictive ability of Elixhauser or Charlson change according to ICU patient subgroups?

Objective 3: To determine whether Charlson and Elixhauser contain all comorbidities those are associated with 1 year mortality in ICU patients.

Primary Research Questions:

- Are there comorbidities missing in Charlson and Elixhauser that can improve risk adjustment further in the ICU population?

CHAPTER 3: METHODS

3.1. Study sample

The Department of Critical Care Medicine at the University of Calgary and the Calgary Health Region (CHR) administer care to critically ill adult patients in a cardiovascular care ICU and three multidisciplinary ICUs. During the study period, the multidisciplinary ICUs consisted of 22 beds from a regional trauma and neurosurgical referral centre, 12 beds from a vascular surgery referral center and 10 beds from a multidisciplinary ICU. The study population consisted of all adult ($\geq$18 years) patient with an index admission into any multidisciplinary ICU in the CHR between April 1st 2000 to March 31st 2004. We excluded cardiovascular surgery patients because these patients were different from other patients admitted into ICU with respect to their physiologic derangement and mortality rate. In the development of APACHE II, cardiovascular patients had high initial APS scores and very low mortality rate, thus they were excluded from the analysis (3, 20).

Table 3: Exclusion Criteria

• Age <18 years old
• Not a Calgary Health Region resident during the study period
• CV ICU patients
• Repeat admissions (only patients index admission was considered)

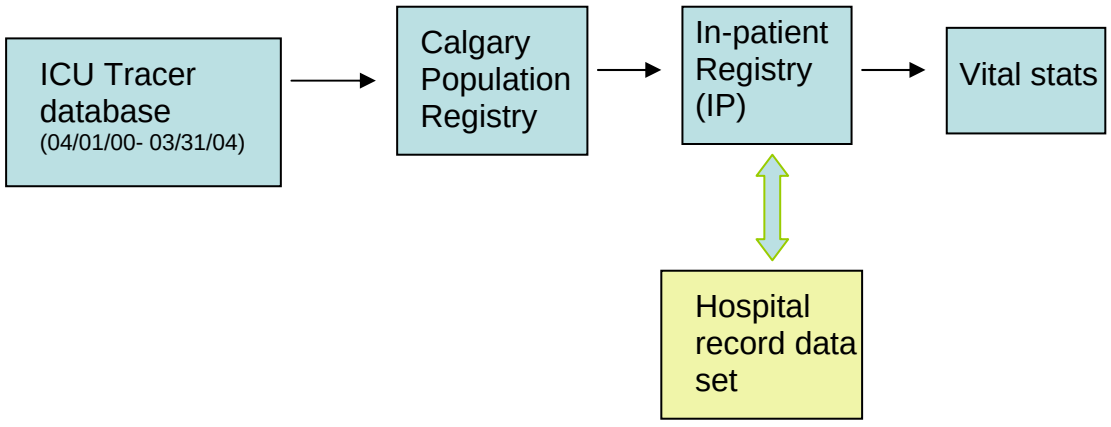
3.2. Data collection and study protocol

This study used a population-based historical cohort design that linked clinical outcomes data to large administrative databases. The ICU Tracer database collects patient specific demographic, clinical, APACHE II scores on the first day of ICU admission for all patients admitted to a Calgary Health Region adult ICU. The Tracer database was linked with the CHR Population Registry database to identify by postal code those patients who were residents in the CHR. Patients remaining in the cohort were

then linked to the in-patient (IP) registry through a deterministic record linkage approach by their unique hospital identification number, followed by last name, sex, and date of birth . Next, the IP registry was linked to a hospital record dataset that contained up to 16 diagnosis codes and diagnosis types as well as 10 procedure codes, to derive the Charlson index and the Elixhauser comorbidities for each patient. We used the ICD-10-CM codes for fiscal years 2002-2004 and ICD-9-CM codes for fiscal year 2000-2002 to identify diagnoses and procedures. Post-admit comorbidities, identified by a diagnosis type indicator, were excluded because these complications occurred after hospital admission. Finally, the dataset was linked with Vital Statistics data (2000-2005) to collect information on 1 year mortality. Since some Vital Statistics record could not be linked with PHN, a deterministic linkage approach using surname, sex, and date of birth was applied to ensure maximum linkage rates. This method of deterministic linkage with Vital Statistics was chosen because of it had previously been validated using these databases and had been found to have a high rate of accuracy (93.1%), with a trend to increased linkage by fiscal year. (81). Through this process, we were able to identify and link 97% (5159/5311) of individual patient records within our cohort. Therefore, the final study sample is 5159 subjects.

Appendix 4: Coding documentation for data cleaning

Figure 3: Database linkage



3.3. Study variables

APACHE II

APACHE II score (Appendix 1) is composed of three components: physiological variables, age, and chronic health points (CHP), and has a potential range of 0-73 points with a higher score corresponding with an increased risk of death. Age is stratified into 5 levels ranging from 0-6 points. CHP is stratified into 3 levels (0, 2, 5 points) based on elective or emergent surgery or non-operative case and the presence of either clinically significant immuno-suppression or chronic end-organ disease. Within the physiological component, the model consists of the most abnormal value of 12 physiological variables recorded within the first 24 hours after admission into the ICU. The APACHE II score was collected in accordance with the original published methods with the following modifications: (1) Glasgow coma scale was determined by the ICU attending physician and was the best estimate of the patient's underlying response due to the acute process excluding the potential confounding effect of sedation, (2) acute renal failure was considered present if the serum creatinine was acutely increased above 132 mmol/L, (3) chronic health status was determined by the most responsible ICU attending physician at the time of ICU admission, and (4) laboratory and physiologic variables were downloaded from bedside electronic devices, after being verified by bedside nursing staff for accuracy, to a bedside electronic charting system that automatically calculated the Apache II score. Previous studies had verified the accuracy of the data download and the calculation of the score (82). The APACHE II score and its components for this study were from this data collected in the ICU Tracer database.

Charlson comorbidities

In the present study, we used the ICD-10 and ICD-9 coding algorithms developed by Quan *et al*. (50), to derive a comorbidity score for the Charlson index for each patient.

Table 4: Charlson index comorbidities

Weight	Clinical condition
1	Myocardial infarct
	Congestive cardiac insufficiency
	Peripheral vascular disease
	Dementia
	Cerebrovascular disease
	Chronic pulmonary disease
	Conjunctive tissue disease
	Slight diabetes, without complications
	Ulcers
	Chronic diseases of the liver or cirrhosis
2	Hemiplegia
	Moderate or severe kidney disease
	Diabetes with complications
	Tumors
	Leukemia
	Lymphoma
3	Moderate or severe liver disease
6	Malignant tumor, metastasis
	Aids

Elixhauser comorbidities

We used the ICD-10 and 9 coding algorithms developed by Quan *et al*. (50), to derive the Elixhauser comorbidities for each patient.

Table 5: Elixhauser comorbidities

Condition	
Fluid and electrolyte disease	Pulmonary circulation disorders
Hypertension	Hypothyroidism
Cardiac arrhythmias	Valvular disease
Chronic pulmonary disease	Metastatic cancer
Congestive heart failure	Diabetes, complicated
Other neurological disorders	Paralysis
Alcohol abuse	Rheumatoid arthritis
Coagulopathy	Drug abuse
Diabetes ,uncomplicated	Obesity
Peripheral vascular disorders	Weight loss
Solid tumor without metastasis	Deficiency anemias
Renal failure	Psychoses
Depression	Blood loss anemias
Liver disease	Peptic Ulcer Disease
	AIDS

3.4. Outcome measures

The primary outcome in this study was mortality at 365 days. Time to mortality was calculated from the date of death found in Vital Statistics and subtracted from the ICU admission date to determine the number of days. Secondary outcomes, 90 days and 180 days, were explored initially but were abandoned as there was very little difference in discrimination across outcomes.

3.5.0 Descriptive baseline statistics in the ICU study sample

Baseline characteristics were calculated on common demographic and clinical variables such as age, APACHE II, Charlson index, length of stay. Primary admitting ICU diagnosis for each patient was classified under five categories (cardiovascular, respiratory, gastrointestinal, neurological, trauma, other & missing). In general, normally distributed variables were reported as means with standard deviations (SD) and non-

normally distributed variables as medians with inter-quartile ranges (IQR). Prior to analysis, all variables were assessed for underlying distribution using histograms, box plots and summary statistics.

3.5.1. Describing comorbidities in the ICU study sample

Several analyses were conducted to demonstrate how the coding algorithms for Charlson index and Elixhauser differ from each other. First, the prevalence of Charlson and Elixhauser comorbidities were calculated from the study sample. As defined by Rothman, prevalence is a measure of disease burden in the population. Prevalence is an appropriate measure for comorbidities since many comorbidities, such as diabetes, hypertension, have insidious onset; therefore it would be difficult to define the onset. The prevalence was ranked from highest to lowest for both Charlson and Elixhauser, with a rank one representing the comorbidity with the highest prevalence in the ICU sample. This ranking can be a crude way to assess whether comorbidity algorithms differ from each other. To provide a crude assessment of the relationship between comorbidity and 365 day mortality, several graphs were developed to also show the number of patients with comorbidity and the number of patients that died from a specific comorbidity. Finally, the Charlson and Elixhauser methods were compared in their ability to assess disparities in the determination of comorbidities among ICU patients.

3.5.2. A comparison between CHP, Charlson index , and Elixhauser by outcomes and descriptive variables

In this section, a stratification analysis was used to assess how well outcomes and descriptive variables differed between patients classified with no comorbidities and patients with any comorbidity, using the Charlson and Elixhauser method. When a comorbidity measure differentiated well in terms of the descriptive variables, such as age or sex, this provided a better indication that the comorbidity measure may be related to the descriptive variable as a confounder.

For all stratified analyses, a decision was made post-hoc to collapse the CHP component of APACHE II, into 2 groups: zero and two chronic health points or five chronic health points. The decision to collapse was based on the distribution of subjects

(only a small proportion of subjects were categorized as having 2 chronic health points), and a similar mortality rate between 0 and 2 CHP points, compared to 5 CHP points.

3.5.3. A comparison between survival curves by Charlson index score and count of Elixhauser comorbidities

Kaplan-Meier survival curves were graphed by comorbidity score or count to show the proportion of patients surviving up to four years of data collection. These graphs display differences between survival function of ICU patients with varying comorbidity scores by examining the slope of the curve and distance between the curves. To graph survival curves by Charlson index score, the score was calculated as the original Charlson *et al*, method and then combined into fewer categories based on the methods described by D'Hoore *et al*, (10, 52). To graph survival curves by Elixhauser comorbidities, a count of the comorbidities was calculated and patients with 7 or more comorbidities were combined into a single group.

3.6. Statistical techniques for model assessment

To investigate how well a risk adjustment method may perform, the C statistic (C stat) is the most commonly used measure of performance of models predicting dichotomous outcomes (83). The C stat represents the area under the curve for the probability that the mortality status of a randomly selected survivor and non survivor is identified correctly by a given prediction model (84). The maximum value of 1.0 for the C stat indicates that the model has 100% ability to discriminate mortality status. In contrast, when the model cannot discriminate better than chance, the C stat will be 0.5. In the literature, C stat of 0.8 to 0.9 are regarded as very good, C stat of 0.7 to 0.8 considered adequate and C stat below 0.7 are regarded as poor (85). It should be noted that the C stat has limited sensitivity to detect additional improvements in prediction at a threshold level and is insensitive to the prevalence of the condition.

As part of the model assessment process, the models were calibrated to assess how well the model measures actual and expected values of mortality across a sample subgroups. Most ICU severity illness scores had some deterioration in calibration over time, especially in subgroups, therefore, it would be important to assess whether

comorbidity scores are susceptible to the same issues(7, 8, 31). Calibration is determined by the Hosmer-Lemeshow Goodness of Fit test, which divides the population into subgroups by deciles of risk and a total Chi Square (χ^2) value is calculated. The null hypothesis is that the model does not fit with p values>0.05, whereas when the model fits the data the p <0.05. Low χ^2 values are indicative of good model performance. In contrast, x^2 greater than 15.51 are regarded as poor fit (25).However, it should be cautioned that the Goodness of Fit test is sensitive to large sample sizes which can produce erroneous significant p values. Therefore, we relied on a graphical means to assess this component.

To test whether the addition of predictor variables into the models changed significantly, likelihood ratio tests (LR) were used. For a given model, the likelihood is calculated as the joint probability of the observed outcome expressed as a function of the chosen regression model. The model coefficients were unknown quantities and are estimated by maximizing this probability. The LR test was computed as twice the difference between the log-likelihood from the two models, and is assessed by examining the X^2 distribution for significance testing (25).

Since a model performance can be severely degraded by the inclusion of many predictors, we calculated the Akaike's information criterion (AIC). This is a method for penalizing the log likelihood to gain an unbiased assessment of the model's performance. To calculate the AIC, we added twice the number of parameters to the -2 log likelihood and examine the relative change of AIC between similar models (25).

Finally, it is important to compare how a model performs in the data on which it was developed on and how well it performs when applied to new data. When fitting complicated models with many covariates, a model fit to one data set is unlikely to predict outcomes equally well in a new data set (25). When there are few covariates in model or when the data set is large and representative of future data, then over fitting is not a major concern. However, since the Charlson and Elixhauser were developed on non-ICU study samples and there are more than 17 covariates in a model, over fitting can be a potential problem. Charlson index was developed based on 559 medical patients and Elixhauser was developed based on a sample size of 1,779,167 adult inpatients. For

predicting dichotomous outcomes, the literature advises that there should be no more than one predictor per 18 events, which may not be sufficiently conservative (25).

To evaluate model fit, Efron's enhanced bootstrapping method was used to estimate the amount of shrinkage in the models (83). Shrinkage was calculated to obtain unbiased estimates of future model performance in similar populations. By shrinking the regression coefficients, over-fitting the model maybe corrected, and transportability of the models to similar populations improved. Shrinkage values close to one indicate very little shrinkage and are regarded as excellent, while shrinkage values much less than one can represent over fitting of the data. There were one hundred replications for bootstrapping coded to calculate mean shrinkage.

3.7.0. Primary Research Question:

How well does Charlson index predict 1 year mortality in ICU patients?

For the analysis of the primary outcome, variations of the Charlson score were explored in univariate logistic regression models with the dependent variable being the natural logarithm of 365 day mortality. First, the Charlson index was explored using the original weighting scheme and calculation methods (Model CI1). For this approach, after taking the exponent of the regression coefficient, these odds ratios were added rather than multiplied for a patient with multiple comorbidities, as described by the methods used by Charlson. Next, the Charlson index was collapsed into scores six and higher, since the relationship was non linear for high scores and the sample size was low for these high scores (Model CI2). D'Hoore also determined that the Charlson index had a non-linear relationship with the log odds of death and the Charlson index was based on highly skewed scores. D'hoore recoded the index into 5 categories (0, 1-2,3-4,5-6,>6) and found that this improved model performance (52). A model using the ICU specific regression coefficients was explored to calculate the Charlson Index (CI3) by multiplying regression coefficient if a patient had more than a single comorbidity. Lastly, a model where Charlson comorbidities were entered as dummy variables was explored (Model C1) to determine whether not fixing comorbidity weights would produce the same predictive accuracy as the previous models.

Model CI1 and CI2: Calculation of Charlson index using original weights and methods described by Charlson

Charlson Index= w_1*(c.cob$_1$)+ w_2*(c.cob$_2$)+…… + w_n*(c.cob$_n$)

w_n corresponds to the original weights derived by Charlson *et al*.

Model CI3: Calculation of Charlson index using ICU specific regression coefficients

Charlson Index= γ_1*(c.cob$_1$) + γ_2*(c.cob$_2$) + …… + γ_n*(c.cob$_n$)

γ_1 ... γ_n are the empirically derived regression coefficients

Model C1:Charlson comorbidity

The following logistic regression equation was used for the Charlson comorbidities models:

Log (p/1-p)= β0+ β1*(c.cob$_1$)+ β2*(c.cob$_2$)+… βn*(c.cob$_n$)

*p is the probability of death within 365 days for a patient

*c.cob are the comorbidities defined in the Charlson list entered as dummy variables

Model CI1-CI3: Comparison between Charlson index variations in univariate models to predict 365 days mortality

The following logistic regression equation was used for the Charlson index models:

Log (p/1-p)= β0+ β1 *(CI)

*p is the probability of death within 365 days for a patient

* CI is the Charlson index calculated for Models CI1 to CI3

3.7.1. Secondary Research Questions:

How much is gained for mortality prediction when Charlson comorbidities are added to age and sex?

Multivariate analysis was used to compare the performance of adding age and sex to the Charlson index to predict 365 day mortality. Age and sex were the only clinical variables chosen for this study because these variables are readily available in administrative data. First, a baseline model with age and sex was assessed on its own and different comorbidity measures were added to determine the change in discrimination.

Models M1- M5: Comparison between Charlson index variations and chronic health points added to a baseline model

The following logistic regression equation was used to determine whether discrimination could improve if variations of the Charlson index were added to a baseline model

M1: Log (p/1-p)= β0+ β1 *(age) + β1*(sex) baseline model

M2: Log (p/1-p)= β0+ β1*(age)+ β2*(sex) + β3 *(CHP)

M3: Log (p/1-p)= β0+ β1*(age) + β2*(sex) + β3 *(CI1)

M4: Log (p/1-p)= β0+ β1*(age) + β2*(sex) + β3 *(CI3)

M5: Log (p/1-p)= β0+ β1*(age) + β2*(sex)+ β3*(c.cob$_1$)+ β4*(c.cob$_2$)+… βn*(c.cob$_n$)

*p is the probability of death within 365 days for a patient

* CHP is the chronic health points of APACHE II

* CI1 is the Charlson index calculated using the original Charlson's method

* CI3 is the Charlson index calculated using regression coefficients

* c.cob is the Charlson comorbidities entered as dummy variables

The separate components of APACHE II were also explored to compare how well comorbidity scores perform to the severity illness score used in the ICU (Model A1, A2, and A3).

Model A1, A2, and A3: Comparison between APACHE II components in univariate models to predict 365 day mortality

The following logistic regression equation was used to model APACHE II components:

A1 : Log (p/1-p)= β0+ β1 *(APACHE II)

A2: Log (p/1-p)= β0+ β1 *(APS)

A3: Log (p/1-p)= β0+ β1* (CHP)

*p is the probability of death within 365 days for a patient

3.7.2. How does the effect of comorbidity on mortality depend on age?

Next, to determine whether adding an interaction term could improve discrimination for severity scores, a series of analysis were undertaken. In epidemiological analysis, an interaction term is known as an effect modifier and there is some evidence in the literature that suggests adding interaction terms between physiological measures can improve discrimination for severity of score measures(86).

First to examine the Charlson index and Elixhauser distribution across age groups, the mean and median number of comorbidities was plotted for each method. Age groups were divided as 18-24, 25-34, 34-44, 45-54, 55-64, 65-74, 75-84, 85 and over. Age groups were combined to create floating age groups in order to assess differences in discrimination between the age groups, while maintaining an adequate sample size. The C stat was calculated for the Charlson index, using the original weighting scheme (Model CI1), by age groups. Lastly, an interaction term of age and Charlson index was created and entered into a baseline model of age and sex o determine whether any improvement in discrimination could be made.

Model M3 and M6: Comparison between Charlson index model with an interaction term for age and comorbidity

The following logistic regression equation was used to address this question:

M3: Log (p/1-p)= β0+ β1(age) + β1(sex) + β3 (CI1) baseline model
M7: Log (p/1-p)= β0+ β1(age) + β1(sex) + β3 (CI1) + β4 (age*CI1)

*p is the probability of death within 365 days for a patient

* CI1 is the Charlson index calculated using the original Charlson's method

3.7.3. How do the new ICU specific weights differ from the original Charlson index weights?

In order to derive empirical study weights, a similar approach to Charlson *et al*, method was undertaken. Regression coefficients were transformed into odds ratios by taking the exponent and these resulting odds ratios were equivalent to the weights. Comorbidities with odds ratio $\geq$1.2<1.5 were assigned a weight of 1; conditions with a ratio of $\geq$1.5<2.5 a weight of 2; conditions with $\geq$2.5<3.5 a weight of 3; conditions with $\geq$3.5<4.5 a weight of 4; and those conditions with weights of 6 or more were assigned a weight of 6.

3.8.0. Primary Research Question:

How well does Elixhauser predict 1 year mortality in ICU patients?

As described by Elixhauser, the comorbidities were entered into a logistic regression model as dummy variables. When the exponent of the coefficient was transformed, this represented the odds ratio, where the odds that a patient with a particular comorbidity would die in 365 days, holding constant all other factors measured. For example, if a patient had odds ratio of 2.5 for congestive heart failure, then this would be interpreted as the odds of death is 2.5 times higher for patients with congestive heart failure, compared to patients who had no congestive heart failure recorded but were similar in all other factors.

Model E1: Elixhauser comorbidities

The following logistic regression equation was used for the Elixhauser model:

Log (p/1-p)= β0+ β1*(e.cob$_1$)+ β2*(e.cob$_2$)+… βn*(e.cob$_n$)

*p is the probability of death within 365 days for a patient

*e.cob are the comorbidities defined in the Elixhauser list

3.8.1. How much is gained in terms of mortality prediction when Elixhauser comorbidities are added to age and sex?

Similar to the Charlson analysis, likelihood ratio tests, C stat, AIC, and shrinkage was calculated to evaluate model performance compared to a baseline model. This question was explored to determine how much discrimination could be gained when Elixhauser comorbidities was added to a model that contained age and sex.

