1-1-2013

Improving the Diagnostic Accuracy of Failure-to-Rescue

Amelia Marie Joseph
University of South Carolina - Columbia

IMPROVING THE DIAGNOSTIC ACCURACY OF FAILURE TO RESCUE

by

Amelia M. Joseph

Bachelor of Science
Southern Connecticut State University, 1978

Master of Business Administration
The Citadel, 1993

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Nursing Science

College of Nursing

University of South Carolina

2013

Accepted by:

Rita Snyder, Major Professor

Swann Adams, Committee Member

Kevin Bennett, Committee Member

Sue Heiney, Committee Member

Lacy Ford, Vice Provost and Dean of Graduate Studies

DEDICATION

This work is dedicated to nurses everywhere who serve as a 24-hour human surveillance system for hospitalized patients. Through their scientific knowledge, competent practice, and caring compassion untoward events are recognized and treated thus rescuing patients from failures in care.

ACKNOWLEDGEMENTS

I would like to thank the many people who contributed towards my successful completion of this program of study. First, to the faculty and staff at the College of Nursing who always believed in me and never stopped encouraging me to stay the course. Particularly, I would like to thank Dr. Rita Snyder for her guidance and support but especially for her caring and compassion. She is the epitome of a professional nurse. I would like to thank the other members of my committee for all of their wisdom and advice.

I would like to thank my cheerleaders at the Veterans Medical Centers. Dr. Sue Haddock who started me on this journey and Dr. Charlene Pope who kept me on the journey to completion. Thanks also to my many friends and supporters, too numerous to list.

A special thanks to my family. My parents, Joe and Anne Ricciardi, who instilled in me a lifelong love of learning. I wish you could have been here today. Certainly thanks to my husband who kept me from quitting my first statistics course and my children who ate many take-out meals. Last, but certainly not least, my faithful companions, Rocky and Riley who never left my side and spent many late nights at my feet while I typed and typed. We can go to the dog park now.

ABSTRACT

The ability to successfully rescue patients from complications has been shown to be a good measure of quality care processes in hospitals. Failure-to-rescue (FTR) has been defined using secondary International Classification of Disease (ICD-9 CM) codes. Studies of FTR using these codes have demonstrated satisfactory accuracy when compared to clinical events documented in the medical record. However, a subset of the original codes for FTR, thought to be sensitive to nursing care, have failed to show the same level of accuracy. This study examined the possibility of using clinical predictors to identify failure-to-rescue.

Secondary analysis of a previous dataset was used to establish and improve diagnostic accuracy of FTR using ICD-9 CM codes as compared to the gold standard of record review. These ICD-9 CM codes performed poorly in terms of diagnostic accuracy (sensitivity and specificity) when compared to record review. A variety of clinical predictors were then tested for accuracy in the measurement of FTR compared to record review. Transfer to a higher level of care in combination with a variety of clinical predictors as well as complications following a procedure demonstrated strong sensitivity and fair specificity. Combining these clinical predictors with secondary ICD-9 CM codes did not enhance diagnostic accuracy. While specificity for clinical predictors was not robust, high levels of sensitivity for certain predictors warrants an increased level of surveillance for patients who exhibit these signs and symptoms.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The cost of errors in the care of patients in the hospital setting is estimated to be between $17 and $29 billion nationally, including health care costs, lost income and production (Kohn, Corrigan, & Donaldson, 1999). National groups such as the Institute of Medicine, Joint Commission for the Accreditation of Healthcare Organizations, and the Agency for Healthcare Research and Quality (AHRQ) have reported on this problem, provided funding to study the problem, and suggested interventions to address these errors in the delivery of patient care (Shojania, Duncan, McDonald, Wachter, & Markowitz, 2001). Yet errors in care persist with an estimated 13.5% of Medicare beneficiaries experiencing at least one adverse event during an inpatient hospital stay. Physician reviewers rated 44% of these errors as preventable leading to prolonged hospitalizations, pain and suffering with unexpected deaths found to occur in an estimated 1.5% of reviewed cases (“Adverse Events in Hospitalized Patients,” 2010).

Mortality has been the historical quality measure of choice when evaluating hospital care processes. Evidence, however, supports that mortality is more consistently linked with patient co-morbidities than hospital characteristics (Silber, Williams, Krakauer, & Schwartz, 1992). Concerns about quality occur when a hospital fails to identify and rectify complications in a timely manner. Silber and colleagues argue that some hospitals are better prepared to care for patients after a complication because they invest in quality resources and infrastructure. Therefore, quality organizations are those that can rescue the patient (Silber et al., 2007; Silber, Rosenbaum, Schwartz, Ross, & Williams, 1995). Using record review to validate complications, they established 15 broad-category secondary International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9 CM) codes that reflected possible instances of failure-to-rescue (FTR). The premise was to include most deaths that were preceded by a complication so that the number of undocumented complications would be minimal (Silber et al., 2007; Silber et al., 1995; Silber et al., 1992). Establishing the quality measure of FTR provided hospitals with an opportunity to evaluate their response to patient complications.

The original identification of FTR was modified in two independent studies in an attempt to link the concept to nursing care and resources. Expert panels were used to identify five broad categories that might be sensitive to nursing care from the original 15 proposed by Silber, et al. (Aiken, Clarke, Sloane, Sochalski, & Silber, 2002; Needleman, Buerhaus, Mattke, Stewart, & Zelevinsky, 2002). Continued study using the abridged version of FTR supported a relationship to nurse staffing in the acute care setting (Boyle, 2004). Friese and colleagues confirmed this association in surgical oncology patients (Friese, 2005). Nurse education and nurse work environments have also been linked to FTR in both acute care and surgical settings (Aiken, clarke, Cheung, Sloane, & Silber, 2003; Aiken, Clarke, Sloane, Lake, & Cheney, 2008; Friese, Lake, Aiken, Silber, & Sochalski, 2008).

The AHRQ adopted this abridged version of FTR as a patient safety indicator (PSI) with the addition of renal failure. As record reviews are expensive and time consuming, the use of ICD-9 CM codes serves as an efficient proxy for hospital quality. However, these revised versions of FTR have failed to show strong accuracy when compared to clinical events as determined by record review in several studies (Horowitz, Cuny, Cerese, & Krumholz, 2007; Silber et al., 2007; Talsma, Bahl, & Campbell, 2008). ICD-9 CM codes were not intended as measures of quality but rather for billing and workload purposes. Several studies supported a variance between codes and clinical events (Iezzoni, 1997, 1994). The addition of clinical data has improved the accuracy of ICD-9 CM codes (Iezzoni, Schwartz, Ash, & Mackieman, 1995) .

The purpose of this study was to determine if the addition of demographic and clinical predictors to the existing ICD-9 CM codes enhanced the overall sensitivity and specificity of FTR prediction. If FTR predictor sensitivity and specificity can be improved, then FTR can be used as an indicator of quality care and an early warning system of potentially untimely deaths. The research questions were:

1. How well do AHRQ secondary ICD-9 CM codes perform (sensitivity, specificity) in identifying FTR?
    a. Does the inclusion of all age groups improve the diagnostic accuracy (sensitivity, specificity) of the AHRQ secondary ICD-9 CM codes?
    b. Does the exclusion of do-not-resuscitate (DNR) status improve the diagnostic accuracy (sensitivity, specificity) of the AHRQ secondary ICD-9 CM codes?
2. How well do clinical predictors perform (sensitivity, specificity) in identifying FTR?
3. Is performance (sensitivity, specificity) improved by a combination of secondary ICD-9 CM codes and clinical predictors in identifying FTR?

1.1 CONCEPTUAL MODEL

Quality is typically broadly defined as a reflection of the values and goals in the health care system and larger society in general. The dimensions and criteria that are selected to measure quality will affect the approaches and methods used in the assessment of care. Empirical quality of care research focuses on goals. Although there are a variety of different theoretical frameworks used to examine quality, the more commonly used is that of Donabedian. Structures, processes and outcomes comprise the basic concepts of this model (Donabedian, 2005b; Yen & Lo, 2004). Structures are represented by the technologies and infrastructure capacity of an organization. Examples include: education and certifications of clinical staff; staffing levels; computerized medical record and other technology resources; teaching capacities; and other components that provide the framework in which care is delivered. Processes are the administrative and clinical methods by which care is delivered. Outcomes are the final product of the processes of care that occur within the existing care delivery structure (Donabedian, 2005a). This study focused on the care process used to rescue hospitalized patients from co-morbidity related complications.