Model M1 & M7 Comparison between an Elixhauser model added to a baseline model to predict 365 day mortality

The following logistic regression equations was used to determine if discrimination could improve when Elixhauser was added to a baseline model

M1: Log (p/1-p)= β0+ β1* (age) + β1*(sex) baseline model

M7: Log (p/1-p)= β0+ β1* (age) + β1*(sex) + δ$_1$*(e.cob$_1$)+….+ δ$_n$*(e.cob$_n$)

*p is the probability of death within 365 days for a patient

* e. cob is the 30 Elixhauser comorbidities entered as dummy variables

3.8.2. What is the impact of a specific comorbidity on mortality?

Forest plots of univariate and multivariate models for Charlson and Elixhauser comorbidities were explored in this analysis. This graphical assessment allows a quick assessment of odds ratios and 95% CI for a given comorbidity. The length of the 95% CI can provide an indication of the variability of the odds ratio estimate. In the scenario where the 95% CI was wide and crossed an odds ratio of one, this suggested that the odds of dying in 365 days were higher or lower for a person with a given comorbidity than a

person without that comorbidity. In other words, the odds ratio was not statistically significant ($p>0.05$). Odds ratios and 95% CI less than one suggest that the odds of death are less in a person with a comorbidity compared to without the comorbidity, whereas, odds ratios and 95% CI greater than one suggest the opposite effect.

Model E1 and C1

The following logistic regression equation was used to address this question:

E1: Log (p/1-p)= $\beta 0 + \beta 1 * (e.cob_1) + \beta 2 * (e.cob_2) + \ldots \beta n * (e.cob_n)$

C1: Log (p/1-p)= $\beta 0 + \beta 1 * (c.cob_1) + \beta 2 * (c.cob_2) + \ldots \beta n * (c.cob_n)$

*p is the probability of death within 365 days for a patient

*e.cob are the comorbidities defined in the Elixhauser list

* c.cob are the comorbidities defined in the Charlson list

3.8.3. Does the predictive ability of Charlson index and Elixhauser change according to ICU patient subgroups?

This was explored by examining comorbidity profiles and physiological profiles between risk groups: trauma vs. non trauma patients, surgical vs. non surgical patients. A model with only ICU survivors was explored because it was hypothesized that comorbidities may play an important role in survival for long term mortality outcomes rather than short term mortality, therefore, comorbidity models would perform better when only ICU survivors are examined. This analysis could provide health service researchers an indication of which comorbidity scores would be most appropriate for risk adjustment within a group of ICU patients. Logistic regression models were constructed to examine the performance of the C stat between each ICU group using Elixhauser and Charlson index. The original Charlson index weighting scheme was used for this analysis.

3.9. Primary Research Questions:

Are there comorbidities missing in Charlson and Elixhauser that can improve risk adjustment further in the ICU population?

3.9.1. Background of Delphi Method

A modified Delphi technique was used to develop a list of clinical comorbidities that may be relevant for predicting one year mortality in ICU patients. Because Elixhauser and Charlson index were developed on non ICU patients, this technique would assess whether there are additional comorbidities that could be added to the list, in order to improve comorbidity risk adjustment in the ICU population. The Charlson index was developed on less than 600 patients and may not have the adequate sample size to detect statically significant comorbidities.

The Delphi technique is a consensus method that provides a way to synthesize information and combine the insights of appropriate experts to enable informed decision making or health service development. It was first developed in the 1950s by the Rand Corporation in California as an attempt to eliminate interpersonal interactions during decision making for meetings(87). The aim of the Delphi technique is to determine the extent to which experts agree about a given issue. Agreement is evaluated in three different methods. First, agreement is the extent to which each respondent agrees with the issue under consideration, usually assessed by a numerical or categorical method. Secondly, agreement can be the extent to which respondents agree with other. Lastly, in this research question context, agreement can be determined by how well the responses agree with the conventional list of comorbidities.

There has been some controversy over the validity of the Delphi method. For example, Harold Sackman argued that the Delphi method fails to meet the standards normally set for scientific methods. For example, although the anonymity feature of the Delphi promotes an honest expression of views and protects the panel from penalty or mockery of their opinion, it can lead to a lack of accountability for the view expressed(88). Further, he argues that the method forces consensus and is weakened by not allowing participants to discuss issues.

However, there are distinct advantages in using a Delphi technique to obtain expert opinion. For example, using the Delphi technique can overcome problems, such as a dominant individual and their view being overly represented. Also, a feedback method in the Delphi technique allows individuals to reevaluate their responses in a non confrontational environment, while considering the responses from the rest of the group. In a practical sense, the Delphi method allows a large group of experts to be contacted cheaply, such as by email or mail, with few geographical limitations on the sample.

The Delphi technique has been used widely in health service research, fields of technology assessment, education and training, and in developing nursing and clinical practice (89-92). In health service research, researchers may face the problem of trying to make decisions in situations where there is insufficient information. Therefore, the purpose would be to identify whether additional comorbidities could be added to Charlson and Elixhauser that could improve risk adjustment from a group of expert ICU physicians.

The Delphi method is characterized by four features: anonymity for all respondents; iteration; controlled feedback; and statistically interpretable group response(93).

Table 6: Features of Delphi method

Anonymity	To avoid dominance; achieved by use of a questionnaire in Delphi and private ranking in nominal group
Iteration	Processes occur in "rounds", allowing individuals to change their opinions
Controlled feedback	Showing the distribution of the group's response (indicating to each individual their own previous response in Delphi)
Statistical group response	Expressing judgment using response summary measures of the full group response, giving more information than just a consensus statement

Adapted from Pill and Rowe (94, 95)

3.9.2. Study sample and recruitment

A convenience sample of nineteen practicing adult intensivist across Canada were identified and asked to participate in the study.

3.9.3. Survey administration and steps

Physicians were given an introductory letter with the purpose of the project, time requirements, standard definitions and steps. Efforts were made to contact physicians at research rounds and through email. The questionnaires were reviewed for face validity by 2 independent physicians with content and method expertise prior to administration.

Appendix 5: Delphi Cover Letter

Appendix 6: Delphi Procedure

The modified Delphi Technique was implemented in the following three steps:

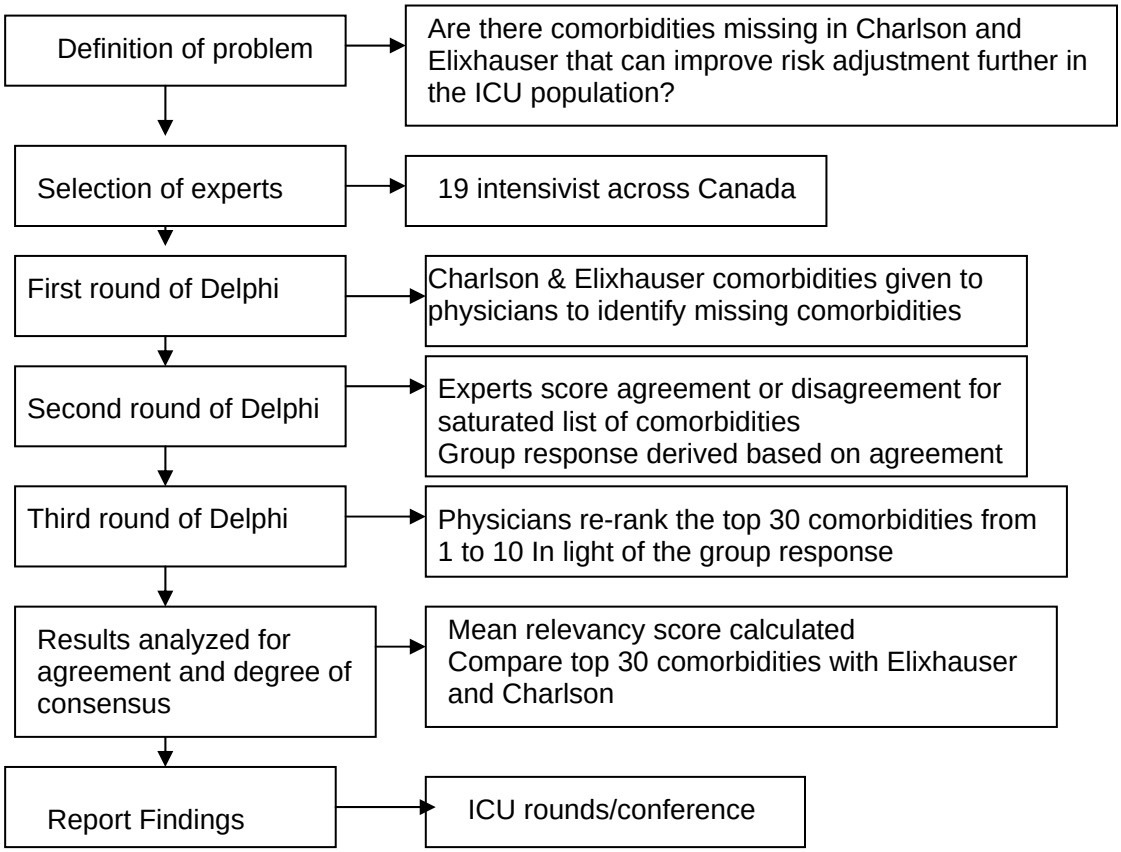
Step 1: A list of Charlson and Elixhauser comorbidities were given to the physicians for review. Experts were asked to provide a list of additional comorbidities that would predict mortality in an ICU population, based on their clinical experience and knowledge.

This list was combined with the all the comorbidities from Charlson and Elixhauser to form a saturated list.

Step 2: These experts provided a response on each of the comorbidities from the saturated list as either yes or no, according to its importance for predicting 90, 180, and 365 days mortality. Based on these results, the percentage agreement for each comorbidities was derived, also known as the group response.

Step 3: The group response was sent back to the expert panel for a second ranking of the top 30 comorbidities . Physicians ranked the comorbidities from a scale of 1 to 10 with high numbers representing the most relevant comorbidities for predicting 90 and 365 day mortality. The final list of comorbidities were formulated based on the mean relevancy score obtained from the questionnaire in Step 4. This list was compared with the Charlson index and Elixhauser list of comorbidities.

Figure 4: Delphi Technique in the ICU research study



3.9.4. Loss to follow up

Since the Delphi technique requires two follow up questionnaires, physicians were provided with more than a month to respond to each questionnaire. One week after administering the questionnaire, reminder emails were sent and physicians were approached at research rounds to facilitate the responses.

3.9.5. Outcome measure

Although physicians were asked to provide response for 90 and 180 day mortality, this was eliminated in subsequent steps because it was cumbersome, and could inhibit the number of responses received. Furthermore, the responses were similar across

the different endpoints (Delphi I & II) and previous analysis showed consistent results across endpoints. This provided justification to focus on the primary outcome.

3.9.6. Statistical analysis

Additional comorbidities were identified by ICU physicians and the percentage agreement was derived (i.e. group response). Mean Relevancy Rank was calculated from the second ranking where physicians ranked comorbidities from 1 to 10. The top 30 comorbidities were compared with the original Charlson and Elixhauser for differences. Odds ratios obtained from multivariate models for Elixhauser (Model E1) was compared with the rankings in a correlation analysis. In a subsequent correlation analysis, comorbidities that were present in <5% of the ICU population were excluded because ICU physicians may not have an accurate assessment of the comorbidity influence on mortality. Comorbidities identified by the physicians that were not on Elixhauser or Charlson's comorbidities list were tested for discrimination by the following logistic regression equations.

Model D1: Delphi comorbidities

The following univariate logistic regression equation was used for the comorbidities identified by the Delphi process:

$$\text{Log } (p/1-p) = \beta 0 + \beta 1 * (\text{d.cob})$$

*p is the probability of death within 365 days for a patient

*d.cob are the comorbidities identified by the ICU physicians in the Delphi process

Model D2: Delphi comorbidities

The following multivariate logistic regression equation was tested whether the Delphi comorbidities could add any discrimination over the Elixhauser model:

$$\text{Log } (p/1-p) = \beta 0 + \beta 1 * (\text{e.cob}_1) + \beta 2 * (\text{e.cob}_2) + \ldots \beta n * (\text{e.cob}_n) + \beta n * (\text{d.cob}_n)$$

*p is the probability of death within 365 days for a patient

*e.cob are the comorbidities defined in the Elixhauser list

*d.cob are the comorbidities identified by the ICU physicians in the Delphi process

3.9.6. Statistical Programs

Data cleaning was conducted using SAS version 9.1 (Cary, NC). Statistical analysis was primarily done in STATA 9.2 (College Station, Texas). GraphPadPrism 5 application was used to develop statistical graphs.

3.9.7. Ethics Approval

Ethics approval for this study was obtained from the conjoint health research ethics board at the University of Calgary and the Calgary Health Region.

Appendix 7: Ethics Approval Letter

CHAPTER 4: RESULTS

4.1.1. Study sample characteristics

A total of 5159 patients met the inclusion criteria for the study. The data cleaning and screening process is summarized in Figure 5. Less than three percent of patients (n=152) could not be linked to Vital Stats data or CHR population registry because of invalid PHNs, which suggests these residents were not part of the CHR.

Figure 5: Data cleaning and screening process

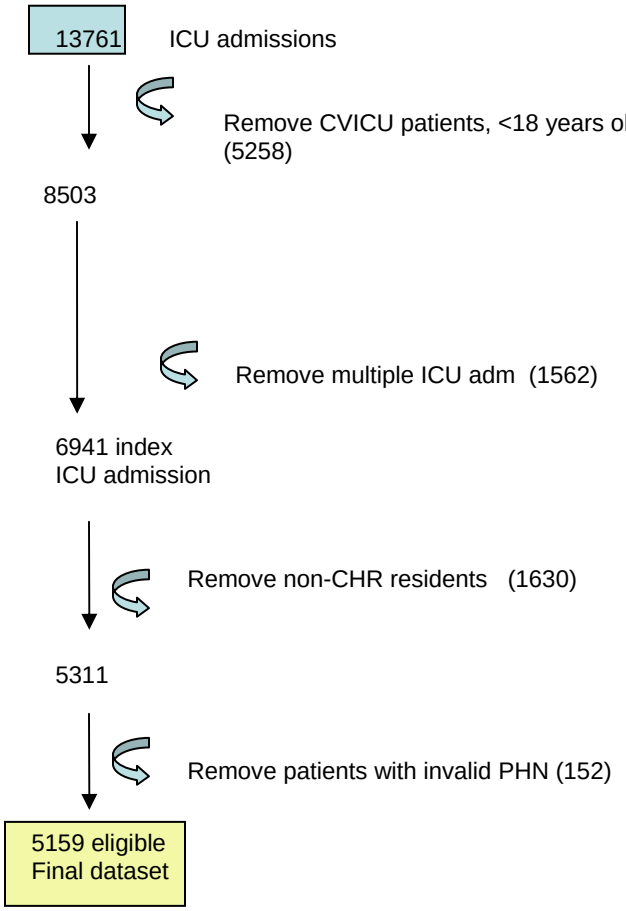


Table 7 provides an overview of the characteristics of the study sample that met all inclusion and exclusion criteria. The average age of patients in the study sample was approximately 59 years old and 56% of the sample consisted of males. Both the Charlson index and length of stay in the ICU were positively skewed; for this reason, medians with inter-quartile ranges were presented. The burden of comorbidities was low based on the CHP (66.6% had 0 CHP) and Charlson index (50% of the ICU population had index between 0 and 3).

Table 7 also provides a list of the primary admitting diagnosis in the ICU. The majority of ICU admissions were related to a primary admitting diagnosis (such as septic shock) in the cardiovascular system (28.2%). Nineteen percent of patients died in the ICU (n=986). Close to 24% (1184/4994) of patients had a hospital stay greater than 30 days, of which 18% (214/1184) died in the hospital. Approximately 5% (261/4994) of patients had a hospital stay greater than 90 days, of which 12% (30/261) died in the hospital. There were missing data on hospital length of stay for 165 patients.

Table 7: Patient Characteristics

Age (y)		
Mean ± SD		58.97±18.7
Sex		
Male Frequency (%)		2874 (55.7)
Length of ICU stay (d)		
Median (IQR)†		2.5 (1.1,5.9)
Admitting APACHE II Score		
Mean ± SD*		19.3 ± 8.6
First TISS score‡		
Mean ± SD*		35.0± 14.4
In-hospital Mortality		
Frequency (%)		1432 (27.8)
ICU mortality		
Frequency (%)		986 (19.1)
Charlson Index		
Median (IQR; maximum)†		2 (0,3;15)
CHP (%)		
0		3370 (66.6)
2		174 (3.4)
5		1517 (30.0)
APS		
Mean ± SD		14.4±7.9
Surgery Type (%)		
Non-Operative		3429 (67.3)
Elective Surgery		542 (10.6)
Emergency		1121 (22.0)
Max SOFA ± SD		7.7 ±4.6
90 day mortality (%)		1623 (31.5)
180 day mortality (%)		1746 (33.8)
365 day mortality (%)		1899 (36.8)
Admitting Diagnosis Category	Cardiovascular (%)	1457 (28.2)
	Respiratory (%)	1613 (31.3)
	Gastrointestinal (%)	542 (10.5)
	Neurological (%)	601 (11.6)
	Trauma (%)	372 (7.2)
	*Other & Missing (%)	874 (16.9)

Total sample size= 5159, unless otherwise stated.
*SD = standard déviations
†IQR= inter-quartile range
‡TISS= Therapeutic Intervention Scoring System
SOFA= Sequential Organ Failure Assessment Score
CHP= Chronic Health Points (APACHE II)

Other includes poisoning, musculoskeletal, hematological, genitourinary, and dermatological.

4.1.2. A descriptive overview of the comorbidity profile in the ICU sample

In order to further explore the comorbidity profile of the study sample, the Charlson and Elixhauser comorbidities were tabulated and ranked in respect to prevalence in the study sample. Table 8 provides the percentages and ranks of ICU patients with Charlson comorbidities, and Table 9 Elixhauser comorbidities derived from ICD- 9 and 10 coding algorithms. The most frequent comorbidity identified by Charlson was chronic pulmonary disease (N=1212). Cardiovascular diseases such as congestive heart failure and myocardial infarction represented 37% of comorbidities.

Table 8: Percentage and ranks of patients with Charlson Comorbidities

Condition	Percent	Rank
Chronic pulmonary disease	23.5	1
Congestive heart failure	18.9	2
Myocardial infarct	18.3	3
Diabetes without complications	12.8	4
Cancer(lymphoma & leukemia)	12.4	5
Cerebrovascular disease	11.8	6
Peripheral vascular disease	10.3	7
Moderate or severe renal disease	9.7	8
Metastic solid tumor	5.7	9
Ulcer disease	5.3	10
Hemiplegia + paraplegia	4.7	11
Diabetes with complications	4.5	12
Mild Liver disease	4.4	13
Moderate or severe liver disease	3.7	14
Connective tissue disease	2.6	15
Dementia	2.5	16
AIDS	0.3	17

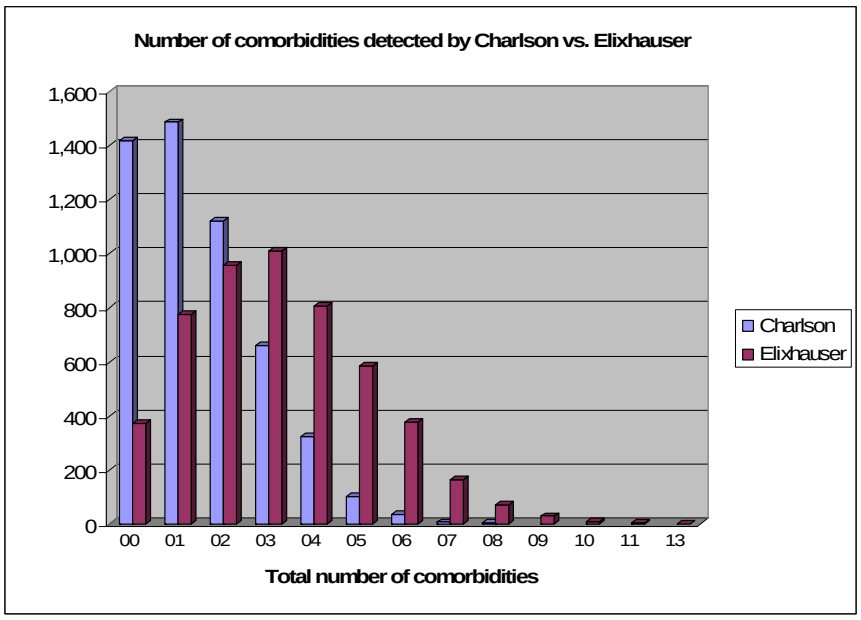
The most frequent comorbidity identified by Elixhauser was fluid and electrolyte imbalances (N=2116). Among comorbidities common to both methods, very similar proportions of patients affected were identified, for instance 18.9% of patients were identified with congestive heart failure (CHF) by the Charlson method, while 18.9% were identified using Elixhauser.

Table 9: Percentage and ranks of patients with Elixhauser comorbidities

Condition	Percent	Rank
Fluid and electrolyte disease	41.0	1
Hypertension	33.0	2
Cardiac arrhythmias	28.0	3
Chronic pulmonary disease	23.5	4
Congestive heart failure	18.9	5
Other neurological disorders	17.5	6
Alcohol abuse	15.5	7
Coagulopathy	13.2	8
Diabetes ,uncomplicated	12.6	9
Peripheral vascular disorders	10.3	10
Solid tumor without metastasis	9.9	11
Renal failure	9.6	12
Depression	8.9	13
Liver disease	8.3	14
Pulmonary circulation disorders	8.1	15
Hypothyroidism	6.7	16
Valvular disease	6.3	17
Metastatic cancer	5.7	18
Diabetes, complicated	5.1	19
Paralysis	4.7	20
Rheumatoid arthritis	4.4	21
Drug abuse	4.4	22
Obesity	3.9	23
Weight loss	3.3	24
Deficiency anemias	2.9	25
Psychoses	2.2	26
Blood loss anemias	1.8	27
Peptic Ulcer Disease	1.8	28
Lymphoma	1.7	29
AIDS	0.3	30

The Charlson and Elixhauser methods were further compared in their ability to identify multiple comorbidities among ICU patients. Compared to Elixhauser, Charlson identified more than three times the number of patients without any comorbidity. In addition, the distribution of Elixhauser comorbidities was fairly normally distributed compared to the Charlson index.

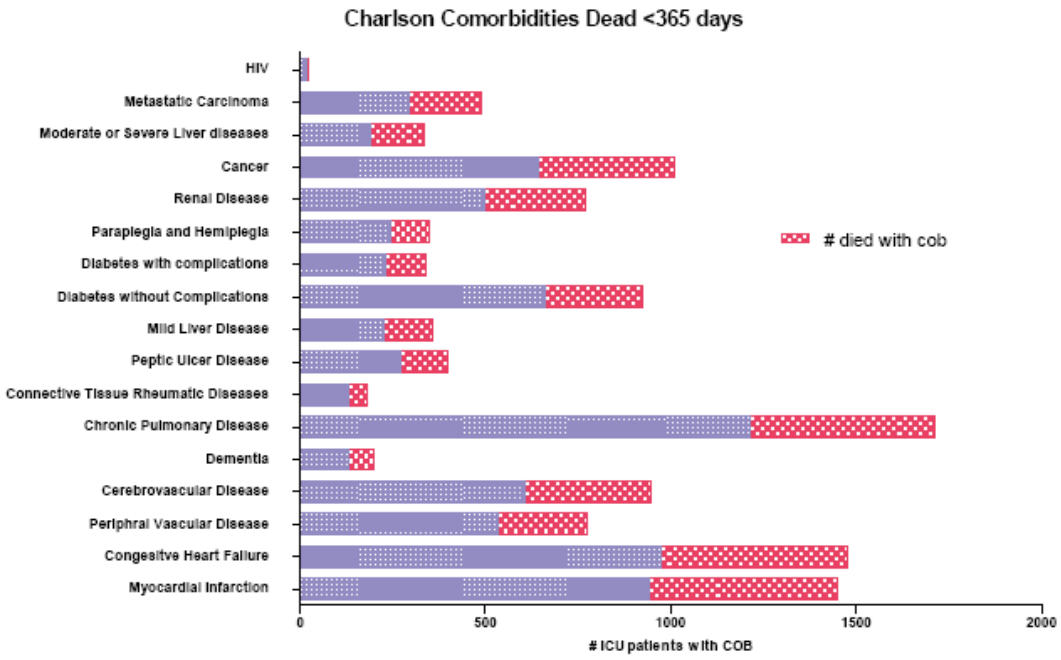
Figure 6: A comparison of the number of comorbidities detected by Charlson and Elixhauser



Note: Each patient could code for more than one comorbidity.

A crude overview of the comorbidities that had the highest proportion of death within 365 days can be observed in Figure 7 and Figure 8. Congestive heart failure, myocardial infarction, and chronic pulmonary disease are the top three comorbidities in the ICU population, with the most number of deaths within 365 days.

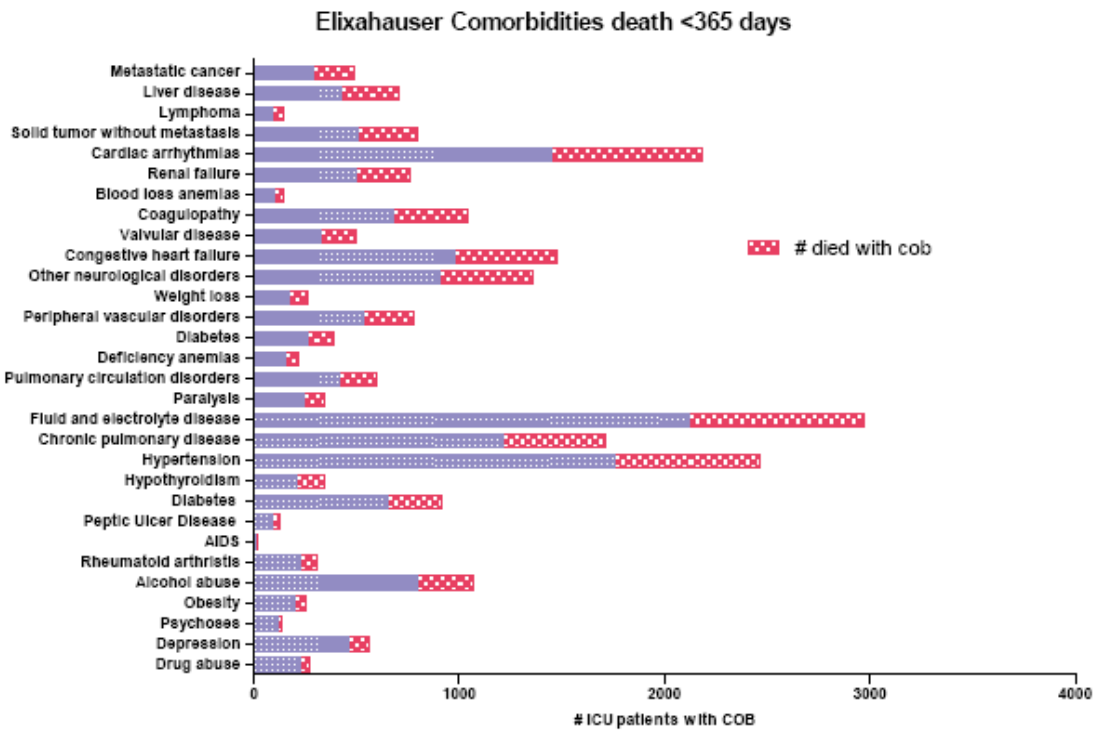
Figure 7: Proportion of ICU patients that died in 365 days by Charlson comorbidities



Note: Purple bar shows the number of ICU patients with a certain comorbidity that survived. Pink bars shows the number of ICU patients with a certain comorbidity that died.

A similar graphical representation of mortality was also presented for Elixhauser comorbidities. Figure 8 shows that cardiac arrhythmia, fluid and electrolyte imbalance, and hypertension are the top three comorbidities in ICU, with the most number of deaths within 365 days. Almost one third of patients who had other neurological disorders died, which was not captured by any of the Charlson comorbidities. Similar to the Charlson method, very few patients had HIV in the ICU.

Figure 8: Proportion of ICU patients that died within 365 days by Elixhauser comorbidities



Note: Purple bar shows the number of ICU patients with a certain comorbidity that survived. Pink bars shows the number of ICU patients with a certain comorbidity that died.

4.1.3. A comparison between CHP, Charlson index , and Elixhauser by outcomes and descriptive variables

Next, to assess how well clinical outcomes and descriptive variables differed between a patients classified without any comorbidities and some comorbidities, analysis between these groups were done for Elixhauser and Charlson index.

A different approach was undertaken for CHP where patients with zero and two CHP were collapsed into the same group because these patients had similar proportions of death (28%, 30% vs. 58%). The majority of patients had CHP of 0 and 2 (70%), while the mortality rate was roughly twice has high for those with 5 CHP.

Table 10: Comparison between CHP, Charlson index, and Elixhauser by clinical and descriptive variables

	CHP		Charlson index		Elixhauser	
	0,2	5	0	>=1	0	>=1
Patients (%)	70.0	29.9	27.5	72.5	7.2	92.8
Mean age ± SD	56.8 ± 19.7	64.1±14.9	45.8 + 19.1	64. 1+ 15.9	43.0 + 19.1	60.21 + 18.1
In-hospital death (%)	21.7	42.0	11.6	33.9	13.4	28.9
365 day mortality (%)	27.9	57.6	15.1	45.0	15.3	38.5
Length of stay ICU (Median , IQR)	2.4 (1.1,5.7)	2.9 (1.3 6.7)	1.9 (0.9, 4.5)	2.9 (1.2, 6.4)	1.6 (0.9, 3.1)	2.6 (1.1,6.0)
APACHE II +SD	17.7 + 8.0	23.1 + 8.8	15.9 ±8.0	20.5± 8.5	14.4+7.5	19.6+ 8.6
Male (%)	56.5	54.9	56.0	55.7	54.2	55.9
Max_SOFA (median, IQR)	7 (4,10)	8 (5,12)	6 (3,9)	8(5,11)	5 (2,8)	7(4,11)

Note: 365 day mortality represent the number of ICU patients that died at 365 days. Max_SOFA represents the maximum sequential organ failure assessment score during an ICU stay

The majority of patients (72.5%) had some Charlson index points and their average age was older. The mortality rate was three times as high for ICU patients with some Charlson index points, compared to patients without any points. Accordingly,

patients with Charlson index points also had higher APACHE II, Max SOFA scores and longer lengths of stay compared to those without points.

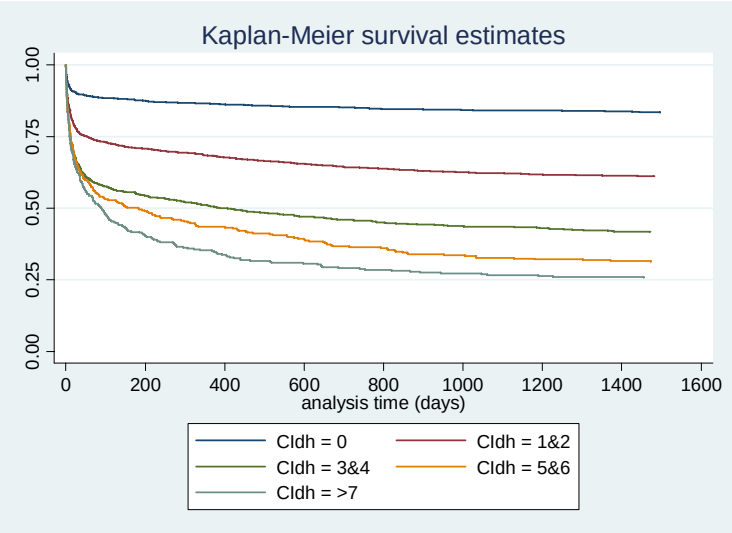
The majority of patients (92.8%) had some Elixhauser comorbidities, compared to 72.5% for Charlson comorbidities 33.4% for CHP. The mortality rate was more than twice as high for ICU patients with some Elixhauser comorbidities, compared to patients without any comorbidity. Patients with Elixhauser comorbidities had higher APACHE II, Max SOFA scores and longer lengths of stay compared to those without any comorbidity.

4.1.4. A comparison between survival curves by Charlson index score and count of Elixhauser comorbidities

Kaplan-Meier survival curves were graphed by comorbidity score or count to show the proportion of patients surviving up to four years of data collection. Figure 9 shows each survival curves for patients with Charlson index score from 0 to 7 and higher. This graph shows that survival curves begin to separate between patients with different Charlson index scores within the first 50 days after ICU stay. Less than 40% of patients with a Charlson index score of 7 or higher survived at 200 days after they were admitted into the ICU. In contrast, more than 80% of patients with a score of 0 survived at 200 days after ICU admission. The change in slopes was most apparent in patients with a score of 7 or higher and this trend continued until the follow up period of the study ended.

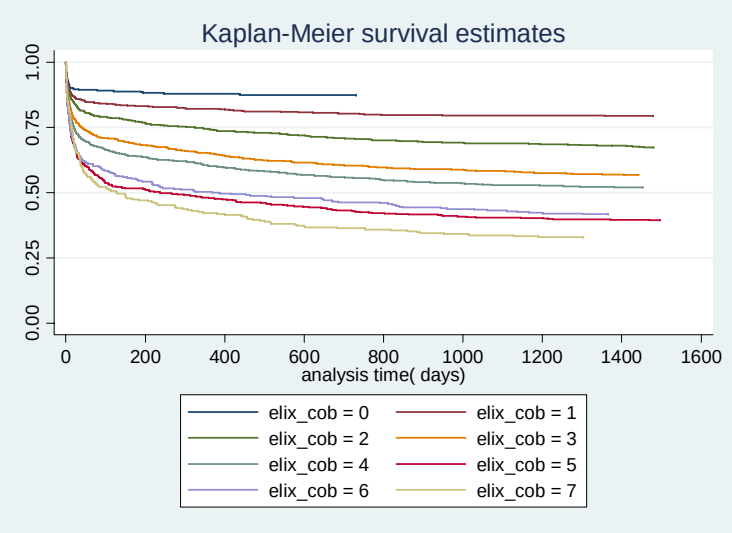
Figure 10 displays survival curves plotted by a count of Elixhauser comorbidities. A similar trend was shown here where patients with more than 7 Elixhauser comorbidities had the worst survival curve. For instance, close to 40% of patients with seven or more comorbidities were still surviving at 200 days after ICU admission, while almost 90% of patients with no comorbidities were surviving at this point in time. STATA graphs the survival curves up to maximum days of follow up for non-censored patients. For example, the survival curve for patients with zero Elixhauser comorbidities was followed up to 730 days. Overall, these survival curves show that patients with a high Charlson index score or high number of Elixhauser comorbidities had poor survival functions after being admitted in the ICU compared to patients with no comorbidities.

Figure 9: Kaplan-Meier survival curves by Charlson index score



Note: CIdh represents Charlson index recoded in the same categories as D'hoore(96).

Figure 10: Kaplan-Meier survival curves by Elixhauser count of comorbidities



Note: Elix_cob represents the total number of Elixhauser comorbidities.

4.2.0. Primary Research Question:

How well does Charlson index predict 1 year mortality in ICU patients?

To address the primary research question of the study, we undertook an investigation of the performance of the Charlson index for predicting 365 day mortality among ICU patients in the CHR.

The first step of this analysis was to explore Charlson index in univariate logistic regression models.

Predictive Performance

In these models, the Charlson index was explored in several variations to determine if the variations differed from the original weighting scheme: a recoded version where scores six and above were combined, a version using the regression coefficients and a version where comorbidities were entered as dummy variables. Scores above six were combined because the Charlson index had very few patients with scores above six (Model CI2).

Results show that the original Charlson Index and variations had adequate ability to discriminate between those who lived and died in the ICU (C=0.70 to 0.73), with minimal differences between the variations. When exact ICU specific regression coefficients were used, as suggested by Moon and Harrell, this achieved better discrimination than using the original weights.

Table 11: Comparison of discrimination between Charlson Index variations in univariate models to predict 365 day mortality

Model	C stat	OR	95% CI
CI1	0.70	1.34	1.31 , 1.38
CI2	0.70	1.46	1.41, 1.50
CI3	0.73		
C1	0.73		

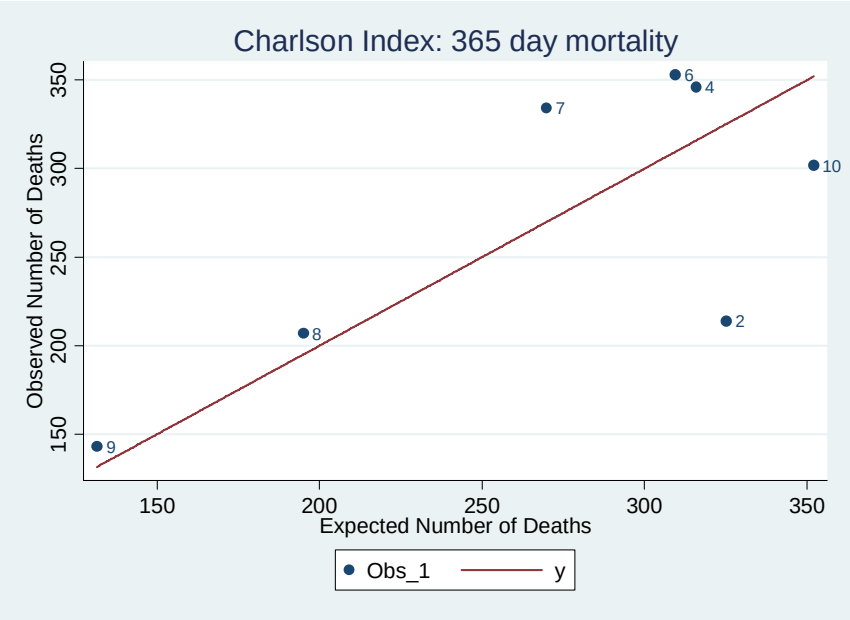
Note: CI1 refers to the Charlson index using the original weighting scheme.
CI2 refer to Charlson index using original weights, recoded as scores higher and equal to six collapsed into one category.
CI3 refers to Charlson index assigned the exact regression coefficient weights.
C1 refers to the Charlson comorbidities entered as binary variables into the model.
OR refers to odds ratio

Next to assess calibration, observed risk groups were plotted against the expected risk groups. The Hosmer Lemeshow Goodness-of-Fit Test divides subjects into deciles based on predicted probabilities, then computes a chi-square statistic from observed and expected frequencies. Points falling directly or close to the line show perfect calibration for the risk group. Points falling above the line indicate that the model underestimated the predicted risk of death in the ICU population. Likewise, when the points fell below the line, this suggests that the model over estimated the predicted risk of death. Low numbers indicate the risk groups that had a small predicted risk of death, while higher numbers present the risk groups that had a high risk of death.

Figure 11 calibration plot refers to the model with the original Charlson index. In this case, only 7/10 risk groups were identified, suggesting that the Charlson index was poor for differentiating risk groups. Three risk groups were dropped from STATA because they were ties (1,3,5). The model predicted well for high risk group (8&9), since these two points were closest to the line. However, the model was poor for low risk group (2) since this was far from the line. Overall, the model underestimated the risk of death for most of the groups, except two and ten. The other calibration graphs for the Charlson

index were not shown since they had similar calibration graphs to the original Charlson index.

Figure 11: Calibration of Charlson index using original weighting scheme to predict 365 day mortality



Note: Labels represent the risk groups.
Points represent the observed risk groups.

4.2.1. Secondary Research Questions:

How much is gained for mortality prediction when you add Charlson comorbidities to age and sex?

Next, to investigate how much gain in discrimination could be achieved when the Charlson index was added to other clinical variables, a series of logistic regression models were analyzed. In the first set of models (M1-M5), a baseline model consisting of age and sex was explored. Different variations of comorbidity measures were added to the baseline model to assess predictive ability. Without any comorbidity measures, the baseline model had fair discrimination (C=0.69). However, adding different variations of the Charlson score did increase the predictive power substantially but this increase was similar to the gain with the CHP component. In fact, using exact regression coefficients

(Model M4) did not substantially increase predictive power over when original Charlson index (Model M3). All models were statistically significant ($p<0.0001$) when the comorbidity measure was added to the baseline model based on the likelihood ratio test. To estimate the amount of shrinkage in the models, 100 bootstrap replications were done. This provides an estimate of the degree of over fitting within the given sample. Overall, shrinkage estimates were excellent in the models because the values were close to one. Additional analysis was conducted using the same models, excluding those who died within the ICU (Model N1- N5). No substantial differences in discrimination were observed. Overall, these models demonstrate that discrimination can be improved substantially by adding comorbidities to age and sex.

Appendix 8: Charlson index Model N1-N5 based on ICU survivors

Table 12: Comparison between Charlson index variations added to a baseline model to predict 365 day mortality (Models M1 to M5)

Models	LL	LR test	P value	C stat	AIC	Mean Shrinkage
Model M1: age sex	-3102.2			0.69		
Model M2: Model M1 + CHP	-2754.5	695.4	<0.0001	0.73	693.4	
Model M3: Model M1 + CI1	-2959.9	284.5	<0.0001	0.74	282.5	1
Model M4: Model M1 + CI3	-2875.1	454.2	<0.0001	0.76	452.2	1
Model M5: Model M1 + c.cob	-2838.7	527.1	<0.0001	0.77	493.1	0.98

Note: CHP is chronic health points.

CI1 is the Charlson index model using the original weights

CI3 is Charlson Index using newly derived exact regression coefficients.

C. cob is Charlson Comorbidities

To investigate how well APACHE II performs, compared to the Charlson index, APACHE II model and components were entered separately into a logistic regression model. The full APACHE II had good discrimination for predicting 365 day mortality. Acute physiology status and CHP each had poor discrimination when analyzed separately.

Table 13: Comparison of Apache II components in univariate models to predict 365 day mortality (Model A1 to A3)

Model	Covariate	C stat	OR	95% CI
A1	APACHE II	0.78	1.04	1.04, 1.04
A2	APS	0.63	1.07	1.06, 1.07
A3	CHP	0.63	3.15	2.79, 3.56

Note: APACHE II is the full APACHE II model including admitting diagnosis in the ICU.