FTR reflects processes of care that are either untimely or missing, which result in the adverse event of death for the patient (outcome). The inability to process a successful rescue from patient- related complications may also refer back to structures of care in a given organization. Hospitals with a higher investment in quality outcomes may put structures in place (e.g., board certified physician staff, improved nurse staffing and staffing mix, more educated nurse workforce) that facilitate processes concerned with early identification and intervention when patient complications occur (Silber et al., 1992). The ability to accurately predict an impending failure, through ICD-9 CM codes, clinical/demographic patient characteristics, or some combination of the two, can result in improved patient outcomes, i.e., less unexpected deaths. Accurate prediction is dependent upon the sensitivity and specificity of the process that is used.

1.2 OPERATIONAL DEFINITIONS

The following operational definitions were used in the study to address the research questions:

FTR: the death of patient from an unanticipated adverse event on an acute care unit within 30 days of admission (Silber et al., 1992).

Reference range for FTR: FTR determined by record review that serves as the best available method for establishing the presence or absence of FTR in detecting accuracy of a new or proposed diagnostic tool (FTR by AHRQ ICD-9 CM codes) (Bossuyt et al., 2003; Bossuyt et al., 2004).

Test for FTR: FTR determined by AHRQ ICD-9 CM codes. A test or diagnostic tool for prediction is defined as any systematic method for obtaining additional information regarding the current or probable future health status of the patient based on a measurable value or criteria (Bossuyt et al., 2003; Bossuyt et al., 2004).

Sensitivity: the proportion of true positive responders that have a positive test result (Griner, Mayerwski, Mushin, & Greenland, 1981).

Specificity: the proportion of true negative responders that have a negative test result (Griner et al., 1981).

AHRQ secondary ICD-9 CM codes: patients discharged with a disposition of “deceased” with a potential complication of pneumonia, DVT/PE, sepsis, acute renal failure, shock/cardiac arrest, or GI hemorrhage/acute ulcer ("PSI Technical Specifications," 2007). A complete list of these codes is provided in Appendix A.

Do-not-resuscitate (DNR): a physician order that indicates that, in the event of cardiac or pulmonary arrest, the patient is not to have aggressive intervention such as chest compressions and intubation.

Patient demographic predictors: include age, gender, race/ethnicity/ length of stay, number of transfers during the hospitalization, and body mass index. Predictor definitions are provided in Appendix B.

Clinical predictors: include clinical events described in the Institute for Healthcare (IHI) Global Trigger Tool (DTT) which is designed to identify potential patient related complications. The general categories of these predictors include: laboratory values; medications; procedures; surgery; and patient specific complications. A complete list of these predictors is found in Appendix C.

1.3 SUMMARY

The ability to successfully rescue patients from adverse complications has been shown to be an important measure of hospital quality care processes. The measure has also been shown to be related to nursing resources. However, the use of ICD-9 CM codes for quality purposes show serious limitations. There is a growing body of evidence that fails to demonstrate a strong relationship between these ICD-9 CM codes and actual clinical events. For FTR to be a useful measure of quality processes for hospitals, the accuracy of the sensitivity and specificity of proxy measures, such as ICD9-CM codes, must attain a level of acceptance. This study investigated the sensitivity and specificity of the secondary ICD-9 CM codes recommended by AHRQ, as well as sensitivity and specificity of patient clinical predictors, and then finally evaluated if sensitivity and specificity could be improved by combining ICD-9 CM codes with clinical predictors.

CHAPTER 2

REVIEW OF THE LITERATURE

The safe care of hospitalized patients has received significant attention in the last ten years. Patients come into a hospital with the expectation that no harm will come to them and potentially that they may benefit from the care that they receive. Yet evidence demonstrates that this is not always the case, with estimates of 98,000 deaths due to iatrogenic injuries occurring each year with 40-70% found to be preventable (Kohn et al., 1999; Michel, 2004). A projected 13.5% of discharged Medicare patients experienced at least one adverse event during hospitalization ("Adverse Events in Hospitalized Patients," 2010). Only one-third of these events were documented as unpreventable with the remaining two-thirds associated with errors in treatment. In reviewing surgical adverse events, 74% were found to be preventable (Shojania et al., 2001). In addition, 93% of errors reviewed in the Emergency Department were also termed preventable (Leape, 2002). An additional confounding factor is increased emphasis on efficiency that may sacrifice comprehensive care (Leape et al., 1991). A safe environment, then, is one where there are structures and processes in place to reduce the probability of errors from exposure to the delivery of healthcare.

2.1 FAILURE-TO-RESCUE

Failure- to-rescue is the probability of death after a complication and was originally identified using 15 broad categories based on secondary ICD-9 CM codes (Silber et al., 2007; Silber et al., 1995; Silber et al., 1992). These conditions were tested using the gold standard of record review by abstracting 5900 patients records with an admitting diagnosis of cholecystectomy or transuretheral prostatectomy (Silber et al., 1992). In 1995, Silber and colleagues evaluated the complication rate for patients undergoing coronary artery bypass surgery using record review. They hypothesized that if patient complications were related to hospital care, then they should be highly correlated with hospital mortality rates. However, they found that the opposite was true. The correlation between hospital rank using death rate and complication rate was .21 (95% CI 0.04-0.38). This lead to the idea that patient complications had little to do with the processes of care in a hospital (Silber et al., 1995). Rather, the early recognition and intervention of these complications by some hospitals, preventing further patient demise and death, was the true measure of quality. Thus, the ability to rescue the patient was an important clinical indicator of quality (Silber et al., 1995).

There is compelling evidence that hospital characteristics play only a small role in patient complication rates. Particularly among surgical patients that have fairly uniform care, the adverse event rate has not been associated with the mortality rate (Green, Passman, & Winfield, 1991; Green, Winfield, & Sharkey, 1990). Rather, patient characteristics such as age, history of congestive heart failure or obstructive pulmonary disease are more likely to predict an adverse event. Whether or not the adverse event progresses on to death is more closely associated with the ability of the hospital to provide early intervention. Quality organizations are those that have the resources and ability to rescue the patient from complications (Silber et al., 1992).

These studies by Silber and colleagues provided the initial groundwork for developing a new quality indicator. Preliminary work by investigators demonstrated positive relationships between specific ICD-9 CM codes and failures in care that was confirmed by record review. Only elective surgical cases were included in the sample records that were studied because of the tendency towards uniform care for surgical patients.

2.2 FAILURE-TO-RESCUE AS A MEASURE OF NURSING QUALITY

The 24-hour presence of nurses within the hospital setting constitutes a patient surveillance system (Shever, 2007) making FTR highly sensitive to nursing care. Using expert panels, Needleman, et al, (2002) took the original 15 complications suggested by Silber (1992) and developed a subset of five broad categories composed of ICD-9 CM codes thought to be related to nursing care: cardiac arrest/shock; upper gastrointestinal bleeding; deep vein thrombosis/pulmonary emboli; pneumonia; and patient safety indicators. In a large national study (2002) with over one million discharges, Needleman, et al. demonstrated relationships between nurse staffing and FTR using ICD-9 CM codes. Specifically they found that a higher proportion of RNs but not greater numbers of RN hours were associated with lower rates of FTR among medical patients. In surgical patients, a greater number of RN hours per day was associated with a lower rate of FTR (Needleman et al., 2002). In a follow-up study, staffing that was less than 8 hours of the targeted nursing hours per patient day was significantly associated with a 2% odds of an increase in mortality (OR 1.02 95% CI 1.01-1.03) (Needleman et al., 2011).

In a separate study among 232,342 surgical patient records in Pennsylvania hospitals, Aiken and colleagues (2002) showed similar relationships between nurse staffing and FTR also using secondary ICD-9 CM codes. An additional patient for each RN resulted in a 7% increase in the odds of FTR (OR 1.07 95% CI 1.02-1.11) (Aiken et al., 2002). A one hospital study of 11,496 patient records found an inverse relationship between nurse measures of autonomy and collaboration and incidence of FTR (r=0.28) which explained 24% of the variance in FTR (Boyle, 2004). Among oncology patients, Friese, et al., looked at FTR and practice environment using the Practice Environment Scale. There was a 48% increase in the odds of FTR (OR 1.48 95% CI 1.07-2.03) among hospitals with poor work environments (Friese et al., 2008).

The evidence continues to grow supporting FTR as a nurse sensitive indicator. Nursing characteristics such as staffing, skill mix, and practice environment have shown significant associations with FTR. If measures of FTR using secondary ICD-9 CM codes can demonstrate sufficient levels of sensitivity and specificity, then this patient outcome can be used to support the work of nurses. However, there is minimal evidence, to date, that has evaluated how well these secondary codes perform.

2.3 FAILURE-TO-RESCUE USING ICD-9 CM CODES

Using record review to predict cases of FTR is a time consuming and expensive process. AHRQ, therefore, has taken the secondary ICD-9 CM codes proposed by Needleman and Aiken (Aiken et al., 2002; Needleman et al., 2002) an adopted FTR as a patient safety indicator. An additional category of renal failure was added to the existing five categories previously identified as sensitive to nursing care. The final categories which encompass FTR by AHRQ definition are: acute renal failure; deep vein thrombosis/pulmonary embolism; gastrointestinal bleed; shock; sepsis; and pneumonia. ("Overview: Patient safety indicators from the Agency of Healthcare Quality and Research," 2007). However, the effectiveness of these secondary ICD-9 CM codes to measure FTR has not been well established. A quality indicator must meet several criteria to be useful: face validity; minimum bias by adequate control of confounding variables, criterion validity; precision; ease of measurement and resistance to falsification (Halfon, 2006). The Institute of Medicine Committee on Regional Health Data Network has mandated the absolute requirement of reliability and validity of data prior to public dissemination of derived quality measures (Iezzoni, 1997). An international panel of experts in quality of care identified the development and validation of algorithms to verify the logic and internal consistency of coding of hospital abstract data as one of the highest priorities for future research (DeCoster, 2006). If FTR using secondary ICD=9 CM codes is used as a measure of hospital quality of care, then they must demonstrate adequate sensitivity and specificity.

There is significant evidence that the use of secondary ICD-9 CM codes, as measures of quality, may have limitations. These codes are retrospective, determined at patient discharge and reflect conditions that were diagnosed or detected at any time during the hospitalization. The codes were not intended nor designed for the purpose of identifying adverse events. The over 15,000 diagnostic codes do not provide a clinical description to define each code. Hospitals code differently with varying degrees of accuracy and thoroughness (Iezzoni et al., 1994). Several studies have demonstrated a lack of correlation between ICD-9 CM codes and clinical documentation. In a review of 974 patient records in California, at least one clinical risk factor for heart attack was missing in 65% of the records, and 35% of records were coded with a risk factor that was not found in the medical chart (Iezzoni, 1997). In a review of 485 randomly sampled hospitalizations of Medicare beneficiaries, McCarthy, et al., (2000) found that, except for acute myocardial infarction, ICD-9 CM codes were confirmed by clinical data in less than 60% of charts that were reviewed. When adding information, such as laboratory data or nursing assessment data, ICD-9 CM codes were found to have a better predictive ability of mortality (Davis, 1995; McCarthy, 2000; Pine, 1997). Iezzoni and colleagues (1995) compared two models using clinical data with two models using administrative data from discharge abstracts of patients with myocardial infarction and found that the measures based on the discharge abstracts provided better mortality predictions than the measures using clinical data. Risk adjusted models that include hospital acquired complications that typically precede death usually predict death better. However, a risk adjusted model for disease severity at admission that includes potentially fatal hospital acquired complications may mask inadequate hospital care by increasing the measure of risk for patients whose condition deteriorates during hospitalization (Iezzoni et al., 1995).

Although it is highly labor intensive, retrospective record review has been used as an alternative to using administrative databases for measuring quality and safety. Postoperative care, medical injury, and malpractice litigation, as well as readmission rates, have all received significant attention in large, multi-site studies where record abstraction was used (Gawande, Thomas, Zinner, & Brennan, 1999). Although standardized tools were not used, in general, the studies incorporated the use of clinical experts for record review associated with a mechanism for inter-rater reliability (Horowitz et al., 2007). Kashner (1998) used patient treatment files that included ICD-9 CM codes and compared them to a random sample of 414 inpatient discharges. Records were abstracted in a uniform way by review nurses, and medical record coders were blinded to administrative file entries. The discharge summaries in the patient treatment files showed higher estimates of disease prevalence than record review by reporting an additional diagnosis per discharge (Kashner, 1998).

The use of AHRQ identified secondary ICD-9 CM codes for predicting FTR may result in misclassification for two reasons. First, the links between these codes and actual events, as documented in the medical record, remain untested. Secondly, the codes reflect a non-clinician's (medical coder) interpretation of clinical events as recorded in the medical record. Silber, et al., (2007) found that 42% of deaths were omitted when comparing patient AHRQ classifications of FTR to his original record abstraction. Reliability was better for the original measure of FTR with a correlation of 0.32 compared to correlation of 0.18 when using patient AHRQ secondary ICD-9 CM codes.

Horowitz and colleagues (2007) used data from University Health Consortium hospitals to compare patient safety indicator FTR measures with record review and found almost 50% false positives. As only charts identified as FTR by ICD-9 codes were used for the record abstraction, review of all deaths might have determined an even high rate of false positives as well as identifying false negatives (Horowitz et al., 2007). An additional limitation of this study was the use of facility staff for record review without standardized measurement tools.

In five mid-western hospitals, the FTR rate using record review was 0.03%, which is significantly lower than other reported rates using patient safety indicator ICD-9 CM codes. In addition, specific patient factors, heart rate, respiratory rate, temperature, serum sodium level, and urine output, were found to be significant predictors of FTR (Bobay, 2008). Talsma and colleagues (2008) found almost half of all patients identified as FTR had the complication present on admission (Talsma et al., 2008).

As a response to these studies, the AHRQ has added conditions that are present on admission in determining the patient safety indicator of FTR. In the most recent release of patient safety indicator software, the measure for FTR has been replaced with death in low mortality DRGs and surgical deaths. Yet the conceptualization of failure-to-rescue remains an important measure of hospital quality and more specifically, nursing care. Continued work in this area is necessary to refine measures of FTR so that they provide hospitals with ease of measurement while maintaining a sufficient level of accuracy.

2.4 THE USE OF CLINICAL AND DEMOGRAPHIC PREDICTORS WITH ADMINISTRATIVE DATA

The addition of clinical data, such as laboratory values, has been shown to improve measures of mortality when using ICD-9 CM codes. Iezzoni, et al., found that by adding specific laboratory values and information from the nursing admission assessment, discharge abstracts were a better predictor of mortality than clinical data alone (Iezzoni, 1994). Adding clinical laboratory data and patient demographic data to diagnosis-related groups improved the ability to predict length of stay (Goldman, Easterling, & Sheiner, 1989).

While not specifically related to FTR, patient demographics, such as race, ethnicity, age and gender, have all been shown to be associated with outcomes of care. Minority men reported less quality of life after prostate surgery (Coffey, Andrews, & Moy, 2005). Among patients undergoing coronary artery bypass surgery, the unadjusted mortality rate was higher in blacks at 30, 90, and 365 days than whites post-surgery, and continued to be higher even after adjusting for patient characteristics (Konety, Vaughan Sarrazin, & Rosenthal, 2005). Mortality was also greater for men than women (Konety et al., 2005). For patients with idiopathic pulmonary fibrosis waiting for lung transplant, adjusted mortality rate was higher for non-Hispanic blacks and Hispanics than non-Hispanic whites (Lederer, Caplan-Shaw, & O'Shea, 2006). In multiple studies, age was a determinant of 30-day mortality rates (Fleisher, Pasternak, Herbert, & Anderson, 2004; Iezzoni, 1997). Using the AHRQ patient safety indicator software, across sixteen states, there was a significantly higher rate of adverse outcomes based on racial, ethnic, and socioeconomic subgroups (Romano et al., 2003). Non-Hispanic Blacks in particularly showed higher rates of FTR, even when controlling for socioeconomic levels (Trivedi, Sequist, & Ayanian, 2006).

The use of secondary ICD-9 CM codes as a measure of FTR provides a ready source of data. However, if these codes do not provide an accurate reflection of clinical events as recorded in the medical record, then any measures of quality or associations with nursing care are suspect. There is a sufficient body of evidence that demonstrates that the addition of clinical and demographic patient characteristics have improved the diagnostic accuracy of ICD-9 CM codes in other measures of quality. If this finding can be applied to measures of FTR, then it may be used as an indicator of quality nursing care.