APS is the acute physiology score, component of the APACHE II score.

CHP is the chronic health points, component of APACHE II score

Summary:

- Charlson index and variations has adequate ability to predict 1 year mortality ICU patients (C stat= 0.70 to 0.73) in univariate models.
- Adding the Charlson Index to models with age and sex provides statistically significant improvement in the model fit ($p<0.0001$) based on likelihood ratio test but does not provide much improvement in discrimination over a model with the chronic health points.
- Using empirically derived study weights or exact regression coefficients provides minimal increases in discrimination
- The Charlson index has poor calibration and most often underestimated the risk of death.
- The full APACHE II provides good discrimination for predicting 365 day mortality but performs poorly when only CHP or APS is examined.

4.2.2. How does the effect of comorbidity on mortality depend on age?

It is widely recognized that comorbidities depend on age. This section explores this relationship in greater detail and provides evidence on whether an interaction term could improve discrimination.

First to examine the Charlson index and Elixhauser distribution across age groups, the mean and median number of comorbidities was plotted for each method. As shown below, the mean number of comorbidities in Elixhauser and mean Charlson index increased with age, showing the 65-84 year old patients with the greatest burden of comorbidities. Interestingly, patients above 85 years old had a lower burden of comorbidities detected by both methods compared to 65-84 year olds. When the Charlson index was tested as a univariate predictor of 365 day mortality in different age groups, the best discrimination was achieved for age group 25-54 and declined for older age groups (Table 14). To gain a better understanding of how these age groups differed in terms of their comorbidities, we compared the comorbidity profile between those 85 years and older and those 25-54 years old. As expected, patients 85 years and older had a higher prevalence of comorbidities, especially congestive heart failure and myocardial infarction, than younger patient age groups.

Figure 12: A relationship between the mean and median Charlson index across various age groups

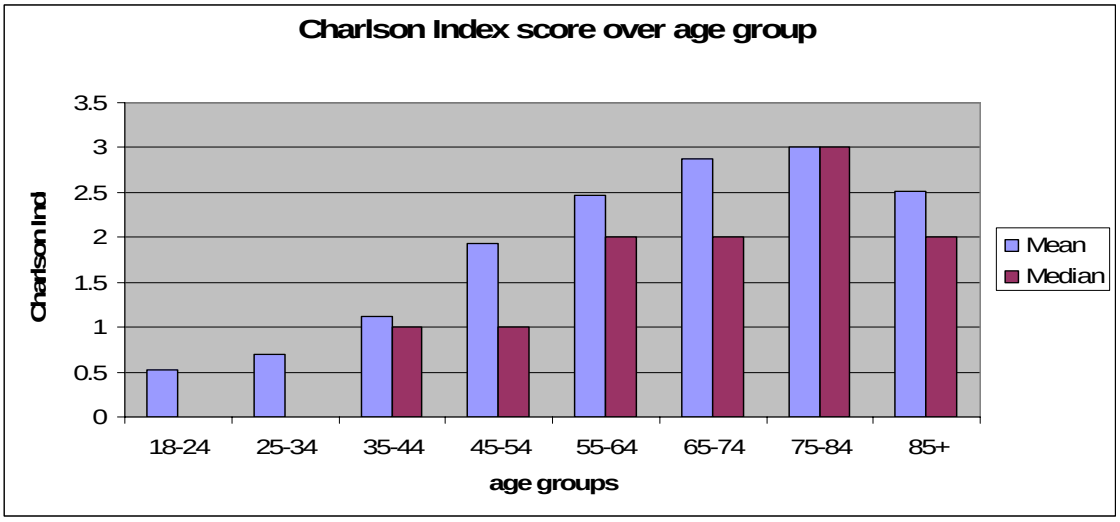


Figure 13: A relationship between mean and median total number of Elixhauser comorbidities across various age groups

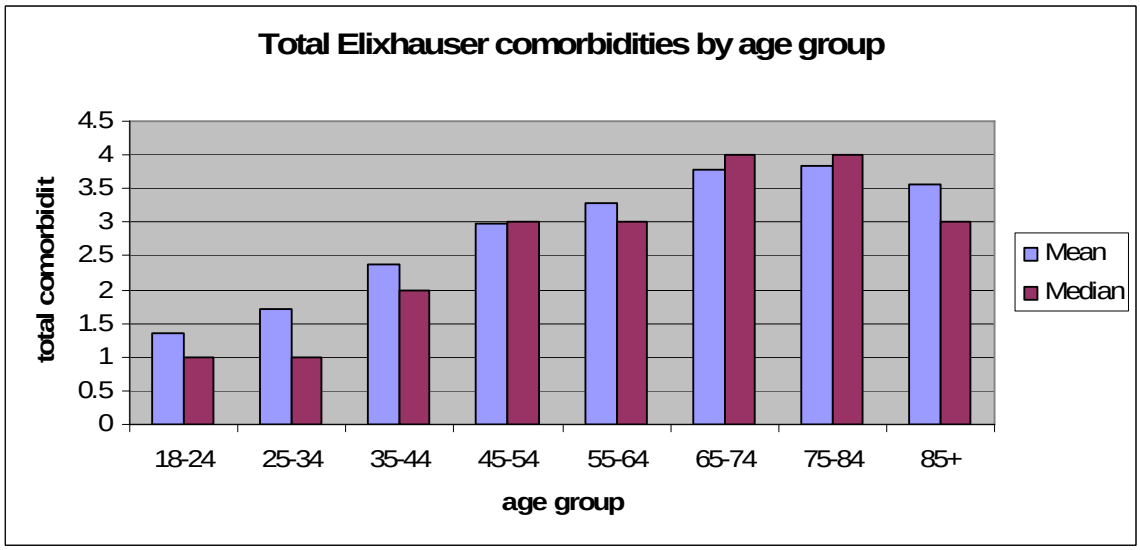
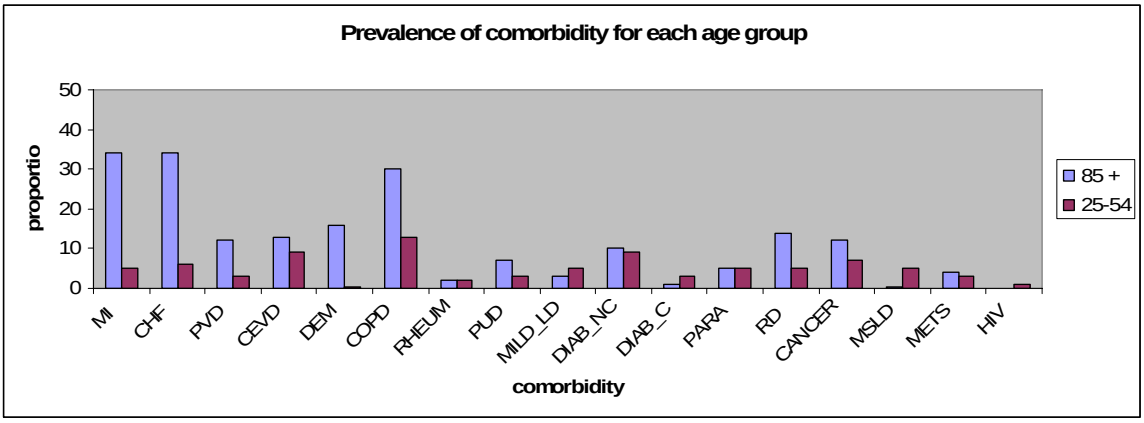


Table 14: Performance of Charlson index to predict 365 day mortality in specific age groups

Age group	C stat
18-44	0.67
25-54	0.71
35-64	0.70
45-74	0.68
55-84	0.64
65+	0.62

Figure 14: Prevalence of comorbidity for patients aged 85 years and older and 25-54 years old



Lastly, the effect of comorbidity on mortality changed according to age, so age and comorbidity was added as an interaction term to a model with age, Charlson index, and sex. Although it was statistically significant (LR test= 20.52, $p<0.01$, discrimination did not change much from a model without the interaction term (Model M6=0.74 vs. M3= 0.73).

Appendix 9: C stat matrix of all models

4.2.3. How do the new ICU specific weights differ from the original Charlson index weights?

As shown in the previous analysis, the models using empirically derived weights had slightly better discrimination than the models using the original Charlson index weights. Several studies have shown that the original weights were not applicable to their study population and deriving new study specific weights could improve predictive performance. In this section, the comparison between original weights and empirically derived weights is presented in greater detail.

Table 15 shows empirically derived Charlson index weights assigned to the ICU study sample, to predict one year mortality. All ICU weights derived were rounded to integer weights, using the same rounding scheme as Charlson. While the original Charlson index was based on medical inpatients using a Cox proportional hazards models, a logistic regression model to derive the regression coefficients for the ICU study sample was used. The odds ratio for 365 day mortality, or weight assigned, represents the odds of dying in 365 days for those who have that comorbidity compared to those without that comorbidity, adjusting for the other comorbidities in the model. Out of 17 comorbidities, 12 comorbidities were significant and received ICU weights to predict death. For newly assigned ICU weights, 6/17 comorbidities overestimated the original Charlson index weights. For example, moderate or severe liver disease was three times higher in the ICU odds ratio compared to the odds ratio of three. In contrast, metastatic carcinoma was only half the original Charlson weight (3 vs. 6). HIV had an odds ratio of one in the ICU, suggesting a small impact on mortality. However, the confidence intervals were wide suggesting the sample size was small to detect a significant impact on death.

When we plotted the odds of death using the new Charlson index weights, the relationship was almost linear. Charlson index of 10 or higher were collapsed into one category because the sample size was small (n=367) and highly skewed (skewness coefficient=10.0) In fact, the median Charlson index was 2 and the inter quartile range was 1 to 4, with a range from 1 to 216. The extreme range was due to the multiplicative properties of logarithms; therefore, comorbidities were multiplied rather than added together for a given patient. For example, if someone had several comorbidities, the rounded odds ratios for each comorbidity was multiplied rather than added together because of logarithmic properties (i.e. log(A) + log (B)= log (A*B)).

Table 15: New empirical ICU weights derived for Charlson

Comorbidity	OR (365 day)	P	95% CI Lower Limit	95% CI Upper Limit	ICU rounded weight	Original Charlson Index
Myocardial infarction	2.04	0	1.74	2.4	2	1
Congestive heart failure	1.68	0	1.43	1.97	2	1
Peripheral vascular disease	1.21	0.06	0.99	1.48	1	1
Cerebrovascular disease	3.02	0	2.47	3.69	3	1
Dementia	1.96	0	1.34	2.86	2	1
Chronic pulmonary disease	1.32	0	1.14	1.53	1	1
Rheumatoid arthritis	1.04	0.83	0.71	1.53		1
Ulcer disease	1.04	0.79	0.79	1.36		1
Mild liver disease	3.08	0	2.31	4.1	3	1
Diabetes non complicated	0.91	0.32	0.76	1.1		1
Diabetes complicated	1.05	0.73	0.78	1.43		2
Paraplegia or hemiplegia	0.87	0.36	0.64	1.18		2
Renal disease	1.91	0	1.55	2.36	2	2
Cancer	2.19	0	1.79	2.68	2	2
Moderate severe liver disease	8.87	0	6.27	12.55	9	3
Metastatic carcinoma	3.01	0	2.25	4.03	3	6
HIV	1.27	0.67	0.43	3.74	1	6

Note: OR- odds ratio
CI- confidence interval
P is the probability of observing a value greater than z.
P> 0.05 was considered not significant.
Blank spaces represent no weights assigned.

Summary:

- Empirically derived ICU new weights did not match the original Charlson index weights
- 12/17 Charlson comorbidities were significant for predicting 365 day mortality
- Derived Charlson index was highly skewed but had an approximately linear relationship with 365 day mortality

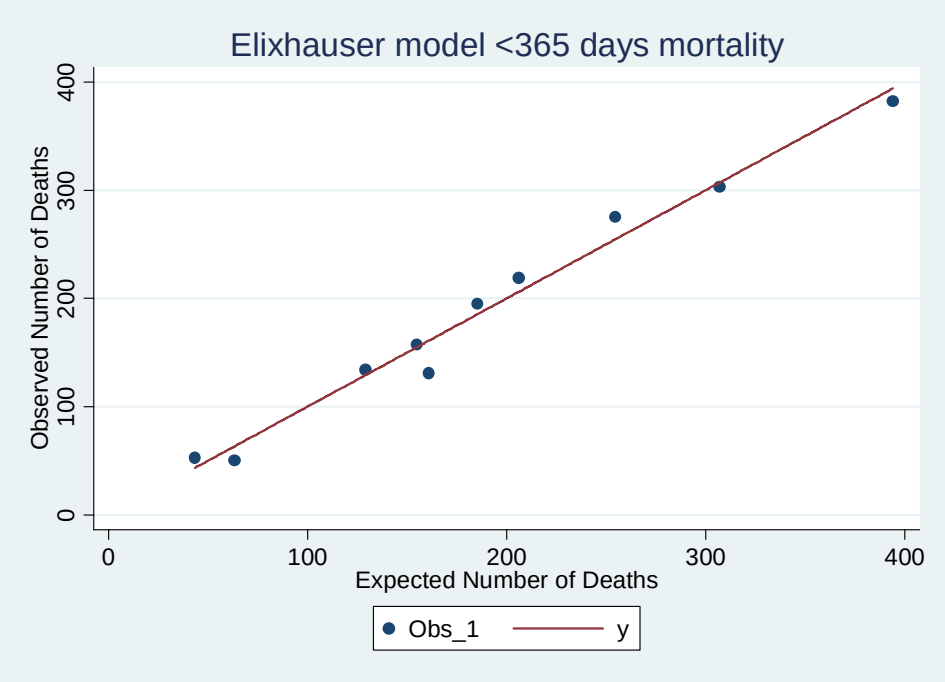
4.3.0. Primary Research Question :

How well does Elixhauser predict 1 year mortality in ICU patients?

To address this question, the predictive performance of Elixhauser on 365 day mortality was examined. A logistic regression model was used by entering the Elixhauser comorbidities as dummy variables into the Model E1. The results show that Elixhauser method adequately predicted 365 day mortality in ICU patients (C=0.74).

As part of the model assessment process, calibration plots were done for Elixhauser. Ten risk groups were identified and all points were close to the line suggesting that the model had excellent calibration.

Figure 15: Calibration of Elixhauser model to predict 365 day mortality (Model E1)



4.3.1. Secondary Research Questions:

How much is gained in terms of mortality prediction when Elixhauser comorbidities are added to age and sex?

In this next assessment, Elixhauser comorbidities were added to a model with age and sex to determine whether the predictive ability could be improved. The results are summarized below in Table 16. Adding Elixhauser increased the C stat by almost 10% for the ICU population (C stat= 0.69 to 0.77) and this was statistically significant based on the likelihood ratio test. These results suggest that the Elixhauser has very good predictive power when it is combined with other clinical variables. When bootstrapping calculations were done, Elixhauser had very little shrinkage despite the model having more than 30 covariates. Additional analysis was conducted using Elixhauser, excluding those who died within the ICU found in Appendix 10.

Appendix 10: Elixhauser multivariate models based on ICU survivors

Table 16: Comparison between baseline model and Elixhauser model to predict 365 day mortality (Model M1 and M7)

Models	LL	LR	p value	C stat	AIC	Shrinkage
Model M1: Age sex	-3102.2			0.69		
Model M7: Model M1 + E. cob	-2799.1	606.2	<0.0001	0.77		0.97

Note: E.cob represents Elixhauser comorbidities entered as dummy variables

LR refers to likelihood ratio test

LL refers to log likelihood

Summary:

- Elixhauser had adequate predict 365 day mortality in the ICU study sample (C stat=0.74)
- When Elixhauser was added to other clinical variables, the increase in discrimination was substantial (C=0.77) and statistically significant (p<0.0001)
- Elixhauser model had excellent calibration

4.3.2. What is the impact of a specific comorbidity on mortality?

To gain a better understanding of the effect of comorbidity on mortality, forest plots were constructed to represent univariate and multivariate models.

The first forest plot shows a list of Charlson comorbidities odds ratio for 95% CIs in univariate models, with 365 day mortality as the outcome Figure 16. In other words, comorbidities were modeled separately. For example, the odds of death were five times higher for those with liver disease compared to those without that comorbidity. (In other words, there is 95% confidence that the true population odds ratio lies between 4 and 9) Most of the comorbidities had an odds ratio greater than 1, except AIDS, connective tissue diseases, hemiplegia/paraplegia and diabetes without complications, which had 95% CI overlapping with one. Some of the CIs were extremely wide, suggesting that the prevalence of the particular comorbidity in the ICU was too small to obtain an accurate estimate. For example, AIDS had a 95% CI that ranged from approximately 0.3 to three.

To assess whether the 95% CI for comorbidities would change when a multivariate model was considered instead, a forest plots of odds ratios and 95% CI are shown for Charlson comorbidities in a multivariate model. For example, there is 95% confidence that the odds of death is between 2.5 and four for someone with mild liver disease compared to someone without mild liver disease, after controlling for the other Charlson comorbidities. In comparison to the forest plots representing the univariate model above, some of the comorbidities shifted left and were no longer significant for predicting 1 year mortality. Specifically, ulcer diseases, peripheral vascular disease, diabetes with complications, and hemiplegia or paraplegia were no longer significant for predicting 1 year mortality when adjusted for other comorbidities. However, some comorbidities shifted right when they were entered in a model with other comorbidities, suggesting a stronger influence on mortality or some effect modification between comorbidities. For example, moderate or severe liver disease had an odds ratio of 5.61 on its own, but increased to 8.87 in the adjusted multivariate model. This increase may suggest effect modification, because some comorbidities have stronger impact when combined with other comorbidities, compared to a model where a particular comorbidity is on its own, which results in a synergistic effect.

Figure 18 depicts a forest plot if the list of Elixhauser comorbidities in univariate models to predict 365 day mortality. In contrast to Charlson comorbidities, Elixhauser detected several comorbidities with odds ratios less than one (ex .alcohol abuse, obesity, psychoses, depression, drug abuse). This “protective” effect may be a result of coding bias where minor comorbidities are less likely to be coded if the patient dies and the more serious comorbidities are coded instead. Some comorbidities such as blood loss anemias, HIV, and peptic ulcer disease had wide confidence intervals and it may difficult to accurately assess the effect of the comorbidity on mortality due to the relatively small sample size.

Similar to the Charlson trend, when the Elixhauser comorbidities were assessed in a multivariate model, several comorbidities shifted left suggesting that they had a smaller influence on mortality when more important mortality predictors were in the model.