2.5 SUMMARY

Measures of safe patient care are important indicators of quality for hospitals. Patient characteristics are more closely aligned with complications than hospital characteristics. Hospitals that have the resources for timely identification with these these complications to prevent patient decline are thought to have a higher quality of care. Nurses constitute 24-hour surveillance specifically designed for early identification and intervention of untoward patient events. Because record review to identify cases of FTR is expensive and time consuming, administrative data such as secondary ICD-9 CM codes provide an efficient mechanism for measuring patient outcomes. However, in limited study, these codes have not performed well when compared to the gold standard of record review in identifying FTR. There is some evidence that the inclusion of clinical data has improved the performance of ICD-9 CM codes with other outcome measures. Therefore, the addition of clinical data to the AHRQ identified ICD-9 CM codes might improve their accuracy in identifying FTR. Further investigation in this area is required before these codes can be used reliably as a measure of FTR. This study evaluated the diagnostic performance of AHRQ secondary ICD-9 CM codes as measures of FTR compared to the gold standard of record review. To address limitations from previous studies, record reviews were conducted by independent experts with no connection to the facilities that were studied. In addition, standardized tools were used for record abstraction.

CHAPTER 3

METHODOLOGY

This chapter describes the research methodology and data that were used to address the following research questions:

1. How well do AHRQ secondary ICD-9 CM codes perform (sensitivity, specificity) in identifying FTR?
    a. Does the inclusion of all age groups improve the diagnostic accuracy (sensitivity, specificity) of the AHRQ secondary ICD-9 CM codes?
    b. Does the exclusion of do-not-resuscitate (DNR) status improve the diagnostic accuracy (sensitivity, specificity) of the AHRQ secondary ICD-9 CM codes?
2. How well do clinical predictors perform (sensitivity, specificity) in identifying FTR?
3. Is performance (sensitivity, specificity) improved by a combination of secondary ICD-9 CM codes and clinical predictors in identifying FTR?

3.1 RESEARCH DESIGN

This study examined whether the addition of demographic and clinical predictors to AHRQ secondary ICD-9 CM codes improved the diagnostic accuracy of FTR. Using a descriptive design to explore these candidate predictors and their relationship to FTR allowed for investigation into an area that has limited evidence. One limitation of descriptive design is that it does not allow for any inference or causality among the predictors. Once determinations can be made about the usefulness of adding clinical predictors to identify FTR, then further study may be conducted to test for causal relationships.

The study was guided by the Standards for Reporting of Diagnostic Accuracy (STARD) model that provides a testing framework for a diagnostic study. Studies of diagnostic performance compare the outcomes from the test(s) with a referenced standard. FTR using AHRQ secondary ICD-9 CM codes served as the test. FTR determined by record review was the reference standard.

3.2 SETTINGS

The Veterans Health Administration (VHA) has an integrated electronic medical record. Both clinical and demographic data from the record interface with large, national databases. The Veterans Integrated Service Network (VISN) 7 consists of VHA hospitals in Georgia, Alabama, and South Carolina. To facilitate performance improvement projects or research at the local level, a corporate data warehouse was created at the VISN level where it is stored on a protected server. This warehouse potentially contains all objective data that is contained in the medical record such as: medication; laboratory values; imaging; ICD-9 CM codes; procedure codes; and demographics. These data are electronically captured from the medical record and as a result, it was assumed that they accurately reflected care as documented in the medical record. Clinical and research staff can access these data by completing a request for the specific information required. The data are then transferred to a protected server at the facility level. Only staff with a specific need to know have access to these data files as they contain both protected health information and patient identifiers. Data may not be removed from the protected server and all analysis of the data must be done on site at the facility once approval is received from the facility Research and Development Committee.

A request was made for clinical data from the five tertiary care centers in VISN 7: Birmingham, Atlanta, Augusta, Columbia, and Charleston. The VHA nationally ranks hospitals based on size and complexity. All five of the hospitals were ranked as Level 1 facilities indicating academic affiliations, extensive surgery and invasive procedure programs, availability of intensive care units, and emergency departments. The five hospitals were fairly homogeneous in size and structure. All hospitals were in urban settings. As these data reflect patients who died within 30 days of admission, mortality rates for the five hospitals were analyzed and no significant difference was found.

3.3 SAMPLE

The study was a secondary analysis of an existing dataset that included records of all patients who died within 30 days of admission at any of five tertiary care centers in the southeastern United States. The sample size for the original study was determined at a level that would result in confidence intervals that were small enough to be clinically important. Based on recommendations by Flahault and colleagues on calculating sample size for studies of diagnostic tests, 624 records were required to reach an expected sensitivity of 85% with a minimal sensitivity of 80% (Flahaut, Cadillhac, & Thomas, 2005). To achieve this number, a request was made for medical records, starting in January, 2010, of all patients who fit the requirements of 30-day mortality at the study hospitals. Using random number generators, 624 records were selected for expert nurse review. Each record contained the following demographic data: race, ethnicity, gender, zip code, birth date, death date, facility, and admission date. All hospitals included in the study had similar numbers of records in the dataset.

Twenty-two percent of patient records were from the Atlanta VA, 20.9% were from the Augusta VA, 23.7% were from the Birmingham VA, 14.7% were from the Charleston VA, and 17.7% were from the Columbia VA. Male patient records composed 98.7% of the sample. The majority of records were from white patients (56.1%) with 33.7% from African Americans. The remainder of records represented small percentages of Asians, Native Hawaiian/Pacific Islander and Native Americans. As race/ethnicity is not a mandatory question on registration, 9.8% of records did not have race recorded.

There are quality measures in place to assure that both demographic and clinical data are as accurate as possible in the medical record. The medical record is an integrated, electronic record which travels with the patient as he seeks care in other VHA facilities. Most demographic data, such as date of birth, is obtained from military discharge papers as the Veteran first registers for care. Gender can be mistakenly entered on initial registration, particularly as most Veterans who receive care are male, but is generally corrected once the Veteran begins to receive care. Race has traditionally not been well captured and was previously not a required field during registration. However, in the past five years, race is consistently recorded and readily available in the dataset that is under analysis. As with gender, race can be entered incorrectly upon registration. As an additional security measure, patient photographs are now included with the patient's medical record. Erroneous entries of race are often found and corrected when the picture does not match the race of record. Date of death as recorded in the medical record must match the official death certificate, leaving little room for error in the recording of this event.

The entry of secondary ICD-9 CM codes was the one predictor that may have lacked accuracy in comparison with clinical events as recorded in the medical record. Despite consistent processes of a large health care system, the accuracy of administrative coding may vary by medical center. Processes were in place throughout VISN 7 for inter-rater reliability of coding with an acceptable rate of greater than 90%.

3.4 MEASUREMENT

This study used two different measures of FTR, secondary ICD-9 CM codes and expert opinion determined through record review, to establish which method provided the best diagnostic accuracy in terms of sensitivity and specificity. Using the STARD model as a framework, secondary ICD-9 CM codes were considered the test and FTR determined by record review was considered the reference standard.

Reference standard – FTR by record review: To address limitations in other studies of FTR using record review, nurse abstractors were asked to use The Global Trigger Tool (GTT). GTT was developed by the Institute of Healthcare Improvement (IHI) to facilitate and standardize the record review process. The GTT, designed as a method for identifying harm over time, was developed using expert panels who reviewed the literature on adverse events throughout various settings in the hospital. The GTT was then tested to prioritize these "triggers" in hundreds of hospitals. Over time, IHI has added, deleted and adjusted triggers to reflect changes in treatment that help identify possible adverse events, whether preventable or not. The GTT thus provides a mechanism to increase the efficiency of record review by focusing first on the identification of potential adverse events which are then evaluated by the record reviewers to determine if FTR has occurred. The GTT is therefore simply used as an aid to guide the record review to increase efficiency of the process (deWet & Bowie, 2009; Griffin & Resar, 2007).

The Test - FTR by AHRQ ICD-9 CM codes: Using the ICD-9 CM codes identified by AHRQ, the database was analyzed for specific cases of FTR within the six large classification of patient conditions: acute renal failure; sepsis; shock; gastrointestinal bleed; deep vein thrombosis/pulmonary embolism; and pneumonia. Patient who were 75 years of age or older were excluded based on existing definitions from AHRQ along with exclusion ICD-9 codes. These codes are listed in Appendix A.

3.5 DATA EXTRACTION PROCEDURES

Because Veteran care is provided across the continuum, it was not possible to include only 30-day mortality that occurred during the admission. As a result, the dataset contained records of patients who died outside of the hospital setting but within 30 days of their last admission. Therefore, the nurses excluded these records during their review and additional records were randomized from the original data pull.