Figure 16: Relationship between Charlson comorbidities and 365 day mortality (univariate)

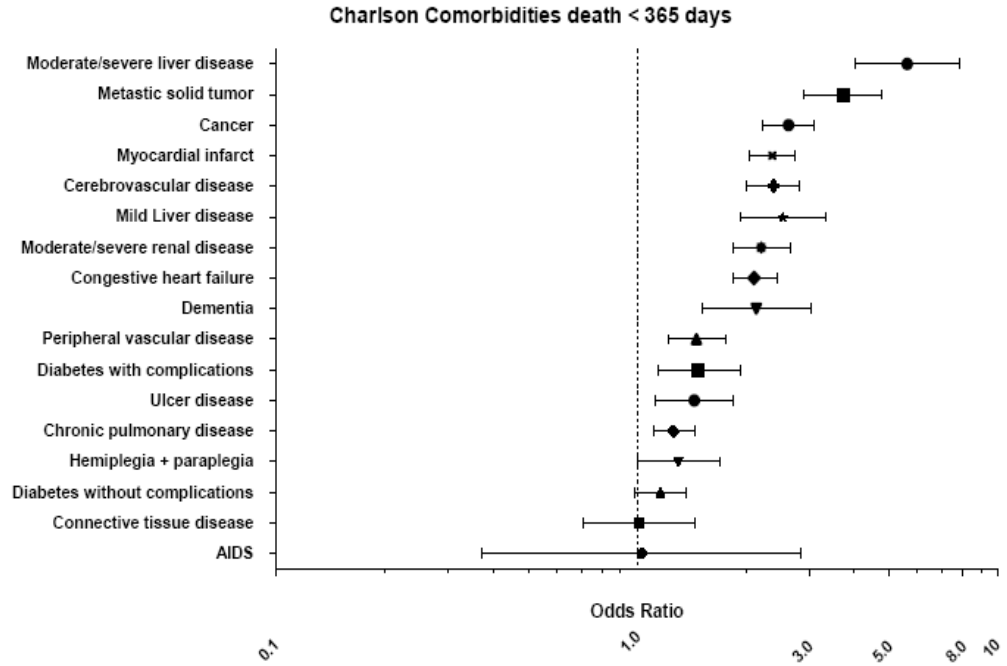


Figure 17: Relationships between Charlson comorbidities and 365 day mortality (multivariate)

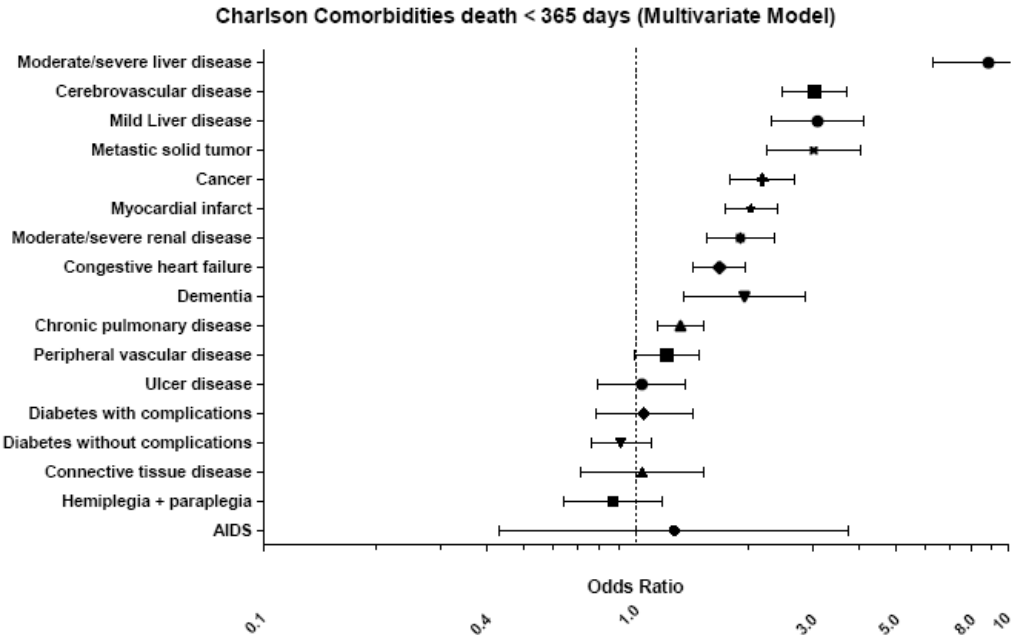


Figure 18: Relationship between Elixhauser comorbidities and 365 day mortality (univariate)

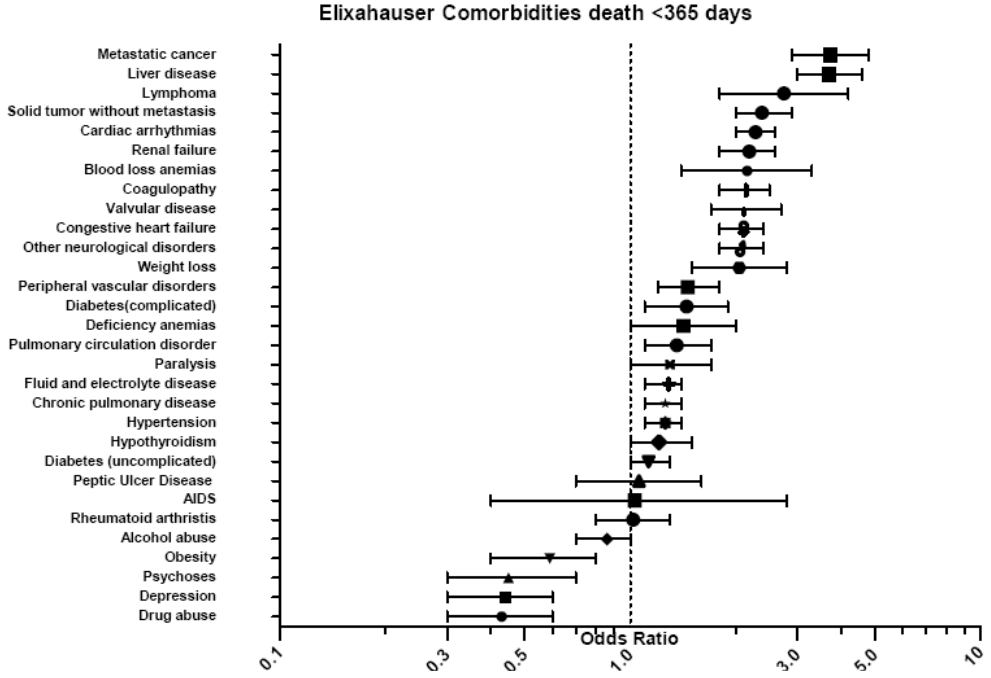
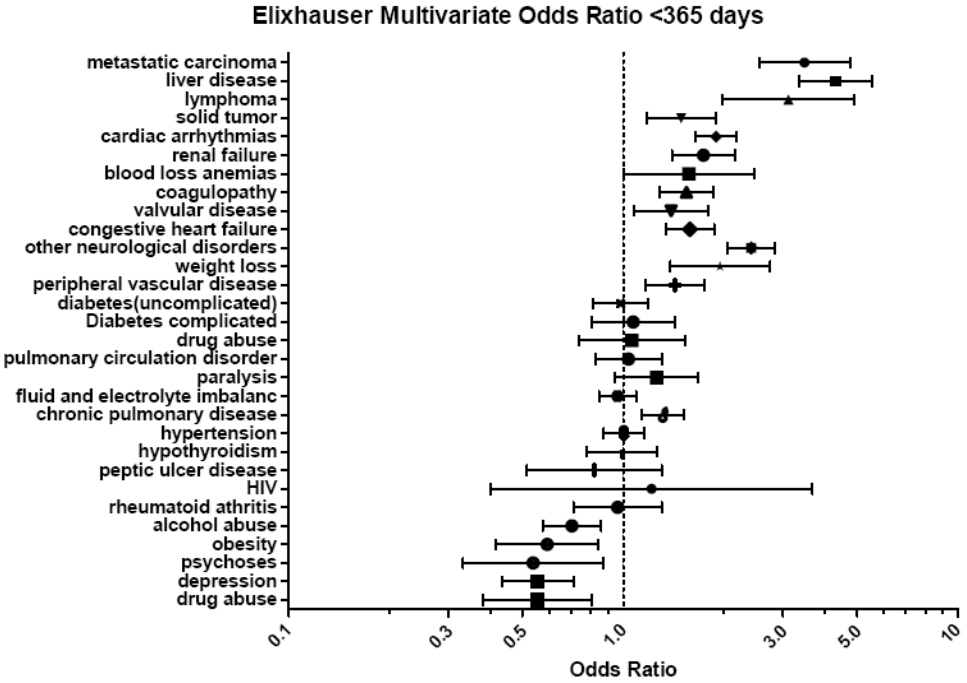


Figure 19: Relationship between Elixhauser comorbidities and 365 day mortality (multivariate)



Summary:

- In univariate models, most comorbidities had an odds ratio greater than one, suggesting that having these comorbidities were associated with an increased odds of death compared to not having the comorbidity.
- In the multivariate models, many comorbidities shifted left or were attenuated when added to the multivariate model, resulting in estimates closer to one and 95% CIs overlapping one.
- AIDS and HIV had confidence intervals had wider confidence intervals, indicating that there high variability of the estimate in the ICU study sample.
- Alcohol abuse, obesity, psychoses, depression, and drug abuse had odds ratios and 95% CIs less than one in the ICU study sample.

4.3.2. Does the predictive ability of Charlson and Elixhauser change according to ICU patient subgroups (trauma, surgical, ICU survivors)?

Given that the ICU study population is a heterogeneous population, comorbidity and physiological profiles would be different depending on the subgroup or risk group of patients explored. This difference may influence how well a comorbidity score performs and can offer insight to which comorbidity measure would be most appropriate for a specific ICU subgroup. To investigate these trends, Charlson and Elixhauser methods were explored by different subgroups using a diagnostic category. Logistic regression models were constructed to examine the performance of the C stat between subgroups.

Table 17, Table 18, and Table 19 show the burden of comorbidity and the differences in other clinical indicators between specific ICU population subgroups. In total, there were 372 trauma patients. Both Elixhauser and Charlson showed that trauma patients had less comorbidity than non trauma patients. As expected, trauma patients were on average younger, mostly male, and had a lower mortality rate than non trauma patients.

In Table 18 non-surgical patients had a slightly higher burden of comorbidity compared to surgical patients. This makes sense since patients with a lot of comorbidities and low physiological reserves would not be good candidates for risky surgical procedures and thus they would have a higher mortality rate, as shown in the data. Finally, both Charlson and Elixhauser showed that those who died in ICU had more comorbidities than those who survived their ICU stay. ICU survivors had lower APACHE II scores and their long term survival may be dependent on comorbidity status, rather than physiological variables at the time of ICU admission.

Table 17: Comparison of descriptive variables, comorbidities, and clinical outcomes between trauma and non trauma patients

	Trauma	**Non trauma**
Patients	372 (7.3%)	4715 (92.7%)
Total number of Elixhauser cob for each patient (Median, IQR)	1 (1, 2)	3 (2,5)
Total number of Charlson cob for each patient (Median , IQR)	0 (0,1)	1 (1,2)
Mean age + SD	42.1+ 19.4	60.35+ 17.9
In-hospital death	66 (17.7%)	1349 (28.6%)
90 day mortality	69 (18.6%)	1532(32.5%)
180 day mortality	70 (18.8%)	1651 (35.0%)
365 day mortality	73 (19.6%)	1799 (38.2%)
APACHE II +SD	17.7± 7.2	19.4± 8.9
Male (%)	285 (76.6%)	2542 (54.1%)

Table 18: Comparison of descriptive variables, comorbidities, and clinical outcomes between surgical and non surgical patients

	Surgical	Non Surgical
Patients	1663 (32.7%)	3429 (67.3%)
Total number of Elixhauser cob for each patient (Median, IQR)	3 (1,4)	3 (2,5)
Total number of Charlson cob for each patient (Median , IQR)	1 (0,2)	1 (0,2)
Charlson index(recoded 6+)	1(0,3)	2(0,3)
Mean age ± SD	59.4 ± 19.5	58.8 ± 18.3
In-hospital death	323 (19.4%)	1090 (31.8%)
90 day mortality	377 (22.7%)	1224(35.7%)
180 day mortality	412 (24.8%)	1310 (38.2%)
365 day mortality	463 (27.8%)	1410 (41.1%)
APACHE II ± SD	19.1 ± 7.9	19.4 ± 8.9
Male (%)	725 (43.6%)	1537 (44.8%)

Table 19: Comparison of descriptive variables, comorbidities, and clinical outcomes between ICU survivors and non ICU survivors

	ICU survivors	Non ICU survivors
Patients	4173(80.9%)	986 (19.1%)
Total number of Elixhauser cob for each patient (Median, IQR)	3 (2, 4)	4 (2,5)
Charlson index(recoded 6+) (Median, IQR)	1 (0,3)	2 (1,4)
Total number of Charlson cob for each patient (Median , IQR)	1 (0,2)	2 (1,3)
Mean age ± SD	57.5 ± 19	65.3±16.5
APACHE II ±SD	17.8 ± 7.8	25.7± 8.9
Male (%)	1841(44.1%)	449(45.6%)

Table 20: Comparison between Charlson index and Elixhauser in ICU patient subgroups for predicting 365 day mortality

Model	C stat	OR	95% CI	Patient population
CI1	0.70	1.33	1.30 1.37	Non-Trauma
CI1	0.64	1.42	1.23 1.64	Trauma
CI1	0.70	1.43	1.38 1.49	Non surgical
CI1	0.70	1.27	1.22 1.32	Surgical
CI1	0.74	1.38	1.34 1.43	ICU survivors
E1	0.74			Non-trauma
E1	0.75			Trauma
E1	0.75			Non surgical
E1	0.75			Surgical
E1	0.77			ICU survivors

Table 20 shows that the predictive performance changes according to ICU subgroup and comorbidity measure. For instance the C stat for trauma patients using the Charlson index was severely degraded, but when the Elixhauser method was used, no substantial differences was seen between the trauma and non-trauma sample. The best discrimination was achieved when Elixhauser was used on ICU patients that survived their ICU stay.

Summary:

- The predictive performance of the Charlson index in trauma patients (C=0.64) was much lower than non trauma patients (C=0.70), in contrast to Elixhauser where the performance remained the same (C= 0.74 to 0.75)
- No difference was observed in surgical and non surgical patients, in terms of predictive performance of the Elixhauser and Charlson index
- The sample including only patients that survived their ICU stay had the best predictive performance using Elixhauser and Charlson index

4.4.0. Primary Research Question:

Are there comorbidities missing in Charlson and Elixhauser that can improve risk adjustment further in the ICU population?

This section reviews the results obtained from the Delphi process when ICU physicians were asked to identify whether Charlson and Elixhauser was missing comorbidities that predicted mortality in the ICU population. Both Elixhauser and Charlson Index were developed on non ICU study samples, therefore, important comorbidities could be missing which could improve discrimination.

Figure 20 presents the participation flow chart diagram outlining the number of physicians that were recruited for the study and the number completing the Delphi processes. While nineteen physicians were recruited for the study, six were loss to follow up, and only thirteen completed the whole process. Several follow up strategies were implemented to encourage participation such as lengthening the time of survey completion, emails, and verbal remainders at ICU rounds.

Figure 20: Study participation in Delphi Process

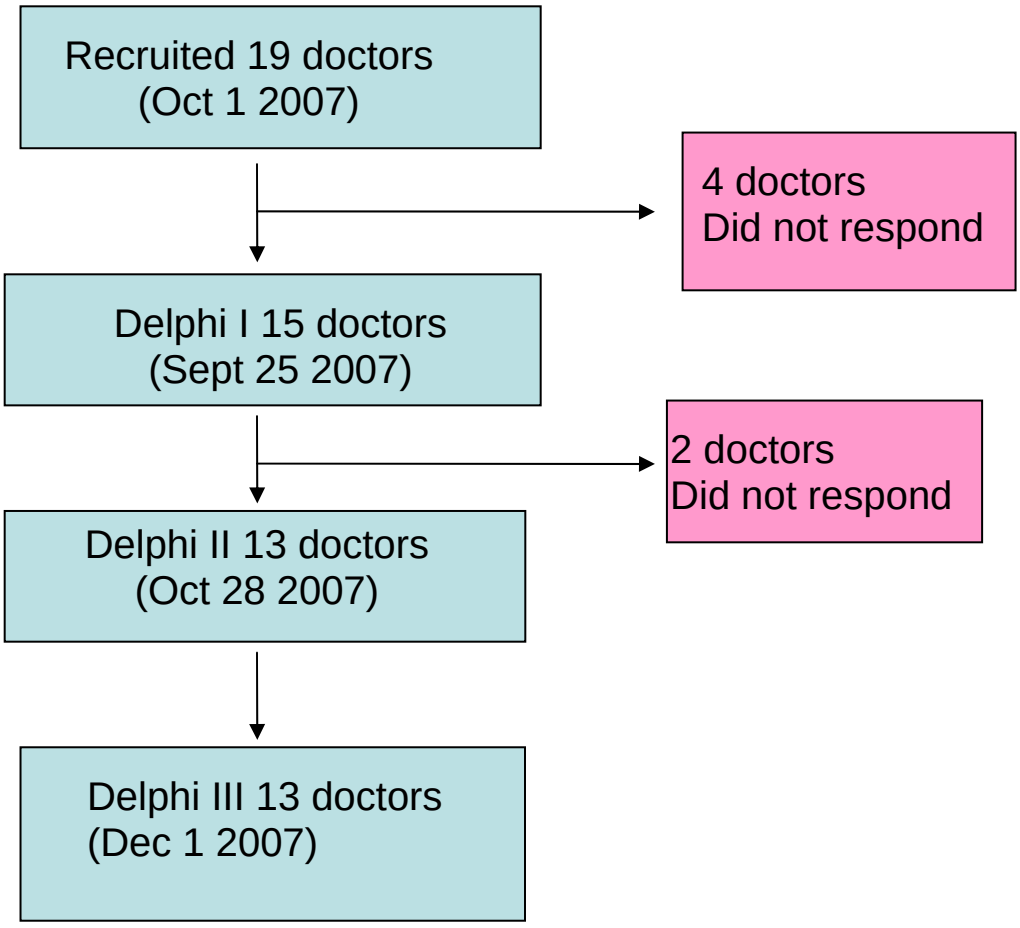


Table 21 presents the results from the first and second step of the Delphi process when ICU physicians were asked to review the list of Charlson and Elixhauser comorbidities and indicate whether any additional comorbidity should be added to the list. This list combined all Charlson and Elixhauser comorbidities and presented to all the ICU physicians. Based on the new saturated list, 19 physicians provided responses on comorbidities as being relevant or not relevant in terms of mortality prediction for 90 day, 180 day, and 365 day mortality. The group response was the percentage of physicians that agreed comorbidity was relevant for predicting mortality. The group response was calculated as the number of physicians that agreed comorbidity was relevant for a mortality outcome divided by the total number of physicians that participated in the first Delphi step (n=19). Brain injury, chronic decubitus ulcer, cystic fibrosis, immunodeficiency, and neuromuscular disorders were additional comorbidities

identified by the majority of physicians as being relevant for mortality, while comorbidities such as hearing loss, GERD, and hyperlipidemia were identified as not relevant. Compared to 90 day mortality, more physicians believed that comorbidity was relevant for long term survival as indicated by the group response, with the exception of a few comorbidities.

Table 21: Additional comorbidities identified by ICU physicians (N=19) for predicting 365 day mortality

Comorbidities	90 day	180 day	365 day
Severe Asthma	30.8	7.7	15.4
Blindness	7.7	7.7	7.7
Hearing Loss	0	0	0
Developmental delay	7.7	7.7	7.7
Inflammatory Bowel disease	7.7	7.7	15.4
Neuromuscular disorders	46.2	46.2	76.9
GERD	0	0	0
Restrictive chest wall disease	15.4	7.7	23.1
Brain injury	61.5	76.9	84.6
Hyperlipidemia	0	0	0
Obstructive sleep apnea	23.1	30.8	38.5
Chronic decubitus ulcer	8.5	53.8	69.2
Cystic fibrosis	38.5	53.8	69.2
Immunodeficiency	38.5	38.5	61.5
Spinal cord injury	30.8	30.8	38.5

Note: GERD is gastroesophageal reflux disease.

In the last round, physicians reviewed the group response and had the opportunity to rank the top 30 comorbidities from 1 to 10, with 10 being the most relevant for predicting mortality. Mean relevancy scores represent the mean rank that ICU physicians (n=13) applied for each of the comorbidities to predict 365 day mortality. Table 22 summarizes the top 30 ranked comorbidities identified through the Delphi process and the odds ratios obtained from the multivariate model. Odds ratios were derived from the multivariate models for Charlson and Elixhauser (Model C1 and E1). In the case when a comorbidity was found in Charlson and Elixhauser, we used the Elixhauser OR instead because the odds ratio adjusted for more comorbidity. From the list derived by the ICU physicians, 18 comorbidities were not included in the Charlson comorbidities and 11 were not included in the Elixhauser list. Some comorbidities such as coagulopathy, cardiac arrhythmias, neurological disorders, valvular disease, and weight loss had low mean ranks, which did not correspond to the odds ratio. In other words, for these comorbidities, physicians ranked these comorbidities as least relevant for predicting

mortality, when the comorbidity corresponded with a high odds ratio in the multivariate model. Pulmonary circulation disorders, on the other hand, received a high rank (Mean rank=6.4) but the OR was small (OR=1.04). In this case, ICU physicians thought the comorbidity was important for long term survival, but the impact was small in the adjusted model. The rest of the comorbidities showed fairly good agreement between OR and mean relevancy ranks. However, blood loss anemias was not identified as one of the top 30 comorbidities by ICU physicians even though the OR was high (OR=1.57). Figure 21 is a graph showing the correlation between comorbidity mean relevancy score and odds ratio. The Pearson product-moment correlation coefficient was 0.2, suggesting a weak linear association between mean relevancy score and the odds ratio. When comorbidities with a prevalence less than 5% in the ICU sample were removed, this did not improve correlation (r^2=0.18). Excluded comorbidities included blood loss anemias, HIV and lymphoma.

Table 22: Mean relevancy scores and odds ratio for comorbidities

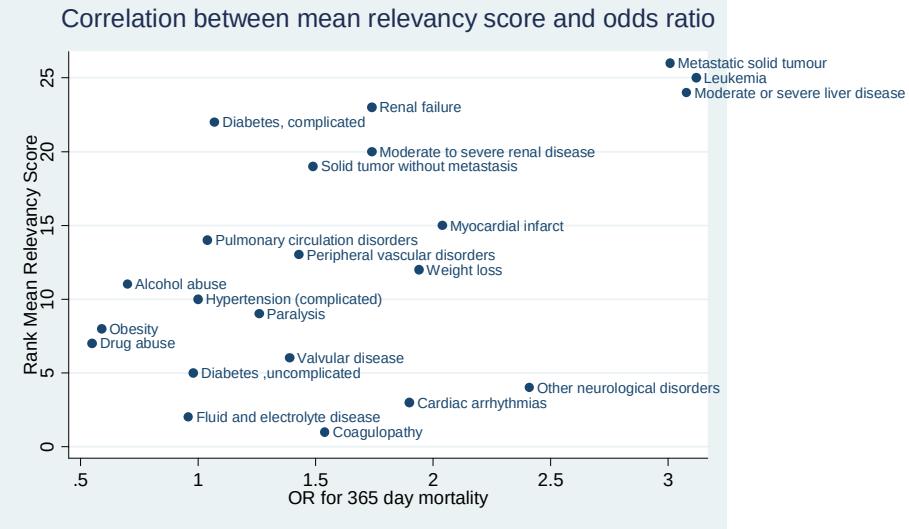
Comorbidities	Mean Relevancy Rank for predicting 365 day mortality	OR (multivariate 365 days)	Charlson	Elixhauser
Coagulopathy	2.4	1.54		X
Fluid and electrolyte disease	2.8	0.96		X
Cardiac arrhythmias	3.4	1.9		X
Other neurological disorders	3.8	2.41		X
Diabetes ,uncomplicated	4.4	0.98	X	X
Valvular disease	4.8	1.39		X
Drug abuse	4.8	0.55		X
Obesity	4.8	0.59		X
Paralysis	4.8	1.26		X
Neuromuscular disorders	5.0			
Immunodeficiency	5.12			
Hypertension (complicated)	5.4	1.00		X
Alcohol abuse	5.6	0.7		X
Brain injury	5.6			
Weight loss	5.8	1.94		X
Peripheral vascular disorders	6.0	1.43	X	X
Lymphoma	6.2	3.12	X	X
AIDS	6.2	1.21	X	X
Pulmonary circulation disorders	6.4	1.04		X
Myocardial infarct	6.6	2.04	X	
Cerebrovascular disease	6.6	3.02	X	
Chronic decubitus ulcer	6.7			
Dementia	6.8	1.96	X	
Chronic pulmonary disease	7.2	1.31	X	X
Solid tumor without metastasis	7.2	1.49		X
Moderate to severe renal disease	7.2	1.74	X	
Cystic fibrosis	7.2			
Congestive heart failure	7.4	1.58	X	X
Diabetes, complicated	7.4	1.07	X	X
Renal failure	8.0	1.7		X

Moderate or severe liver disease	8.2	3.08	X	X
Leukemia	9.0	2.19	X	
Metastatic solid tumor	9.6	3.01	X	

Yellow rows represent the comorbidity not included in Elixhauser or Charlson.