The two expert registered nurse record abstractors were given written definitions for each clinical or demographic predictor to be collected. The nurses were contracted employees from a company nationally recognized for record abstraction. To assure consistency between the abstractors, a web-in-air was held to provide education on data definitions, record manipulation and review. All records were then reviewed independently by the two expert nurse abstractors using the standardized GTT to help identify potential adverse events. Weekly telephone meetings were conducted with the principal investigator (PI) to answer questions and help in record navigation. Prior to the weekly calls, the PI reviewed each of the FTR determinations made by the nurses and identified records where the nurses were not in agreement. The abstractors were asked to hold an additional telephone conference to discuss these cases and arrive at a consensus opinion. Overall, inter-rater reliability was high. Nurses were in initial agreement in their determination of failure or not in 97% of records reviewed.

As a result of the record review, demographic and clinical predictors were collected and used in this study. Patients who had a physician order for “do not resuscitate” upon admission or within 24 hours of admission and prior to any documented adverse event were excluded from record review and counted as “no-failure”.

The individual datasets from each nurse were then merged and reviewed to assure that all required elements were completed. The final dataset contained 610 usable records.

Table 3.1. Final Sample

Hospital	Number in the final sample	Percent in the final sample	Not failure-to rescue (NFTR)	FTR	% FTR
Atlanta	137	22.4%	120	17	12.4%
Augusta	148	24.3%	115	33	22.3%
Birmingham	128	21.0%	115	13	10.1%
Charleston	74	12.1%	64	10	13.5%
Columbia	123	20.2%	115	8	6.5%
Totals	610		529	81	13.3%

The secondary ICD-9 CM codes identified by AHRQ were applied to the final sample of 610 records. Each record was determined to be either FTR or NFTR. The results were compared with the gold standard of record review. As previous studies of FTR using record review did not include all deaths, but only FTR deaths, false negatives could not be evaluated. In this study, all deaths were included in the final analysis, so that false positives as well as false negatives could be determined. This is a necessary step in calculating sensitivity and specificity.

3.6 DATA ANALYSIS

Data were analyzed for all ICD-9 CM codes and clinical predictors. Measures of sensitivity and specificity were used to determine the diagnostic accuracy FTR using the various predictors.

Descriptive Analysis: Descriptive statistics and frequency distributions were calculated as appropriate for all demographic and clinical predictors. A total of 43 different clinical predictors were collected by the nurse abstractors. A list of these predictors is available in Appendix C.

Sensitivity and specificity: Performance characteristics such as the quality and usefulness of a diagnostic test were described through sensitivity and specificity. This approach was used for research questions 1, 2, and 3, at the 95% confidence limit level.

Sensitivity was the proportion of times that a death was labeled FTR by the secondary ICD-9 CM codes (test) compared to true failure cases by the record review (reference standard) among all deaths. Specificity was the proportion of times that a death was labeled as not FTR (NFTR) by these codes compared with true non-failures by the record review among all deaths.

3.7 SUMMARY

Secondary analysis of a previous dataset was used to establish the diagnostic accuracy of FTR using AHRQ secondary ICD-9 CM codes as compared to the gold standard of record review. A variety of candidate predictors, such as age, race, length of stay, and clinical predictors were used to test improvement in the performance of these codes. Sensitivity and specificity, were used to determine the diagnostic accuracy of FTR using ICD-9 CM predictors, demographic and clinical predictors and combinations of predictors.

CHAPTER 4

FINDINGS AND DISCUSSION

FTR has the potential to be an important indicator in the quality of care of the hospitalized patient. A growing body of evidence supports that the current predictors, secondary ICD-9 CM codes, may not be accurate when compared to actual clinical events. This study continues to add to the knowledge base regarding FTR. The specific research questions that were addressed were:

1. How well do AHRQ secondary ICD-9 CM codes perform (sensitivity, specificity) in identifying FTR?
    a. Does the inclusion of all age groups improve the diagnostic accuracy (sensitivity, specificity) of the AHRQ secondary ICD-9 CM codes?
    b. Does the exclusion of do-not-resuscitate (DNR) status improve the diagnostic accuracy (sensitivity, specificity) of the AHRQ secondary ICD-9 CM codes?
2. How well do clinical predictors perform (sensitivity, specificity) in identifying FTR?
3. Is performance (sensitivity, specificity) improved by a combination of secondary ICD-9 CM codes and clinical predictors in identifying FTR?

4.1 DESCRIPTIVE RESULTS

The average age of the all patient in the records that were reviewed was 70.3 ranging from age 20 to 100. On average, hospital length of stay was 9 days (sd =7.0). All records were coded as "death" as their final discharge disposition. Records identified by the expert nurses as FTR cases were significantly younger than non-failure cases (66.3 versus 70.9, p=0.001). They were also more likely to be transferred during their hospital stay (1.11 times versus 0.89, p= 0.03) which supports findings by Shever (Shever, 2007). Records identified as FTR had a longer length of stay (10.06 versus 8.23, p=0.01) which was consistent with previous evidence (Silber, 1998). The records determined to be FTR by the expert nurses were also healthier on admission with lower blood urea nitrogen (BUN) levels (20.0 vs. 23.4, p=0.04), creatinine (1.3 vs. 1.6, p=0.01), hemoglobin (11.9 vs. 11.3, p=0.02) and hematocrit (35.6 vs. 33.9, p=0.03). The group determined to be FTR, however, were also more likely to be overweight (BMI 28.5 vs. 25.6, p=0.002).

4.2 RESEARCH QUESTION #1

How well do AHRQ secondary ICD-9 CM codes perform (sensitivity, specificity) in identifying FTR?

Comparing FTR by record review with FTR by ICD-9 CM codes, sensitivity was 27.7% (95% CI 24.2-31.3) and specificity was 72.3% (95% CI 68.8-75.9). As the ICD-9 CM codes are grouped into specific diagnoses, each of these six was analyzed with the following results:

Table 4.1 Analysis of Specific Diagnostic Classifications.

	Sensitivity	95% CI	Specificity	95% CI
Acute renal failure	9.6%	0.07-0.12	88.7%	0.86-0.91
Deep vein thrombosis/pulmonary embolism	3.6%	0.02-0.05	98.3%	0.97-0.99
Pneumonia	10.6%	0.08-0.13	95.2%	0.94-0.97
Shock	22.3%	0.19-0.26	89.4%	0.87-0.92
Gastrointestinal Bleed	7.1%	0.05-0.09	96.5%	0.95-0.98
Sepsis	22.3%	0.19-0.26	88.6%	0.86-0.91

No individual diagnosis outperformed the full model. As acute renal failure had the lowest findings, it was removed from the full model but performance only improved slightly. The three strongest predictors: pneumonia; shock; and sepsis were evaluated against the full model and again the full model outperformed individual diagnoses. Overall, FTR using AHRQ ICD-9 CM codes has poor sensitivity with good specificity which means that these codes perform well in identifying non-failures but perform poorly at identifying true failures.

Research Question #1a

Does the inclusion of all age groups improve the diagnostic accuracy (sensitivity, specificity) of the AHRQ secondary ICD-9 CM codes?

By removing the age restriction that is currently in place with the AHRQ predictor (75 years or older), performance of ICD-9 codes improved significantly with a sensitivity of 43.5% (95% CI 0.40-0.47) but specificity decreased to 75.1% (95% CI 0.71-0.78). In identifying potentially preventable patient deaths, sensitivity is much more important from a clinical standpoint than specificity. The age restriction should be removed in any future work with secondary ICD-9 codes as measures of FTR.

Research Question 1b

Does the exclusion of do-not-resuscitate (DNR) status improve the diagnostic accuracy (sensitivity, specificity) of the AHRQ secondary ICD-9 CM codes?

Excluding records with a DNR order did not improve the performance of the measure with sensitivity falling to 24.1% (95% CI 0.21-0.27) and specificity also decreasing to 69.4% (95% CI 0.66-0.73). This is an important finding as a DNR order precludes any rescue attempt following serious patient complications. The predictor did not impact the sensitivity of the secondary ICD-9 CM codes.

4.3 RESEARCH QUESTION #2

How well do clinical characteristics perform (sensitivity, specificity) in identifying FTR?

Clinical predictors were evaluated for collinearity prior to any further analysis. Eigen values ranged from 0.006 to 13.32 indicating no concern for collinearity among the predictors under analysis.

The performance of each clinical predictor was evaluated individually against FTR. The clinical predictors that outperformed the sensitivity of AHRQ secondary ICD-9 CM codes were: positive cultures for infection; increase in BUN and creatinine; decrease in hemoglobin/hematocrit; blood transfusion; transfer to a higher level of care;

intubation/reintubation; and oversedation/hypotension. Each of these were less specific, however, than the AHRQ codes. Only complications following a procedure outperformed AHRQ ICD-9 CM code with sensitivity of 65% (95% CI 0.61-0.70) and specificity of 83% (95% CI 0.80-0.87). A complete list of the findings of all of the clinical indicators is available in Appendix C.

Various combinations of clinical predictors were tested for performance. Transfer to a higher level of care with a decrease in hemoglobin/hematocrit or blood transfusion outperformed AHRQ secondary ICD-9 CM codes without an age restriction (sensitivity 44.6% 95%CI 0.41-0.49 and specificity 82.7% 95% CI 0.80-0.86). Transfer to a higher level of care and any of the following predictors: positive culture;, two-fold increase in creatinine; decrease in hemoglobin/hematocrit; increase in BNP; chest x-ray positive for pneumonia; or sedation/hypotension was also tested with good results. Sensitivity was 50.6% (95% CI 0.47-0.55) with specificity at 76.1% (95% CI 0.73-0.79).

Clinical predictors that were related to surgery all performed poorly in terms of sensitivity but this may have been due to the small sample of records with surgical procedures (n=81). These predictors, however, had high specificity indicating good performance for records that were not failures.

Finally, clinical predictors that had the strongest performance were tested against FTR by record review. These included: two-fold increase in creatinine; 25% decrease in hemoglobin/hematocrit; transfusion of blood products; complications following a procedure; transfer to a higher level of care; intubation/reintubation; sedation/hypotension. Any combination of these variables showed strong sensitivity of greater than 90%. However, specificity decreased to less that 40% indicating good performance for true failures but less ability to predict non-failures.

4.4 RESEARCH QUESTION #3

> Is performance (sensitivity, specificity) improved by a combination of secondary ICD-9 CM codes and clinical or demographic characteristics in identifying FTR?

The final analysis was done using the AHRQ secondary ICD-9 CM codes with a model that included any of the strongest predictors. There was no improvement in the diagnostic accuracy of the secondary ICD-9 CM codes for sensitivity (27.7% 95% CI 0.24-0.31) but specificity did improve to 83% (95% CI 0.80-0.86). Combining secondary ICD-9 CM codes with the clinical predictor with the best performance, complication following any procedure, sensitivity fell to 16% (95% CI 0.13-0.19) but specificity improved to 97.7% (95% CI 0.97-0.99). Similar to surgical predictors, this combination does not provide a strong indication of failures but performs very well at identifying non-failures.

4.5 SUMMARY

The six large classes of conditions currently associated with FTR: acute renal failure; deep vein thrombosis-pulmonary embolism; pneumonia; shock; gastrointenstinal bleed; and sepsis show fair accuracy when compared to the gold standard of record review. No one specific diagnosis outperformed the full model. Removing the age restriction did improve performance both for sensitivity and specificity. Clinical predictors with the highest sensitivity had lower specificity than the model with the secondary ICD-9 CM codes. This indicates that clinical predictors are useful at identifying true failures but perform less well for non-failures. Clinical predictors related to surgical records had poor sensitivities with strong specificities. Adding clinical predictors to the secondary ICD-9 CM code model did not provide any improvement in performance in terms of sensitivity but did improve the specificity of the model.

CHAPTER 5

CONCLUSIONS

The measure of quality in the hospitalized setting is an inexact process. Historically, mortality has been used as the basis for comparing adverse patient outcomes among hospitals. But research supports that mortality is much more closely tied to patient characteristics than hospital characteristics. Rescuing patients from complications has gained attention as a strong indicator of hospital quality. The ability to predict FTR has the potential to decrease unexpected mortality among hospitalized patients. The current use of secondary ICD-9 CM codes to identify FTR has failed to show strong diagnostic accuracy when compared to actual clinical events recorded in the medical record. The purpose of this study was to determine if the addition of clinical predictors would improve the accuracy (sensitivity, specificity) of secondary ICD-9 CM codes in identifying FTR.

5.1 FAILURE-TO-RESCUE AND ICD-9 CM CODES

The specificity and sensitivity of the ICD-9 CM codes recommended by AHRQ was consistent with reported findings in the literature. Silber found that 42% of deaths that were true failures were omitted when using the abridged version of ICD-9 CM codes (Silber et al., 2007) and both Horowitz and Talsma reported up to 50% false positives (Horowitz et al., 2007; Talsma, 2008). This study found that ICD-9 CM codes predicted only 27.7% of true FTR, based on record review. ICD-9 CM codes performed substantially better at predicting non FTR cases at 72.3%. When the specified codes are not present, there is a fairly good chance that a failure did not happen. However, if the codes are present, there is a still a 72% chance that a failure did not happen.

Evaluating the individual performance of the large category diagnoses of FTR using secondary ICD-9 CM codes was consistent with findings by Talsma (2008) who analyzed the performance of acute renal failure, pneumonia, and DVT/PE codes. The study supported his findings of sufficient sensitivity but poor specificity with 33% of the records that he reviewed failing to meet the conceptual definition of FTR (Talsma, 2008). Higher sensitivity for DVT/PE codes were identified by Romano, et. Al (2009) than in this study (68%) but only surgical records were reviewed which might account for the discrepancy(Romano et al., 2009).

In a presentation on the validity of the PSIs, Romano calls for national consistency in coding patient discharges (Romano, 2008). In a review of five AHRQ accepted PSIs, only accidental puncture or laceration had a PPV of greater than 90%. Pneumothorax, postoperative DVT/PE, infections due to medical care and postoperative sepsis all had PPV raging between a low of 48% (DVT) to a high of 78% (pneumothorax). Multiple coding errors were identified during the review of these safety indicators (Romano, 2008). A more uniform approach to ICD-9 CM coding has the potential to enhance the use of these codes for quality purposes. Additional study in this area is indicated.

Changes to the software provided by AHRQ to measure patient safety indicators removed FTR as an indicator and substituted death among surgical inpatients with serious treatable conditions ("Patient Safety Indicators Technical Manual Version 4.2," 2010). Although many of the ICD-9 CM codes remain the same, the population is limited to surgical patients. In addition, the age exclusion has been changed from 75 to 90 and the category for acute renal failure has been removed. To accommodate the concept of FTR in the medical patients, death in low mortality DRGs has been included in the PSIs. Whether or not these two revised PSIs reflect actual clinical events has not yet been reported in the literature.

5.2 FAILURE-TO-RESCUE AND AGE

The inclusion of all age groups within the AHRQ secondary ICD-9 CM codes improved the diagnostic accuracy of FTR, increasing sensitivity from 27.7% to 43.5%. Specificity also improved from 72.3% to 75.1%. This is consistent with two previous studies that have linked age with 30-day mortality (Fleisher et al., 2004; Iezzoni, 1997). In the revised version of the AHRQ software, the age exclusion for deaths among surgical patients has been increased to 90. This should improve diagnostic accuracy of using secondary ICD-9 CM codes to identify failures in care although this change has not yet been studied and reported in the literature.

5.3 FAILURE-TO-RESCUE AND DO-NOT-RESUSCITATE

Patients who indicate that they do not wish to be resuscitated in the event of a cardiac or pulmonary arrest must be excluded from any measures of failure of care. It can hardly be a failure if the patient expressly asks that he not be rescued. This patient preference is not currently captured in any secondary ICD-9 CM codes. There are ICD-9 CM codes that are related to palliative care and it is possible that these could serve as a proxy for a DNR order. However, performance of FTR was not improved when excluding DNR records.

5.4 FAILURE-TO-RESCUE AND CLINICAL AND DEMOGRAPHIC PREDICTORS

Demographic predictors, specifically gender and race, have been found to be related to poorer outcomes in hospitalized patients. Minority men reported lower quality of life following prostate surgery (Coffey et al., 2005), there was a higher mortality rate for black men undergoing coronary artery bypass surgery even when adjusted for comorbidities (Konety et al., 2005). Non-Hispanic blacks had a higher mortality rate than whites when awaiting lung transplantation (Lederer et al., 2006) as well as higher rates of FTR (Trivedi et al., 2006). However, in this study, race and gender were not associated with FTR, although the number of women included in the review was very small.

There were significant differences in the study population between FTR and NFTR patients in terms of age, length of stay, number of transfers, and BMI. However, when these predictors were compared with FTR, there was no significant association with any of the demographics. Further study is indicated in the area of demographic predictors.