Since chronic decubitus ulcer had a mean rank above six, the proportion of ICU patients who had this comorbidity was calculated and entered it into a logistic regression model with all the Elixhauser Comorbidities (Model D2). Although, cystic fibrosis had a mean rank of 7.2, this was not further explored because of the relatively low prevalence in the ICU patients, as noted in the literature. For the other comorbidities, such as neuromuscular disorders, brain injury, and immunodeficiency, no further analysis was done because these were poorly defined and overlapped with some comorbidities previously defined. To identify pressure sore in administrative data, we used the code for plaster ulcer, pressure sore and bed sore in ICD 9 as L90 and ICD 10 as 707.0. In this sample, 134 patients (2%) had pressure sores coded in-hospital discharge data. As a univariate predictor of 365 day mortality, chronic decubitus ulcer was significantly associated with mortality (OR=1.53, 95%CI: 1.09, 2.17) (Model D1). However, when chronic decubitus ulcer was added to a model with Elixhauser comorbidities, it was no longer significant (OR1.28, 95% CI:0.87,1.87) and could not provide any predictive power to the model (Model D2). Therefore, these results suggest that no other comorbidities can be added to either the Charlson or Elixhauser list to improve prediction for ICU risk adjustment.

Figure 21: Correlation between comorbidity mean relevancy score and odds ratio



Summary:

- No other comorbidities can be added to both Charlson and Elixhauser list to improve prediction for ICU risk adjustment.
- ICU physicians indicated that most comorbidities were important for predicting 1 year mortality, compared to shorter term mortality outcomes (ex. 90 days, 180 days)
- Elixhauser provides a more complete comorbidity profile of ICU patients than the Charlson comorbidities list.
- ICU physicians identified chronic decubitus ulcer and cystic fibrosis as additional comorbidities relevant for mortality prediction in ICU patients
- Chronic decubitus ulcer was significantly associated with mortality (OR=1.53, 95%CI:1.09,2.17), but was no longer significant in a model with Elixhauser comorbidities (OR1.28, 95% CI:0.87,1.87)
- Although ICU physicians identified other comorbidities not present in Elixhauser or Charlson, these comorbidities were not explored further due to a number of reasons such as low prevalence, poorly defined, poor agreement among other ICU physicians, and some overlap with previously defined comorbidities.

CHAPTER 5: DISCUSSION

5.1. Results summary

This study investigated the use of the Charlson index and Elixhauser comorbidity score as possible risk adjustment methods for the ICU population. This work builds on previous research on comorbidity scores by examining how these scores can be added to administrative data to provide additional predictive power for risk adjustment in the ICU population.

This research adds to the body of literature on using comorbidity scores for risk adjustment in a number of important ways. It is the first study that provides evidence that Charlson and Elixhauser comorbidity lists are complete, with no other obvious comorbidities missing which might provide further explanatory power for risk adjustment in the ICU. Although ICU physicians identified a list of additional comorbidities, most were minor comorbidities that did not receive much agreement from other physicians. Only cystic fibrosis and chronic decubitus ulcer received relatively high rankings (6.7 and 7.2 respectively) but when the latter was explored further in a multivariate model, it could not provide any additional predictive power. Therefore, these findings suggest that the comorbidities scores contain all important comorbidities that are associated with mortality for the ICU population.

This study is the only research that has explored using Charlson index and Elixhauser to predict one year mortality outcome in the ICU population. All the ICU severity illness scores have focused on short term mortality outcomes, failing to recognize the importance of using these risk adjustment methods to examine long term mortality outcomes. The results from this study suggest that both Elixhauser and Charlson index provide adequate ability to predict one year mortality and by adding administrative variables, such as age and sex, this could substantially improve predictive power (C=0.69 vs. C=0.77). Although the conventional method for risk adjustment in the ICU involves using severity illness scores, the real advantage with using comorbidity scores pertain to their availability in administrative data. Using administrative data to collect comorbidities and long term mortality outcomes is a cost-effective methodology because it does not require primary data collection from chart reviews. Further, administrative data captures virtually all patients that enter the health care system and this

allows researchers to apply the same comorbidity score on all study populations, thus facilitating comparisons across ICU populations. In contrast, severity illness scores are not collected consistently across ICUs because they are labor intensive which has limited their uptake across ICUs. As such, this challenge makes it difficult to compare across ICUs. Further, these scores focus on short term mortality outcomes such as in hospital mortality, and were never validated on longer term outcomes.

This research also examined in detail variations of the Charlson index score, in specific subgroups and compared the predictive ability to Elixhauser. Compared to the Charlson index, Elixhauser had better calibration in the ICU population. In this study, much work was done to improve the performance of the Charlson index by readjusting weights. Despite these efforts, the change in predictive power was small and did not provide a practical advantage over the original weighting scheme (C=0.70 vs. 0.73). There are some reasons why only a small increase was observed. First, the degree of difference between the original Charlson index and new derived weights may not affect a large proportion of the population. The comorbidities that had substantially different weights from the original Charlson index were metastatic carcinoma, moderate/severe liver disease and HIV. All three comorbidities occurred in less than 6% of the ICU population, therefore, most of the change using new weights would only occur within this subgroup. Secondly, calibration in the Charlson index was poor as there were only seven risk groups identified and the rest were ties. When the plots for calibration were examined, risk group #3 or #2 substantially overestimated the number of deaths (Figure 11). With the Charlson index, 50% of the sample had scores 2 or less , suggesting that the Charlson index could be poor for differentiating those with a high risk vs. low risk of death. Other researchers have also found readjusting weights provided small increases in predictive ability over using the original weights (45, 46); however, they have suggested to re-adjust weights in order to improve risk adjustment. Contrary to these authors standpoint, the small gain received from readjusting weights does not provide a practical advantage to risk adjustment. Further, the weights derived can only be applied to a specific population which can limit comparisons across different study populations where new weights must be derived again. This is similar to the problem when new versions of severity illness scores such as SAPS are updated and the weights are re-estimated,

thereby limiting comparisons between ICUs that use different versions of the same scoring system.

Elixhauser, on the other hand, does not assign weights to comorbidities and still provides adequate discrimination (C stat=0.74) in the ICU population. When added to other demographic variables like age and sex, Elixhauser had good discrimination to predict 1 year mortality (C=0.79). Further when ICU subgroups were compared, Elixhauser provided stable predictive power (C=0.74 to 0.77), while the Charlson index varied greatly according to ICU subgroup population (C=0.64 to 0.74). Certain comorbidities like drug abuse, fluid and electrolyte disease, hypertension, drug and alcohol abuse were not included in Charlson because they were not statistically significant for predicting mortality in medical inpatients. However, these comorbidities may be found in ICU subgroups such as trauma patients, and therefore Elixhauser may perform better than Charlson because these comorbidities are accounted for. For example, Charlson did not find coagulopathy to be independently associated with mortality among medical patients, but Romano found it to be an important predictor of trauma death (54). Similarly, congenital coagulopathy and hypertension were identified as comorbidities that were significantly associated with mortality in studies of trauma patients (97, 98) but were not included in the Charlson index. As shown in this study, failing to account for comorbidities that are associated with mortality can severely degrade the ability to risk adjust in certain sub-groups of the ICU.

Our results are similar to Johnston *et al*,(22), such that adding clinical variable to comorbidity scores can improve discrimination substantially. However, this differs from a recent study done by Ho *et al* and Nuttall(19, 99), where they found that adding comorbidity scores did not provide any significant improvement in discrimination . When Johnston added other clinical variable such as age, laboratory values, principal diagnosis and admission source to their model with Elixhauser, this improved discrimination significantly (C stat=0.70 vs. 0.88). Likewise, in our study when age and sex were added to a model with Elixhauser comorbidities, this also provided a significant increase (C stat=0.69 vs. 0.77). One possible reason to explain why the other study could not detect any significant gain in discrimination when comorbidities were added could be related to the completeness of comorbidity coding in their datasets. For example, Ho *et al*,

demonstrated in Western Australia hospital morbidity database that only 615 (14.9%), 10 223 (42.1%), and 11 597 (47.7%) patients were identified as having at least one comorbidity, as defined in the APACHE II score, Charlson comorbidity index, and Elixhauser comorbidities, respectively. These results are quite different than our study and Johnston where 93% and 85.1% of patients had at least one comorbidity as defined by Elixhauser. These results could suggest that the burden of comorbidities in the Western Australian ICU was much less compared to our study or there may be some under coding of comorbidities in the Western Australia database. In fact, many studies have also cited a problem with under-coding of comorbidities in administrative databases compared to chart review. Comorbidities such as cerebrovascular disease, malignancy, myocardial infarction, peripheral vascular disease were citied as being under-reported in administrative data using hospital discharge data from the Calgary Health Region(72). Researchers have suggested that diseases that are asymptomatic are less likely to be coded in administrative data compared to chart review(84). Therefore, researchers should be aware of these discrepancies when using administrative data and understand that the completeness and accuracy of comorbidity coding in administrative data will determine how useful the scores are for risk adjustment.

Another interesting finding from this study showed that five comorbidities had protective effects on mortality (alcohol abuse, depression, drug abuse, obesity, and psychoses), only two of these comorbidities were common with Johnston's results. Drug abuse and alcohol abuse had protective effects on mortality for both studies and this may be a result of age. Alcohol and drug problems are more prevalent in the younger ages, by controlling for age, this protective effect may disappear. In our sample, the mean age for a patient with these protective comorbidities was less than the average age of the sample (Age 43-57 vs. 59). Interestingly, in a multivariate model with only comorbidities, the odds ratios for alcohol abuse, drug abuse, obesity were significant, 0.70(95%CI: 0.57-0.86), 0.55(95%CI 0.38-0.80),0.59 (95% CI: 0.42-0.84), respectively. When the same model was run again with age, all three comorbidities were no longer significant, 0.82 (95%CI: 0.66- 1.01), 0.77 (95%CI: 0.52 -1.14), 0.76(95% CI: 0.53-1.09), respectively. These findings suggest that a "protective effect" seen by some comorbidity may be explained by confounders such as age. Other studies have noted a protective effect seen

with some comorbidities and have suggested that it may be a result of systematic bias in coding(22). Specifically, Iezzoni and Jencks *et al,* identified an issue known as recording bias where the likelihood of a chronic diagnosis being reported is reduced if the patient dies(60, 100, 101). Jencks hypothesizes that coders may substitute acute complications for comorbidities in these situations. However, when Romano *et al.* explored this issue they only found reduced sensitivity in comorbidity detection in those who died, when they truncated into five diagnosis abstracts, compared to 9 and 25 diagnosis abstracts(102). For our study, there was up to 16 diagnosis codes available but we are unaware if this same effect was present in Canadian administrative data.

In this study, we explored the relationship between comorbidity and age in the ICU study sample. The finding showed the mean and median Charlson index score increases up to age 84 and decreases for those 85 years and older. Similarly, this relationship was also apparent when we examined the total count of Elixhauser comorbidities. Further analysis showed that the relationship between comorbidities and one year mortality was strongest in the age group 23-54, with the highest discrimination (C= 0.71), and weakest in those 85 years and older (C= 0.62). Based on intuition, one would expect that the number of comorbidities would increase with age and those with the oldest age would be in the most severe stage of the comorbidity; therefore the relationship between comorbidity and death would be strongest in the oldest age category. However, this was not the case and there are a few potential reasons to explain the discrepancy in the results. First, a generation effect may be apparent in those 85 years and older, such that these individuals represent a group with less comorbidities or an earlier stage of comorbidity compared to their later generation counterparts. Since Elixhauser and Charlson provide no indication of the severity of the comorbidity, this limitation makes it difficult to assess the severity of the comorbidities across age groups. Secondly, selection bias may occur when two individuals with the same comorbidity but different ages are permitted to enter to the ICU based on their age. For example, ICUs may be more likely to admit a 25 year old with valvular disease rather than a 105 year old with the same disease and severity.

A more recent study conducted by Froehner *et al*, (2008), also demonstrated differences in discrimination between Charlson Index and age groups for patients

selected for radical prostatectomy. In their study, they found that when patients 70.0 years or older were included in the model, this resulted in a decrease in discrimination. However, the discriminative ability of Charlson index comorbidity was greatest in the age group that was 63.0 to 69.9 years old, compared to those younger than 63 years old and older than 70 years old. These results differ from the ICU study sample and further studies are required to determine if these trends are persistent in the ICU population.

5.2. Strengths and limitations of the study

The present study had several strengths and some limitations. First, Canadian hospital discharge databases are less susceptible to coding biases compared to American databases because of a single payer government infrastructure. Therefore, the data is less susceptible to "up coding" practices which have been a problem in Medicare data. Both Canada and the United State's hospital reimbursement rely on the extent of coding in administrative data. This provides a financial incentive to encourage comprehensive coding of comorbidities.

Secondly, this research was the first ICU study that examined the comorbidity scores using ICD -10 algorithms in Canada. Around half of the study data was based on ICD-10 codes for comorbidities but we did not notice any improvement in coding for comorbidities when the ICD-10 was introduced. The prevalence of comorbidities between ICD-9 and ICD-10 were similar (data not shown), despite ICD-10 having more specifically coded conditions. Since ICD-10 was introduced in 2002, it may take some time for coders to be familiar with the new coding scheme and improvements in coding may occur in the future.

Next, this study was able to exclude complications from being coded as comorbidity by using the "diagnosis type indicator" found in the administrative data. This variable occurred with every diagnosis code to flag whether the diagnosis was present at time of hospitalization or after admission. The administrative hospital discharge data contained up to 16 diagnosis codes and types as well as 10 procedure codes, which provides sufficient coding spaces for comorbidities to be entered in the administrative record. Given that only the hospital discharge record associated with the ICU admission was used to collect data on comorbidities, this may miss some comorbidity coded only in

previous records. However, we didn't expect this to make a significant impact on the performance of our model since some studies found that additional information from other hospital admissions would yield only minor improvements in model performance, using either comorbidity methods(18).

In our study, we collected ICU clinical data prospectively and linked 97% of our study sample to their ICU clinical data, hospital discharge record and mortality outcome. Mortality outcome data were verified by two sources, vital statistics and the hospital discharge record. All critically ill patients admitted to an ICU in the CHR during the study period were included in the sample, therefore reducing susceptibility to selection bias.

The study did examine the extent of over fit by calculating the amount of shrinkage in the multivariate models. Over-fitting can be a problem when models are derived on small sample sizes using many variables that are specific to a particular sample(25). Our results show that there were very little shrinkage in the models and the estimated coefficients suggest these results can be applied to similar populations to predict outcomes accurately. In our study, 37% of the patients died within a year which provided sufficient numbers to minimize over-fitting the data.

There are some limitations to this study that should be acknowledged. First, it is unknown whether coding errors such as unbundling or misspecification are present within the study. These two forms of errors would represent a source of systematic bias in a study because they could be repeated several times for a specific coding scheme for comorbidity. Systematic error would have the potential to bias the risk in either direction away from the null hypothesis, depending on the type of diagnostic coding error. When coders assign codes for all the separate parts of a diagnosis instead of assigning a code for the overall diagnosis, this is known as unbundling(71). This might occur as a result of coder training, workload issues, or a lack of standardization across facilities for coding certain comorbidities.

Second, our sample consisted of only CHR ICUs which can limit its generalizability to rural hospitals with ICUs or other ICUs across the country. Changes in coding practices or patient population characteristics across provinces can affect the

performance of a risk adjustment method; therefore, more studies are needed across the country to determine if the results can be generalized to other Canadian hospitals. Similarly, the Delphi process was conducted in the Calgary Health Region and the majority of physicians in the Delphi are from the same region (10/13). It is unknown whether these results would be generalizable to other ICUs where comorbidity profiles can be different. We did not obtain any information about demographic variables of the ICU physicians which could be useful to assess whether these characteristics are similar across ICU facilities. An additional meeting with the expert panel may have been useful for exploring ambiguities between comorbidities defined by Charlson and Elixhauser and new ones identified by the ICU physicians.

Finally, administrative data does not contain any information about the severity of the comorbidity or duration of comorbidity. This would be useful to explain the relationship between comorbidity and death since it is expected that more severe comorbidities would increase the risk of death for a person. Other studies have relied on assigning severity scores for comorbidities based on a certain comorbidity and age group(103, 104). Information on comorbidity severity could be obtained by medical chart review, however, this information would be expensive to collect and require a lot of time. In the future, electronic medical records can be a source of this information because it may contain detailed clinical information on a patient that can be retrieved electronically for research needs.

5.3. Recommendations on using research results

This research focused on applying the Charlson index and Elixhauser as possible risk adjustment methods used in the ICU population. The results of this study will be important to health service researchers who are conducting ICU performance assessments, as well as clinicians who are designing clinical studies. It also provides physicians a better understanding of the comorbidity profile of ICU patients.

Since there are existing coding algorithms for Elixhauser and Charlson index in administrative data, the best comorbidity risk adjustment method will depend on the study population, sample size, outcome and objectives of the study. Elixhauser may be suitable to use when an ICU study sample contains a large proportion of trauma patients.

As shown, Elixhauser contains more comorbidity that is relevant to trauma patients and would produce a better model than Charlson index. When researchers have limited resources and time to derive study specific weights, then Elixhauser would be a suitable option that can still provide fairly good discrimination on its own. In contrast, the Charlson index may be a suitable risk adjustment method over Elixhauser in certain situations. For example, when researchers want to add a single summarized score as a confounder into a model, rather than a series of individually weighted comorbidities. This would be the case when the study sample is small and researchers want to avoid over-fitting the model. Researchers have recommended that there should be no more than one predictor per 18 events when predicting dichotomous outcomes(25). Charlson index can also be modeled as a time varying covariate in survival studies because the score is summarized into one signal measure. When researchers want to explore the interaction of comorbidity and other variables such as age, Charlson index may be the preferred method. Finally, this study showed that Charlson index and Elixhauser were both valid methods to risk adjust in the ICU population and the choice of risk adjustment method will depend on the study population, sample size, outcome and objectives of the study.

5.4. Future research

This work provides the groundwork for exploring comorbidity scores in an ICU study sample. There is still much work to validate these comorbidities scores across Canada in the ICU study population. ICUs vary across the country in terms of coding practices and study sample characteristics, therefore, researchers should investigate whether these scores perform the same across jurisdictions. In this study, short term mortality outcomes were initially explored (ex. 30 days and 90 days) with very little differences in outcomes. However, as the Brussel's Roundtable Report recommended, it may be useful to explore mortality outcomes greater than a year. This could provide valuable information on the role that comorbidities play for a long term survival in ICU patients. Finally, as the health care system begin to adopt electronic medical records, valuable information on clinical and physiological measures can be captured in this data source. The data may provide important information about comorbidities such as the severity and stage of the disease. This information could be combined with pharmacy data to gather information about medication dose and duration of illness to provide a better understanding of the burden of comorbidities in the ICU population. All three data sources remain potentially valuable for risk adjustment and researchers should continue to explore how each can be used to enhance risk adjustment in the ICU population.

5.5. Conclusion

In summary, this study has demonstrated the potential for Charlson and Elixhauser as comorbidity risk adjustment methods in the ICU population. Both methods proved to be adequate on their own, and could provide meaningful improvement when they were combined with other demographic data such as age and sex. In contrast to ICU severity illness scores, comorbidity scores can be easily collected through administrative database which is an inexpensive and efficient method to obtain data on large populations. The comorbidity coding algorithms can be applied uniformly across most ICU study samples without the need to derive new scoring systems or update existing ones, which is a common problem among existing ICU severity illness score. Therefore, this study provides justification that Charlson index and Elixhauser can be used to risk

adjust in the ICU population, as an alternative to APACHE II, when long term mortality outcomes are considered.

Finally, as the demand of critical care services continue to grow over the next few decades, there will be the need to evaluate health outcome disparities between different providers. As the 2002 Brussels Roundtable Report suggested, understanding the effects of ICU interventions on long term health and well being are important, therefore, ICU clinical trials of ICU therapies should examine long term follow up of survivors. This information is essential to provide a high quality of care across providers and to ensure that therapies have beneficial long term outcomes in order to further advance the field of critical care medicine.

References

1. Jacob P, Noseworthy T. National estimates of intensive care utilization and costs: Canada and the United States. Crit Care Med. 1990;18(11):1282-6.
2. Leeb K, Jokovic A, Sandhu M, Zinck G. CIHI Survey: Intensive Care in Canada. Healthcare Quarterly. 2006;9(1):32-3.
3. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: A severity of disease classification system. Crit Care Med. 1985;13(10):818-29.
4. Zimmerman J, Kramer A, Douglas S, Malila F. Acute Physiology and Chronic Health Evaluation (APACHE IV): Hospital mortality assessment for today's critically ill patients. Crit Care Med. 2006;34(5):1297-310.
5. Knaus W, Zimmerman J, Wagner D, Draper E, Lawrence D. APACHE- acute phsyiology and chronic health evaluations: A physiologically based classification system. Crit Care Med. 1983;9(8):591-7.
6. Lemeshow S, Teres D, Avrunin JS, Pastides H. Predicting the Outcome of Intensive Care Unit Patients. JSTOR. 1988;83(402):348-56.
7. Lemeshow S, Teres J, Klar J, Avrunin J, Gehlbach S, Rapoport J. Mortality probability models (MPM II) based on an international cohort of intensive care patients JAMA. 1993;270:2478-86.
8. Moreno RP, Metnitz PGH, Almeida E, et al. SAPS 3—From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. Intensive Care Med. 2005;31:1345-55.
9. Schneeweiss S, Wang P, Avorn J, Glynn R. Improved comorbidity adjustment for predicting mortality in Medicare populations. Health Serv Res. 2003;38(4):1103-19.
10. Charlson Mary E, Pompei Peter, Ales Kathy L, Ronald M. A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation. J Chron Dis. 1987;40(5):373-83.
11. Charlson ME, Szatrowski TP, Peterson J, Gold J. Validation of a combined Charlson Index. J Clin Epidemiol. 1994;47(11):1245-51.
12. Cleves MA, Sanchez N, Draheim M. Evaluation of Two Competing Methods for Calculating Charlson's Comorbidity Index When Analyzing Short-Term Mortality Using Administrative Data. J Clin Epidemiol. 1997;50(8):903-8.
13. Geissler HJ, Holzl P, Marohl S, et al. Risk stratification in heart surgery: comparison of six score systems Eur J Cardiothorac Surg 2000;17(4):400-6.
14. Kieszak S, Flanders W, Kosinski A, Shipp C, Karp H. A comparison of the Charlson comorbidity index dervied from medical record data and adminstrative billing data. J Clin Epidemiol. 1999;52(2):137-42.
15. Librero J, Peiro S, Ordinana R. Chronic Comorbidity and Outcomes of Hospital Care: Length of Stay, Mortality, and Readmission at 30 and 365 Days. J Clin Epidemiol. 1999;53(3):171-9.
16. Rochon PA, Katz JN, Morrow LA, et al. Comorbid Illness Is Associated with Survival and Length of Hospital Stay in Patients with Chronic Disability: A Prospective Comparison of Three Comorbidities Indices. Med Care. 1996;34(11):1093-101.

17. Southern DA, Quan H, Ghali WA. Comparison of the Elixhauser and Charlson/Deyo Methods of Comorbidity Measurement in Adminstrative Data. Med Care. 2004;42(4):355-60.
18. Stukenborg George, Wagner Douglas, Jr. CA. Comparison of the Performance of Two Comorbidity Measures, With and Without Information from Prior Hospitalization. Medical Care. 2001;39(7):727-39.
19. Ho K, Knuiman J, Webb S. Combining multiple comorbidities with Actuve Physiology score to predict hospital mortality of critically ill patients : a linked data cohort study. Anaesthesia 2007;62(11):1095-100.
20. Poses RM, McClish DK, Smith WR, Bekes C, Scott WE. Prediction of Survival of Critically Ill Patients by Admission Comorbidity. J Clin Epidemiol. 1996;49(7):743-7.
21. Needham DM, Scales DC, Laupacis A, Pronovost PJ. A systematic review of the Charlson comorbidity index using Canadian administrative databases: a perspective on risk adjustment in critical care research. J Crit Care. 2005;20(1):12-9.
22. Johnston JA WD, Timmons S, Welsh D, Tsevat J, Render ML. Impact of different measures of comorbid disease on predicted mortality of intensive care unit patients. Medical care. 2002;40(10):929-40.
23. Norena M, Wong H, Thompson WD, Keenan SP, Dodek PM. Adjustment of intensive care unit outcomes for severity of illness and comorbidity scores. J Crit Care. 2006;21(2):142-50.
24. Angus D, Carlet J. Surviving Intensive Care: a report from the 2002 Brussels Roundtable. Intensive Care Med. 2003;29(3):368-77.
25. Iezzoni L. Risk adjustment for measuring health care outcomes. Chicago; 2003.
26. Hennekens C, Buring J. Epidemiology in Medicine. Boston, Little: Brown and Company; 1987.
27. Wunsch H, Linde-Zwirble W, Angus D. Methods to adjust for bias and confounding in critical care health services research involving observational data. J Crit Care. 2006;21(1):1-7.
28. Apgar V. A proposal for a new method of evaluation of the newborn infant. Anesth Analg. 1953;32(4):260-7.
29. Le Gall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European / North American multicenter study JAMA. 1993;270(24):2957-63.
30. Lemeshow S, Teres D, Avrunin JS, Pastides H. Predicting the Outcome of Intensive Care Unit Patients. Journal of the American Statistical Association. 1988;83(402):348-56.
31. Knaus W, Wagner D, Draper E, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. Chest. 1991;100:1619-36.
32. Pittet D, Thiévent B, Wenzel R, Li N, Gurman G, Suter P. Importance of pre-existing co morbidities or prognosis of septicemia in critically ill patients. Intensive Care Med. 1993;19(5):265-72.
33. Iezzoni L, Ash A, Coffman GA, Moskowitz MA. Predicting In-Hospital Mortality: A Comparison of Severity Measurement Approaches. Med Care. 1992;30(4):347-59.

34. Costa Joel I, Amaral Gomes do, et al. Severity and prognosis in intensive care: prospective application of the Apache II Index. Sao Paulo Medical Journal. 1999;1999(117):205-14.
35. Hsu CW WS, Chiang HT, Lin CH, Kung MH, Lin SL. Comparison of the APACHE II and APACHE III scoring systems in patients with respiratory failure in a medical intensive care unit. J Formos Med Assoc. 2001;100(7):437-42.
36. Chatzicostas C RM, Notas G, Vlachonikolis IG, Samonakis D, Romanos J, Vardas E, Kouroumalis EA. A comparison of Child-Pugh, APACHE II and APACHE III scoring systems in predicting hospital mortality of patients with liver cirrhosis. BMC Gastroenterol. 2003;8(3):7.
37. Wong D, Crofts S, Gomez M, McGuire G, Byrick R. Evaluation of predictive ability of APACHE II system and hospital outcome in Canadian intensive care unit patients. . Crit Care Med. 1995;23(7):1177-83.
38. Zimmerman JE, Wagner DP, Draper EA, Wright L, Alzola C, Knaus W. Evaluation of acute physiology and chronic health evaluation III predictions of hospital mortality in an independent database. Crit Care Med. 1998;26(8):1317-26.
39. Harrison D, Parry G, Carpenter J, Short A, Rowan K. A new risk prediction model for critical care: The Intensive Care National Audit & Research Centre (ICNARC) model. Crit Care Med. 2007;35(4):1091-7.
40. Moreno R, Jordan B, Metnitz P. Intensive Care Medicine. The Changing Prognostic Determinants in the Critically Ill Patient: Springer Berlin Heidelberg; 2007. p. 899-907.
41. Harrison D, Brady A, Parry G, Carpenter J, Rowan K. Recalibration of risk prediction models in a large multicenter cohort of admissions to adult, general critical care units in the United Kingdom. Crit Care Med. 2006;34:1378-88.
42. Lemeshow S, Teres D, Avrunin JS, Gage R. Refining intensive care until outcome prediction by using changing probabilities of mortality. Crit Care Med. 1988;16(5):471-7.
43. Wunsch H, Brady A, Rowan K. Impact of Exclusion Criteria on Case Mix, Outcome, and Length of in Critical Care. J Crit Care. 2004;19(2):67-74.
44. Kaplan M, Feinstein A. The importance of classifying initial co-morbidity in evaluatin the outcome of diabetes mellitus. J Chron Dis. 1974;28(7-8):387-404.
45. Ghali William, Hall RE, Rosen AK, Ash A, Moskowitz MA. Searching for an Improved Clinical Comorbidity Index for Use with ICD-9-CM Administrative Data. J Clin Epidemiol. 1996;49(3):273-8.
46. Romano PS, Roos L, Jollis JG. Further evidence concerning the use of a clinical comorbidity index with ICD-9-CM administrative data. J Clin Epidemiol. 1993;46(10):1085-90.
47. Elixhauser A, Steiner C, Robert H, Coffey RM. Comorbidity Measure for Use with Adminstrative Data. Med Care. 1998;36(1):8-27.
48. Grunau GL, Sheps S, Goldner EM, Ratner PA. Specific comorbidity risk adjustment was a better predictor of 5-year acute myocardial infarction mortality than general methods. J Clin Epidemiol. 2006;59(3):274-80.
49. Stukenborg George WD, Connors Alfred Jr. . Comparison of the Performance of Two Comorbidity Measures, with and without
Information from Prior Hospitalizations. Medical Care. 2001;39(7):727-39.

50. Quan H, Sundararajan V, Halfon P, et al. Coding Alogrithms for Defining Comorbidities in ICD-9 CM and ICD-10 Adminstrative Data. Med Care. 2005;43(11):1130-90.
51. von Korff M, Wagner EH, K.. S. A chronic disease score from automated pharmacy data. J Clin Epidemiol 1992;45:197-203.
52. D'Hoore W, Bouckaert A, Tilquin C. Practical Considerations on the Use of the Charlson Comorbidity Index with Administrative Data Bases. J Clin Epidemiol. 1996;49(12):1429-33.
53. Deyo RA, Cherkin DC, Ciol MA. Adapting a Clinical Comorbidity Index for Use with ICD-9CM Administrative Databases. J Clin Epidemiol. 1992;45(6):613-9.
54. Romano PS, Roos L, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives J Clin Epidemiol. 1993;46(10):1075-9.
55. de Groot V, Beckerman H, Lankhorst G, Bouter L. How to measure comorbidity: a critical review of available methods. J Clin Epidemiol. 2003;56(3):221-9.
56. Gijsen R, Hoeymans N, Schellevis F, Ruwaard D, Satariano W, Van G. Causes and consequences of comorbidity: A review. J Clin Epidemiol. 2001;54(7):661-74.
57. Roos N, Wennberg J, Malenka D, et al. Mortality and reoperation after open and transurethral resection of the prostate for benign prostatic hyperplasia. New England J Med. 1989;321(16):1120-24.
58. Concato J, Horwitz R, Feinstein A, Elmore J, Schiff S. Problems of comorbidity in mortality after prostatectomy. JAMA. 1992;267:1077-82.
59. Humphries K, Rankin J, Carere R, Buller C, Kiely F, Spinelli J. Co-morbidity data in outcomes research: Are clinical data derived from administrative databases a reliable alternative to chart review? J Clin Epidemiol. 2000;53(4):343-9.
60. Iezzoni et al. Comorbidities, complications, and coding bias. JAMA. 1992;267(16):2197.
61. Schneeweiss S, Macture M. Use of comorbidity scores for control of confounding in studies using administrative databases. International Journal of Epidemiology. 2000;29:291-8.
62. Tu JV, Mazer CD, Levinton C, Armstrong PW, CD. N. A predictive index for length of stay in the intensive care unit following cardiac surgery. CMAJ. 1994;151:177-85.
63. Solomon D, Ganz D, Avorn J, Glynn R, Knight E, Gibson D, et al. Which patients with unstable angina or non-Q-wave myocardial infarction should have immediate cardiac catheterization? A clinical decision rule for predicting who will fail medical therapy. J Clin Epidemiol. 2002;55(2):121-8.
64. Harrell FE. Regression Coefficients and Scoring Rules. J Clin Epidemiol. 1996;49(7):819.
65. Moons K. Should scoring rules be based on odds ratios or regression coefficients. J Clin Epidemiol. 2002;55(10):1054-5.
66. Norena M, Wong H, Thompson WD, Keenan SP, Dodek PM. Adjustment of intensive care unit outcomes for severity of illness and comorbidity scores. J Crit Care. 2006;21(2):142-50.

67. Yan Yan B-DE, Radford Martha, Nilasena David, and Gage Brian. Comorbidity Indices to Predict Mortality From Medicare DataResults From the National Registry of Atrial Fibrillation. Medical Care. 2005;43:1073-7.
68. Rubenfield G, Angus D, Pinsky R, Curtis R, Connors A, Bernard G. Outcomes Research in Critical Care: Results of the American Thoracic Society Critical Care Assembly Workshop on Outcomes Research. Am J Respir Crit Care Med. 1999;160(1):358-67.
69. Curtis J, Martin D, Martin T. Patient-assessed health outcomes in chronic lung disease: what are they, how do they help us, and where do we go from here? Am J Respir Crit Care Med. 1997;156:1032-9.
70. Rubenfeld G, et al. Surrogate measures of patient-centered outcomes. In: Angus D, Carlet J, editors.; 2003; Brussels 2003.
71. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnosis: ICD code accuracy. Health serv res. 2005;40(5):1620–39.
72. Quan H, Parsons GA, Ghali WA. Validity of Information on Comorbidity Derived From ICD-9CCM Adminstrative Data. Med Care. 2002;40(8):675-85.
73. J.-L. Vincent RM, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, C. K. Reinhart, P. M. Suter, L. G. Thijs The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure Intensive Care Medicine. 1998 26(11):1793-800.
74. Jollis J, Ancukiewicz M, De Long E, et al. Discordance of databases designed for claims payment versus clinical information systems. Implications for outcomes research. Ann Intern Med. 1993;119(8):844-50.
75. Raiford D, Pérez GS, García RL. Positive predictive value of ICD-9 codes in the identification of cases of complicated peptic ulcer disease in the Saskatchewan hospital automated database. Epidemiology. 1996;7(1):101-4.
76. Quan H, Bing L, Duncan S, et al,. Assessing Validity of ICD-9-CM and ICD-10 Administrative Data in Recording Clinical Conditions in a Unique Dually Coded Database. Health serv res. 2008;43(4):1424-41.
77. Roos L, Stranc L, James R, Li J. Complications, comorbidities and mortality: improving classification and prediction. Health serv res. 1997;32(2):229-38.
78. Malenka DJ, Roos N, Fisher E, et al. Further study of the increased mortality following transurethral prostatectomy; a chart based analysis. J Urol. 1990;144(2):224-28.
79. Quan H, Parsons G, Ghali W. Assessing the accuracy of diagnosis-type indicators for flagging complications in adminstrative data. J Clin Epidemiol. 2004 2004;57(4):366-72.
80. Iezzoni LI, Foley SM, Daley J, Hughes J, Fisher ES, T H. Comorbidities, complications, and coding bias. JAMA. 1992;267(16):2197-203.
81. Li B, Quan H, Fong Aa, Lu M. BMC Health Services Research. BioMed Central. 2006;6(48):1-10.
82. Doig CJ, Zygun DA, Fick G, et al. . Study of clinical course of organ dysfunction in intensive care. Crit Care Med. 2004;32(2):384-90.
83. Harrell FE. Regression Modeling Strategies. 1 ed. Nashville: Springer; 2001.
84. Sackett D, Gent M. Controversy in counting and attributing events in clinical trials. N Engl J Med 1979;301(26):1410-2.

85. Hosmer D and Lemeshow S NY. Applied logistic regression (1st ed): John Wiley & Sons; 1989.
86. Kuzniewicz M, Draper D, Escobar G. Incorporation of physiological trend and interaction effects in neonatal severity of illness scores: an experiment using a variant of the Richardson score Intensive Care Med. 2007;33(9):1501-2.
87. Linstone H, Turoff M. The Delphi Method: Techniques and Applications. 2002 [cited 2008 Jan 1 2008]; Available from: http://www.is.njit.edu/pubs/delphibook/delphibook.pdf
88. Sackman H. A Delphi Critique. Lexington, Massachusetts: Lexington Books; 1975.
89. Beers M, Ouslander J, Rollinger I, Reuben D, Brooks J, Beck J. Explicit criteria for determining inappropriate use in nursing home residents. Arch Intern Med. 1991;151(9):1825-32.
90. Bellamy N, Anastassiades T, Buchanan W, David L, Lee P, McCain G, et al. Rheumatoid arthritis anti-rheumatic trials III. Setting the dela for clinical trials of anti-rheumatic drugs-results of a consensus devlelopment (Delphi) excercise. J Rheumatol. 1991;18:1908-15.
91. Crotty M. The emerging role of the British nurse teacher in Project 2000 programmes: a Delphi suvey. J Adv Nurs. 1993;18(1):150-7.
92. Fraser C, Smith Q, Luchi R. Geriatric fellows' perceptions of the quality of their research training. Avad Med. 1992;67(10):696-8.
93. Service. USDoAUF. Delphi Process Description - Summary. United States Department of Agriculture (USDA) Forest Service Website; 2004.
94. Pill J. The Delphi method: substance, contesxt, a critique and an annoted bibliography. Socio-Economic Planning Science. 1971;5:57-71.
95. Rowe G, Wright G, Bolger F. Delphi: a re-evaluation of research and theory. Technological Forecasting and Social Change. 1991;39:235-51.
96. D'Hoore William, Bouckaert Andre, Charles T. Practical Considerations on the Use of the Charlson Comorbidity Index with Adminstrative Data Bases. J Clin Epidemiol. 1996;49(12):1429-33.
97. Morris J, MacKenzie E, Edelstein S. The effect of preexisting conditions on mortality in trauma patients. JAMA. 1990;263(14).
98. Bochicchio G, Manjari J, Bochicchio K, Shih D, Meyer W, Scalea T. Incidence and Impact of Risk Factors in Critically Ill Trauma Patients World Journal of Surgery. 2005;30(1):114-8.
99. Nuttall M, Meulen J, Emberton M. Charlson scores based on ICD-10 adminstrative data were valid in assessing comorbidity in patients undergoing urological cancer surgery. Journal of Clinical Epidemiology. 2006;59(3):265-73.
100. Jencks S, Williams D, Kay T. The role of length of stay and comorbidities. JAMA. 1988;260:2240.
101. Jencks S, Daley J, Draper D, Thomas N, Lenhart G, Walker J. Interpreting hospital mortality data.. The role of clinical risk adjustment. JAMA. 1988;260(24):3611.
102. Romano PS, Mark DH. Bias in the Coding of Hospital Discharge Data and Its Implications for Quality Assessment. Med Care. 1994;32(1):81-90.

103. Cujec B, Jin Y, Quan H, Johnson D. The province of Alberta, Canada avoids the hospitalization epidemic for congestive heart failure patients. International Journal of Cardiology. 2004;96(2):203-10.
104. Polancyzk C, Rohde L, Philbin E, DiSalvo T. A new casemix adjustment index for hospital mortality among patients with congestive heart failure. Medical Care. 1998;36(10):1489-99.

APPENDICES

Appendix 1: APACHE II calculation formula (3)

Acute physiology and chronic health evaluation (APACHE II) scoring system

Physiology points	4	3	2	1	0	1	2	3	4
Rectal temperature (°C)	≥41.0	39.0-40.9		38.5-38.9	36.0-38.4	34.0-35.9	32.0-33.9	30.0-31.9	≤29.9
Mean blood pressure (mm Hg)	≥160	130-159	110-129		70-109		50-69		≤49
Heart rate (beats/min)	≥180	140-179	110-139		70-109		55-69	40-54	≤39
Respiratory rate (breaths/min)	≥50	35-49		25-34	12-24	10-11	6-9		≤5
Oxygenation (kPa)*:									
Fio_2 ≥50% A-aDo_2	66.5	46.6-66.4	26.6-46.4		<26.6				
Fio_2 <50% Pao_2					>9.3	8.1-9.3		7.3-8.0	<7.3
Arterial pH	≥7.70	7.60-7.59		7.50-7.59	7.33-7.49		7.25-7.32	7.15-7.24	<7.15
Serum sodium (mmol/l)	≥180	160-179	155-159	150-154	130-149		120-129	111-119	≤110
Serum potassium (mmol/l)	≥7.0	6.0-6.9		5.5-5.9	3.5-5.4	3.0-3.4	2.5-2.9		<2.5
Serum creatinine (μmol/l)	≥300	171-299		121-170	50-120		<50		
Packed cell volume (%)	≥60		50-59.9	46-49.9	30-45.9		20-29.9		<20
White blood cell count ($\times 10^9$/l)	≥40		20-39.9	15-19.9	3-14.9		1-2.9		<1

*If fraction of inspired oxygen (Fio_2) is ≥50% the alveolar-arterial gradient (A—a) is assigned points. If fraction of inspired oxygen is <50% partial pressure of oxygen is assigned points.