There is some limited study that has demonstrated that clinical data will enhance the performance of secondary ICD-9 CM codes in predicting quality outcomes. Laboratory data and nursing assessment data have both been found to be useful in improving diagnostic accuracy of ICD-9 CM codes (Davis, 1995; Iezzoni et al., 1995). Several studies have recently shown interesting relationships between deterioration in vital signs and failures (Cei, Bartolomei, & Mumoli, 2009; Lee et al., 2008; Moon, Cosgrove, Lea, Fairs, & Cressy, 2011). Bobay found significant but subtle differences among surgical records with changes in vital signs (heart rate, respiratory rate, and temperature) as well as an increase in serum sodium and decrease in urinary output (Bobay, 2008). Fifty percent of patients died who had a decrease in spot oxygen concentration (spO2) of less than 90% in a study of 6303 patient records following cardiac arrest (Buist, Bernard, Nguyen, Moore, & Anderson, 2004).

In this study, clinical predictors individually did not outperform secondary ICD-9 CM codes except for complications following a procedure. When used in combination, there were some improvements in diagnostic accuracy. Transfer to a higher level of care in combination with drops in hemoglobin/hematocrit, transfusion of blood products, BNP, chest x-ray positive for pneumonia or hypotension/sedation all performed better than secondary ICD-9 CM codes. While Shever (2007) found that multiple transfers during an inpatient stay were associated with FTR, she did not specifically evaluate transfers to a higher level of care (Shever, 2007). Predictors associated with surgical care had strong specificity but lacked adequate sensitivity to identify true failures. This is in contrast with findings by Silber and colleagues (Silber et al., 2007; Silber et al., 1995; Silber et al., 1992) but their studies used the full 15 secondary ICD-9 CM codes. In addition, the sample size for this study among surgical records was very small (n=81) and is not sufficiently adequate to draw conclusions. Other clinical variables that showed strong sensitivity included two-fold increase in creatinine, intubation/reintubation, and sedation/hypotension.

While clinical predictors, except for those related to surgical care, did not show strong specificity, allowing for the identification of records which were not failures, the sensitivity among many of the clinical predicators was quite high. Close monitoring of these clinical predictors could be clinically useful in the early identification and intervention of patient-related complications. While such a process would ultimately include patients who were not at risk for failure, this is outweighed by the opportunity to rescue those who are. With most acute care facilities now using an electronic medical record, alerts could be set to notify the clinician when any of the most high risk clinical predictors are occurring.

5.5 FAILURE-TO-RESCUE, ICD-9 CM CODES, AND CLINICAL PREDICTORS

Adding the clinical predictors to the AHRQ secondary ICD-9 CM code model did not enhance diagnostic accuracy. While specificity was improved through the addition of clinical predictors, no single or combination of clinical information improved the sensitivity of secondary ICD-9 CM codes. This is in contrast to previous studies where clinical information enhanced the accuracy of ICD-9 CM codes for identifying quality measures (Davis, 1995; Iezzoni, 1997; Iezzoni et al., 1994). While secondary ICD-9 CM codes may be an efficient method of identifying FTR rates for hospitals, the lack of sensitivity, even in combination with clinical predictors, does not provide a sufficient level of accuracy for use in further study of nursing care.

5.6 LIMITATIONS

The primary study limitation was the use of VHA as the setting. The integrated electronic medical record allowed for not only efficiency of record review but also a homogenous population. Conversely, the sample size was predominately male (over 97%) which did not allow for evaluation of differences in FTR between genders. Also, the number of records with surgical care was limited and did not allow for comparisons to other findings in the literature. As VHA puts less emphasis on ICD-9 CM codes for billing purposes than the private sector, there may be significant differences in these codes that were not examined in this study.

Bias during record review is another possible limitation. Although efforts were made to control for bias by using two independent reviewers who were not affiliated with the hospitals and a standardized tool, expert opinion is always subjective.

Only records where the patient died were reviewed. Therefore, comparisons between patients with similar secondary ICD-9 CM codes and clinical predictors who did not die with those who did die could not be made. Analysis of these records might provide important information in the processes of care that prevented failures and should be considered for future study.

5.7 SUMMARY

Evidence supports FTR as an important indicator of quality for hospitals. The original 15 secondary ICD-9 CM codes that were confirmed by record review were only studied in elective surgical patients. Subsequent study expanded this work to medical patients, selecting five of the 15 codes that were thought to be sensitive to nursing care. The AHRQ continued this work by providing software with these five codes with the addition of codes for acute renal failure. A strong body of evidence indicates that nursing characteristics, such as staffing and practice environment, influence FTR.

Three additional studies found poor relationships between the AHRQ ICD-9 CM codes and FTR determined by expert opinion through record review. To address these findings, AHRQ amended the software first to exclude diagnoses that were present on admissions and later to modified FTR into two separate indicators: death among surgical patients and death in low mortality DRGs. Effectiveness of these changes has not been reported in the literature.

In the current study, FTR determined by expert opinion through record review was compared to FTR using the AHRQ secondary ICD-9 CM codes. Poor sensitivity and specificity was found although diagnostic accuracy improved when the age restrict was removed. Specific clinical predictors did outperform the secondary ICD-9 CM codes, particularly when used in specific combinations. Transfer to a higher level of care with a variety of different clinical predictors consistently performed well as did the individual predictor of complications following a procedure. The sample of surgical records was not sufficient to draw conclusions. Combining clinical predictors with secondary ICD-9 CM codes did not enhance diagnostic accuracy.

Although many of the clinical predictors lacked specificity, sensitivity has greater clinical importance. The development of electronic alerts, particularly transfer to a higher level of care and complication following a procedure, may provide an increased level of surveillance for these patients. Intensive monitoring may decrease the risk of the patient for failures in care. Decreases in hemoglobin/hematocrit, increases in creatinine, transfusion of blood products, intubation/reintubation, hypotension/sedation should also receive a higher level of surveillance. While patients who exhibit these predictors may not necessarily result in failures in care (low specificity), the potential is great enough (high specificity) that added scrutiny is warranted. These findings may not generalize to private sector hospital care and further study is recommended. Record review of all patients, not just those who died, may provide additional information on processes of care that help rescue patients and should be investigated.

REFERENCES

. Adverse Events in Hospitalized Patients. (2010): Office of the Inspector General Department of Health and Human Services.

Aiken, L. H., clarke, S. P., Cheung, R. B., Sloane, D. M., & Silber, J. H. (2003). Educational levels of hospital nurses and surgical patient mortality. *Journal of the American Medical Association*, *290*(12), 1617-1623.

Aiken, L. H., Clarke, S. P., Sloane, D. M., Lake, E. T., & Cheney, T. (2008). Effects of hospital care environment on patient mortality and nurse outcomes. *Journal of Nursing Administration*, *38*(5), 223-229.

Aiken, L. H., Clarke, S. P., Sloane, D. M., Sochalski, J., & Silber, J. H. (2002). Hospital nurse staffing and patient mortality, nurse burnout, and job dissatisfaction. *Journal of the American Medical Association, 288*, 1987.

Bobay, K. L., Fiorelli, K.L., & Anderson, A.J. (2008). Failure-to-rescue: a preliminary study of patient level factors. *Journal of Nursing Care Quality, 23*(3), 211-215.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., . . . Lijmer, J. G. (2003). The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. *Annals of Internal Medicine*, *138*(1), w1-w12.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., . . . CW de Vet, H. (2004). Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Family Practice, 21*, 4-10.

Boyle, S. M. (2004). Nursing unit characteristics and patient outcomes. *Nursing Economic$, 22*(3), 111.

Buist, M., Bernard, S., Nguyen, T., Moore, G., & Anderson, J. (2004). Association between clinically abnormal observations and subsequent in-hospital mortality. *Resuscitation, 62*, 137-141.

Cei, M., Bartolomei, C., & Mumoli, N. (2009). In-hospital mortality and morbidity of elderly medical patients can be predicted at admission by the Modified Early Warning Score: a prospective study. *The International Journal of Clinical Practice, 63*(4), 591-595.

Coffey, R. M., Andrews, R. M., & Moy, E. (2005). Racial, ethnic, and socioeconomic disparities in estimates of AHRQ PSI. *Medical Care, 43*, SI 48-57.

Davis, R. B., Iezzoni, L.I., Phillips, R.S., Reilly, P., Coffman, G.A., & Safran, C. (1995). Predicting in-hospital morality. The importance of functional status information. *Medical Care, 33*(9), 906-921.