Other points

Glasgow coma scale: Score is subtracted from 15 to obtain points.

Age <45 = 0 points, 45-54 = 2, 55-64 = 3, 65-75 = 5, ≥75 = 6.

Chronic health points (must be present before hospital admission): chronic liver disease with hypertension or previous hepatic failure, encephalopathy, or coma; chronic heart failure (New York Heart Association grade 4); chronic respiratory disease with severe exercise limitation, secondary polycythaemia, or pulmonary hypertension; dialysis dependent renal disease; immunosuppression—for example, radiation, chemotherapy, recent or long term high dose steroid therapy, leukaemia, AIDS. 5 points for emergency surgery or non-surgical patient, 2 points for elective surgical patient.

- Serum creatinine points are doubled if there is acute renal failure present.

Estimation of probability of death in hospital by applying APACHE II for 71 year old man admitted to intensive care from the hospital's accident and emergency department with (a) abdominal aortic aneurysm and (b) asthma attack

Criteria	Value	Points
Primary reason for admission	(a) Abdominal aortic aneurysm (b) Asthma attack	
Age	71 years	5
History	None	0
Physiology:		
Temperature	38.4°C	1
Mean blood pressure	112 mm Hg	2
Heart rate	136 beats/min	2
Respiratory rate	28 breaths/min	1
Oxygenation:		0
Fraction of inspired oxygen	0.4	
Partial pressure of oxygen	21.2 kPa	
Partial pressure of carbon dioxide	4.4 kPa	
pH	7.09	4
Serum sodium	150 mmol/l	1
Serum potassium	5.5 mmol/l	1
Serum creatinine	145 µmol/l	2
Packed cell volume	40%	0
White blood cell count	20×10^9/l	2
Glasgow coma score:		
Eyes	Opening spontaneous	1
Motor	Obeys verbal command	
Verbal	Disoriented and converses	
Total		22

(a) APACHE II probability of hospital death: Abdominal aortic aneurysm (0.731) + APACHE II score $(22 \times 0.146 = 3.212) - 3.517 = 0.426$

$\frac{e^{0.426}}{1+e^{0.426}} = 0.6049182 = 60.5\%$ probability of hospital death

(b) APACHE II probability of hospital death: Asthma attack in known asthmatic (−2.108) + APACHE II score $(22 \times 0.146 = 3.212) - 3.517 = -2.413$

$\frac{e^{-2.413}}{1+e^{-2.413}} = 0.08211867 = 8.2\%$ probability of death

Appendix 2: Charlson index score calculation (10)

Weight	Clinical condition
1	Myocardial infarct Congestive cardiac insufficiency Peripheral vascular disease Dementia Cerebrovascular disease Chronic pulmonary disease Conjunctive tissue disease Slight diabetes, without complications Ulcers Chronic diseases of the liver or cirrhosis
2	Hemiplegia Moderate or severe kidney disease Diabetes with complications Tumors Leukemia Lymphoma
3	Moderate or severe liver disease
6	Malignant tumor, metastasis Aids

Figure 1 - Charlson comorbidity index – weighting of the clinical conditions present among secondary diagnoses.

Appendix 3: Pilot Study Results

Comorbidity	OR	P>|z|*	95% CI		ICU Weight †	Charlson Weight	% of person with COB ‡	Rank	% hospital MT attributable to a COB	% In hospital MT **
Chronic Pulmonary Disease	0.84	0.12	0.67	1.04		1	16.38	1	27.46	4.5
Congestive Heart Failure	1.06	0.63	0.84	1.33		1	12.92	2	35.86	4.63
Diabetes without Complications	0.77	0.036	0.61	0.98		1	12.44	3	29.15	3.63
Myocardial Infarction	1.47	0.003	1.15	1.89	1	1	10.03	4	40.94	4.1
Renal Disease	1.37	0.036	1.02	1.84	1	2	7.41	5	40	2.96
Cerebrovascular Disease	1.75	0	1.28	2.38	2	1	6.14	6	40.94	2.51
Cancer	1.46	0.02	1.06	2.00	1	2	5.64	7	40.38	2.276
Periphral Vascular Disease	1.3	0.109	0.94	1.80	1	1	5.61	8	41.51	2.33
Paraplegia and Hemiplegia	1.27	0.209	0.88	1.83	1	2	4.71	9	32.02	1.51
Metastatic Carcinoma	1.5	0.022	1.06	2.13	2	6	5.64	10	62.96	1.8
Mild Liver Disease	2.5	0	1.69	3.70	2	1	4.08	11	52.6	2.14
Diabetes with complications	0.63	0.031	0.42	0.96		2	4.08	12	29.22	1.19
Moderate or Severe Liver diseases	3.85	0	2.41	6.14	4	3	2.86	13	62.96	1.8
Dementia	0.7	0.152	0.43	1.14		1	2.38	14	33.33	0.79
Peptic Ulcer Disease	0.87	0.61	0.51	1.48		1	2.17	15	31.7	0.69
Connective Tissue Rheumatic Diseases	0.91	0.751	0.52	1.60		1	1.85	16	31.43	0.58
HIV	1.26	0.85	0.11	14.2	1	6	0.11	17	25	0.026

* P>|z| probability that an observed value is greater than 0.05

† ICU Weights calculated from logistic regression model

‡ COB refers to comorbidities

** MT refers to mortality, Ranks refer to most frequent disease, with “1” representing the disease with the highest proportion of cases in the ICU

Appendix 3: Pilot Study Results

Model	Log Likelihood	Degrees of Freedom	P value for LR	Likelihood Ratio Test	C stat	AIC (LR- 2xdf)	Mean Shrinkage
Model A (Baseline model) age sex APS *	-1882.0096				0.743		0.97
Model B age sex APS CHP†	-1843.9226	1	<0.0001	76.174	0.757	74.174	0.97
Model C age sex APS c.cob‡	-1814.6895	17	<0.0001	134.640	0.768	100.64	0.94
Model D (D'Hoore) Age sex APS Charlson index	-1861.5636	1	<0.0001	40.8992	0.752	38.8992	0.98
Model E Age sex APS Charlson index (excluded scores 6 and above)	-1701.1368	1	<0.0001	11.916	0.757	9.916	1.01
Model F Apache II + Charlson Index	-1652.6734				0.813		0.99
Model G Apache II (with diagnosis)	-1670.112				0.808		1.00

* APS is the acute physiology status derived from APACHE II score by subtracting the age score

† CHP is the Chronic Health Points from the Apache II score.

‡ Charlson co-morbidities as individual dummy variables

Appendix 4: Coding documentation for data cleaning

Cleaned ICU data (Sept 24 2007)

1. (8524) ICU admission records (excluded CVICU and <18)
2. In sas, I ran a command to get the first admission based on PHN (7430)
3. I got rid of people who were non CHR and Alberta based on Unit Discharge to variable and postal code. (6793)
4. Looked for chartnum that appeared twice and got rid of those people if they were the same (6395)
5. Cleaned people based on same lastname, phn (6329)
6. Cleaned people with same lastname , bdate (6325)
7. CHRINDEX11, 212 missing phns, and 84 had invalid phn

Link to IP dataset to get all those missing phns and invalid phn cleaned up (Oct 1)

8. Saved 140/212 phns , got rid of 87 phns that were pissing and not found in ip dataset (missing record in ip or blank phn) CHRINDEX11_c (6173)

Link to Population Registry to identify CHR residents by phn (October 5)

9. CHRINDEX12a (6048) – lost 125 records (non CHR residents)
10. Looked at active field to do some estimate checks for mortality and logic checks

Link to IP dataset to get hospital record that corresponds with ICU stay, also need to identify cdr key to link to get diagnosis (Oct 10)

11. Widen admit and discharge dates for IP1 (414793) to include icu dates
12. Link on phn, admit , discharge date→ CHRINDEX 14 (6157)
13. 91 of these icu records had no hospital record
14. 103 had two hospital records for ICU date , TESTCHRINDEX14
15. From TESTCHRINDEX14, I chose the hospital records, that covered the ICU period(made most sense)
16. CHRINDEX15b (5851) + singlechrindex14c (106)→ CHRINDEX16 (5954) unique records with one hospital record for each icu record
17. From the badones91, I manally looked for the chartnum by using MPI file
18. CHRINDEX19 (6040,123) add the diagnosis type codes → CHRINDEX20(6040,174)
19. chrindex20 add the diagnosis → CHRINDEX21 (6040,224)
20. get rid of those who were before april 1 2000 icu admits→ CHRINDEX23 (6017,224)

Link to viral stats (2000-2004 Bing) by PHN to find out death dates

21. TESTC(5160,227) could link okay. They will link to VS if they have a death record (death date will be available). I can assume that the people who did not link to VS did not die yet
22. TESTA(857,227) could not link by phn to VS.
23. TESTA did some manual checks and tried to link to VS again, and population registry, there were no records like these in VS and when they linked to population registry, I looked at the active fields, and they were blank for the study period. This means these people were not CHR residents during that time. This is

also confirmed by looking at MPI resident location, not CHR. These people were dropped from the study.

24. From testc, 348 people are from the cohort Jan2004 to April 2004 admit. If they had died a year later, they would not be found in Bing's vital stats record.
25. Jan2004apr1a link to C_vital2005, only 5 linked by phn (death date confirmed) manually searched through records to see if same lastname and firstname died in vital stats because c_vital2005 was not clean for phn, I found only 2 records that died based on names, combined these 7 deathrecords with the complete dataset
26. CHRINDEX26 (5160,227)

Mortality Rate

Variable	Source	MT
Dead/alive	Icu	19.11%
Exit_alive_code	Hospital	1432/5159=27.76%
Hosp_outcome	ICU	30.41%

Appendix 5: Delphi Cover Letter

September X 2007

Dear Dr X,

I am a Master's student working with Dr Christopher Doig at the University of Calgary, and am conducting a study to compare three risk adjustment methods for predicting mortality in ICU patients. You have been selected as one of the intensivists to join our expert panel for this study. With your help, and using a three step process, we will develop a list of clinical comorbidities that are the most relevant for predicting 90 days, 180 days, and one year mortality in ICU patients.

Two indices are commonly used in research to adjust for comorbidity, the Charlson and Elixhauser methods. These indices however were developed in general medical patients and not the ICU population, thus they may be missing important comorbidities that are clinically known to predict mortality in an ICU population. Therefore we are asking for your input to develop a list of clinical conditions which are most relevant for predicting mortality in the ICU population. This process will take place in three rounds, each of which will take approximately 5-10 minutes to complete.

Round 1: You will be provided with a list of Charlson and Elixhauser comorbidities, and asked to identify additional comorbidities that can increase the risk of death for ICU patients. **Comorbidities are defined as a clinical condition that exists before a patient's admission to the hospital, is not related to the principal reason for the hospitalization, and is likely to be a significant factor influencing mortality.**

Round 2: We will give you a complete list of comorbidities derived from the entire panel and you can check off which comorbidities are relevant for predicting mortality.

Round 3: Based on the responses from the second round, we will provide the group response, and you will rank the relevancy for the comorbidity to predict mortality.

Detailed instructions will be provided at each round.

If you have any questions about the process, you can send an email to susan.quach@gmail.com. **Please email the completed forms within 10 days for each round. You will receive a reminder email close to the deadline.** At the end of the thesis project, you will receive a summary report of the results obtained from this process. Thank you for your assistance in the completion of this project.

Sincerely,

Susan Quach

Master's Student

Department of Community Health Sciences

Comorbidity
Cardiac arrhythmias
Chronic pulmonary disease
Fluid and electrolyte disorders
Congestive heart failure
Diabetes ,uncomplicated
Diabetes, complicated
Alcohol abuse
Other neurological disorders
Coagulopathy
Mild Liver Disease
Moderate or severe liver disease
Depression
Peripheral vascular disorders
Hypertension (complicated)
Hypertension (uncomplicated)
Hypothyroidism
Metastatic solid tumour
Solid tumor without metastasis
Lymphoma
Leukemia
Pulmonary circulation disorders
Valvular disease
Drug abuse
Obesity
Rheumatoid arthritis
Deficiency anemias
Blood loss anemias
Weight loss
Psychoses
Peptic ulcer disease excluding bleeding
AIDS
Myocardial infarct
Cerebrovascular disease
Dementia
Connective tissue disease
Ulcer disease
Hemiplegia/ paraplegia
Paralysis
Moderate to severe renal disease
Renal failure

Appendix 6: Delphi Procedure

Delphi Step I

Step 1: The following is a list of Charlson and Elixhauser comorbidities that have been shown to predict mortality in non-ICU populations. Based on your clinical experience and knowledge, please identify additional comorbidities on this list that are useful to predict mortality in an ICU population.

Comorbidities are defined as a clinical condition that exists before a patient's admission to the hospital, is not related to the principal reason for the hospitalization, and is likely to be a significant factor influencing mortality.

Additional comorbidities:

Please complete this form within the next 5 days and email/fax it to Susan Quach (403)270-4329 (fax)

Susan.quach@gmail.com

Thank you!

Delphi Step II

Round 2 Instructions: Please check off which comorbidities are relevant for predicting mortality in ICU population, for a given endpoint.

If you do not think that the comorbidity predicts mortality in the ICU population for any of the three endpoints, leave these fields blank.

Comorbidities that predict mortality in ICU patients	**90 days**	**180 days**	**365 days**
Cardiac arrhythmias			
Chronic pulmonary disease			7
Fluid and electrolyte disease			
Congestive heart failure			
Diabetes ,uncomplicated			
Diabetes, complicated			
Alcohol abuse			
Other neurological disorders			
Coagulopathy			
Mild Liver Disease			
Moderate or severe liver disease			
Depression			
Peripheral vascular disorders			
Hypertension (complicated)			
Hypertension (uncomplicated)			
Hypothyroidism			
Metastatic solid tumour			
Solid tumor without metastasis			
Lymphoma			
Leukemia			
Pulmonary circulation disorders			
Valvular disease			
Drug abuse			
Obesity			
Rheumatoid arthritis			
Deficiency anemias			
Blood loss anemias			
Weight loss			
Psychoses			
Peptic ulcer disease excluding bleeding			
AIDS			
Myocardial infarct			
Cerebrovascular disease			
Dementia			
Connective tissue disease			
Comorbidities	**90 days**	**180 days**	**365 days**
Ulcer disease			
Hemiplegia/ paraplegia			
Paralysis			
Moderate to severe renal disease			
Renal failure			
Pulmonary fibrosis			
Severe Asthma			
Pulmonary hypertension			
Malnutrition			
Chronic /Active Viral hepatitis (B or C)			
Blindness			
Hearing Loss			

Delphi Step III

Thank you for your participation to determine comorbidities that predict mortality in the ICU. This is the third, and final, round of responses.

Description: Listed below are the top 30 comorbidities ranked, determined by the group response from the last Delphi questionnaire. The group response is the % of doctors that agreed that a specific comorbidity was relevant for predicting mortality in ICU patients at 90 and 365 days.

Instructions: Please rank each of the comorbidities below with a score from 1 to 10, with **10 being the most relevant for predicting mortality** for that given endpoint.
Example: If you think that HIV is one of the strongest predictors of 365-day mortality for ICU patients, you would rank this as a 10. You can apply the same rank for more than one comorbidity, but please refrain from using decimal places for ranking. There are two pages.

Comorbidities that predict mortality in ICU patients	**90 day mortality Rank**	**Group Response (90 day) (% said yes)**	**365 day mortality Rank**	**Group Response (365 days)**
Cardiac arrhythmias				
Chronic pulmonary disease				
Fluid and electrolyte disease				
Congestive heart failure				
Diabetes ,uncomplicated				
Diabetes, complicated				
Alcohol abuse				
Other neurological disorders				
Coagulopathy				
Moderate or severe liver disease				
Peripheral vascular disorders				
Hypertension (complicated)				
Metastatic solid tumour				
Solid tumor without metastasis				
Lymphoma				
Leukemia				
Pulmonary circulation disorders				
Valvular disease				
Drug abuse				
Obesity				
Weight loss				
AIDS				
Myocardial infarct				

Cerebrovascular disease				
Dementia				
Comorbidities				
Paralysis				
Moderate to severe renal disease				
Renal failure				
Pulmonary fibrosis				
Pulmonary hypertension				
Malnutrition				
Chronic /Active Viral hepatitis (B or C)				
Neuromuscular disorders				
Brain injury				
Chronic decubitus ulcer (pressure sores)				
Cystic firbosis				
Immunodeficiency				
Multiple Myeloma				

Appendix 7: Ethics Approval Letter

FACULTY OF MEDICINE | UNIVERSITY OF CALGARY

FILE COPY

2007-09-28

OFFICE OF MEDICAL BIOETHIC
Room 93, Heritage Medical Research Bld
3330 Hospital Drive N
Calgary, AB, Canada T2N 4N
Telephone: (403) 220-799
Fax: (403) 283-852
Email: ombe@ucalgary.c

Dr. Christopher J. Doig
Division of Critical Care
Room EG23, Foothills Hospital
Calgary, Alberta

Dear Dr. Doig:

RE: A Comparison Between Three Risk Adjustment Methods for Predicting One Year Mortality in Critically Ill Adult Patients

Ethics ID: E-21159

Student: Susan Quach

The above-named research, including the Thesis Proposal (Approval of Proposal, July 10, 2007), Protocol has been granted ethical approval by the Conjoint Health Research Ethics Board of the Faculties of Medicine, Nursing and Kinesiology, University of Calgary, and the Affiliated Teaching Institutions. The Board conforms to the Tri-Council Guidelines, ICH Guidelines and amendments to regulations of the Food and Drugs Act re clinical trials, including membership and requirements for a quorum.

Dr. Christopher Doig, one of the investigators for this study, is a member of the CHREB but did not participate in the review, was not present during discussion and did not vote on this protocol.

Please note that this approval is subject to the following conditions:

(1) consent for access to personal identified health information in chart review is not required on grounds considered under Section 50 of the Health Information Act;

(2) a copy of the informed consent form must have been given to each research subject, if required for this study;

(3) a Progress Report must be submitted by **September 28, 2008,** containing the following information:

i) the number of subjects recruited;
ii) a description of any protocol modification;
iii) any unusual and/or severe complications, adverse events or unanticipated problems involving risks to subjects or others, withdrawal of subjects from the research, or complaints about the research;
iv) a summary of any recent literature, finding, or other relevant information, especially information about risks associated with the research;
v) a copy of the current informed consent form;
vi) the expected date of termination of this project.

4) a Final Report must be submitted at the termination of the project.

Please accept the Board's best wishes for success in your research.

Yours sincerely,

Glenys Gjellosseth, BA(Hons), LLB, PhD
Chair, Conjoint Health Research Ethics Board

GG/emcg
c.c. Adult Research Committee Dr. P.J.E. Bolteau (information) Research Services Office of Information & Privacy Commissioner

Appendix 8: Charlson index Model N1-N5 based on ICU survivors

Models	LL	LR test	P value	C Stat	AIC
Model N1: age + sex	-1972.82			0.72	
Model N2 Model N1+ HP	-1825.57	294.51	<0.0001	0.77	292.51
Model N3 Model N1 + Charlson Index	-1841.36	262.93	<0.0001	0.77	260.93
Model N4: Model N1 + CI2	-1835.40	274.84	<0.0001	0.78	272.84
Model N5: Model N1 + C cob	-1779.06	387.54	<0.0001	0.79	353.54

Note: Charlson score models excluding patients that died in the ICU to predict 365 day mortality (only ICU survivors)

Appendix 9: C stat matrix of all models

Model	in-hospital	90 days	180 days	365 days
Sex	0.5	0.5	0.51	0.51
Age	0.67	0.68	0.68	0.69
APS	0.69	0.65	0.65	0.63
CHP	0.6	0.61	0.62	0.63
MPM	0.79	0.78	0.78	0.78
MPM sex	0.79	0.78	0.78	0.78
age sex	0.67	0.68	0.68	0.69
age sex APS	0.75	0.74	0.74	0.73
age sex APS CHP	0.76	0.76	0.77	0.77
age APS CHP	0.76	0.76	0.77	0.77
CI3 (exact weights)				0.73
CI1 (orignal)	0.66	0.68	0.69	0.7
CI*age				0.72
Cirw (5+rounded new weights)				0.72
C1	0.7	0.72	0.72	0.73
CI1 MPM	0.8	0.8	0.8	0.8
E1	0.73	0.73	0.73	0.73
E1 MPM	0.82	0.82	0.82	0.82
age sex APS CI1				0.77
age sex APS C1				0.79
age sex APS Cirw				0.78
age sex CHP		0.71		0.77
age sex CI1		0.71		0.74
age sex CI CI*age				0.74
age sex C1		0.75		0.77
age sex Cirw				0.75
age sex E1		0.77	0.77	0.78
age sex APS E1		0.8	0.8	0.8

Appendix 10: Elixhauser multivariate models based on ICU survivors

Table X. Elixhauser models to predict 365 day mortality with demographic variables

Models	**LL**	**LR test**	**P value**	**C Stat**	**AIC**	**Mean Shrinkage**	**Population**
Model A1: Age sex	-1972.8			0.72			ICU alive
Model B1 : Model A1 + E. cob	-1742.0	461.7	<0.0001	0.81	401.7		ICU alive