DeCoster, C., Quan, H., Finlayson, A., Gao, M., Halfon, P., Humphries, K.H., Johansen, P., Lik, L.M., Luthe, J.C., Romano, P.S., Roos, L., Sundarajan, V., Tu, J.V., Webster, G., & Ghali, W.A. (2006). Identifying priorities in methodological research using ICD-9 CM and ID-10 administrative data. Report from an international consortium. *British Medical Journal, 6*, 77.

deWet, C., & Bowie, P. (2009). The preliminary development and testing of a global trigger tool to detect error and patient harm in primary-care recoreds. *Postgraduate Medicine, 85*(1002), 176-180.

Donabedian, A. (2005a). Evaluating the quality of medical care. *Millbank Quarterly*, *83*(4), 691-729.

Donabedian, A. (2005b). Evaluating the quality of medical care. *The Millbank Quarterly, 83*(4), 691-729.

Flahaut, A., Cadillhac, M., & Thomas, G. (2005). Sample size calculations should be performed for design accuracy in diagnostic test studies. *Journal of Clinical Epidemiology, 58*, 859-862.

Fleisher, L. A., Pasternak, L. R., Herbert, R., & Anderson, G. F. (2004). Inpatient hospital admission and death after outpatient surgery in elderly patients: importance of patient and system characteristics and location of care. *Archieves of Surgery, 139*, 67-72.

Friese, C. R. (2005). *Hospital nursing organization and outcomes for surgical oncology patients.* (Ph.D.), University of Pennsylvania. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=rzh&AN=2009283247&site=ehost-live Available from EBSCOhost rzh database.

Friese, C. R., Lake, E. T., Aiken, L. H., Silber, J. H., & Sochalski, J. (2008). Hospital nurse practice environments and outcomes for surgical oncology patients. *Health Services Research, 43*(4), 1145-1163.

Gawande, A. A., Thomas, E. J., Zinner, M. J., & Brennan, T. A. (1999). The incidence of nature of surgical adverse events in Colorado and Utah in 1992. *Surgery, 126*, 66-75.

Goldman, E. S., Easterling, J., & Sheiner, L. B. (1989). Improving the homogeneity of diagnosis-related groups (DRGs) by using clinical laboratory, demographic, and discharge data. *The American Journal for Public Health, 79*(4), 441-445.

Green, J., Passman, L. J., & Winfield, N. (1991). Analyzing hospital mortality: the consequences of diversity in patient mix. *The Journal of the American Medical Association, 265*(1849-1853).

Green, J., Winfield, N., & Sharkey, P. (1990). The importance of severity of illness in assessing hospital mortality. *The Journal of the American Medical Association, 263*, 241-249.

Griffin, F. A., & Resar, R. K. (2007). IHI Global Trigger Tool for measuring adverse events *IHI Innovative Services White Paper*. Cambridge: Institute for Healthcare Improvement.

Griner, P. F., Mayerwski, R. J., Mushin, A. I., & Greenland, P. (1981). Selection and interpretation of diagnostic tests and procedures: Principles and applications. *Annals of Internal Medicine, 94*, 577-592.

Halfon, P., Eggli, Y., Pretre-Rohrbach, I., Meylan, D., Marazzi, A., & Burnand, B. (2006). Validation of the potentially avoidable hospital readmission rate as a routine indicator of the quality of hospital care. *Medical Care, 44*(11), 972-981.

Horowitz, L. I., Cuny, J. F., Cerese, J., & Krumholz, H. M. (2007). Failure to Rescue: Validation of an algorithm using administrative data. *Medical Care, 45*, 283-287.

Iezzoni, L. I. (1997). Assessing quality using administrative data. *Annals of Internal Medicine, 127*(8), 666-674.

Iezzoni, L. I., Daley, J., Heeren, T., Foley, S. M., Fisher, E. S., Duncan, C., . . . Coffman, G. A. (1994). Identifying complications of care using administrative data. *Medical Care, 32*, 700-715.

Iezzoni, L. I., Daley, J., Heeren, T., Foley, S.M., Fisher, E.S., Duncan, C., Hughes, J.S., & Coffman, G.A. (1994). Identifying complications of care using administrative data. *Medical Care, 32*, 700-715.

Iezzoni, L. I., Schwartz, M., Ash, A. S., & Mackieman, Y. D. (1995). Using severity measures to predict the likelihood of death for pneumonia inpatients. *Journal of General Internal Medicine, 11*(1), 23-31.

Kashner, M. T. (1998). Agreement between administrative files and written medical recores: A case of the Department of Veterans Affairs. *Medical Care, 36*, 1324-1336.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (1999). *To Err is Human*. Washington, DC: National Academy Press.

Konety, S. H., Vaughan Sarrazin, M. S., & Rosenthal, G. E. (2005). Patient and hospital differences underlying racial variation in outcomes after coronary artery bypass graft surgery. *Circulation, 111*, 1210-1216.

Leape, L. L. (2002). What practices will most improve safety? Evidence-based medicine meets patient safety. *Journal of the American Medical Association, 288*(4), 501-507.

Leape, L. L., Brennan, T. A., Laird, N., Lawthers, A. G., Localio, A. R., Barnes, B. A., . . . Hiatt, H. (1991). The nature of adverse events in hospitalized patients: Results of the Harvard Medical Practie Study II. *New England Journal of Medicine, 324*(6), 377-384.

Lederer, D. J., Caplan-Shaw, C. E., & O'Shea, M. K. (2006). Racial and ethnic disparities in survival in lung transplant candidates with idiopathic pulmonary fibrosis. *American Journal Of Transplantation, 6*, 398-403.

Lee, L. L. Y., Yeung, K. L., Lo, W. Y. L., Lau, Y. S. C., Tang, S. Y. H., & Chan, J. T. S. (2008). Evaluation of a simplified therapeutic intervention scoring system (TISS-28) and the modified early warning score (MEWS) in predicting physiological deterioration during inter-facility transport. *Resuscitation, 76*, 47-51.

McCarthy, E. P., Iezzoni, L.I., Davis, R.B., Palmer, R.H., Calahane, M., Hamel, M.B., Mukamal, K., Phillips, R.S., & Davies, D.T. (2000). Does clinical evidence support ICD-9 CM diagnosis coding of complications? *Medical Care, 38*, 868-876.

Michel, P., Quenon, J.L., de Sarasqueta, A.M., & Scemama, O. (2004). Comparison of three methods for estimating rates of adverse events and rates of preventable adverse events in acute care hospitals. *British Medical Journal, 328*, 1-5.

Moon, A., Cosgrove, J. R., Lea, D., Fairs, A., & Cressy, D. M. (2011). An eight year audit before and after the introduction of modified early warning score (MEWS) charts, of patients admitted to a teritary reefrral intensive care unit after CPR. *Resuscitation, 82*, 150-154.

Needleman, J., Buerhaus, P., Mattke, S., Stewart, M., & Zelevinsky, K. (2002). Nurse staffing and quality of care in hospitals in the United States. *New England Journal of Medicine, 346*, 1715-1722.

Needleman, J., Buerhaus, P. I., Pankratz, V. S., Leibson, C. L., Stevens, S. R., & Harris, M. (2011). Nurse staffing and inpatient hospital mortality. *New England Journal of Medicine, 364*(11), 1037-1045.

Overview: Patient safety indicators from the Agency of Healthcare Quality and Research. (2007). Retrieved April 12, 2012

. Patient Safety Indicators Technical Manual Version 4.2. (2010) (Vol. 4.2). Washington, DC: Agency for Healthcare Research and Quality.

Pine, M., Noursis, M., Jones, B., & Rosenthal, G.E. (1997). Predictions of hospital mortality rates: a comparison of data sources. *Annals of Internal Medicine, 126*(5), 347-354.

PSI Technical Specifications. (2007). Retrieved February 4, 2013

Romano, P. S. (2008). *Lessons from the AHRQ PSI validation project*. Paper presented at the AHRQ QI Users Meeting, Washington, DC.

Romano, P. S., Geppert, J. S., Davies, S., Miller, M. R., Elixhauser, A., & McDonald, K. M. (2003). A national profile of patient safety in U.S. hospitals. *Health Affairs, 22*(2), 154-166.

Romano, P. S., Mull, H. J., Rivard, P., e, Shibei, A., Henderson, W. G., Loveland, S., . . . Rosen, A. K. (2009). Validity of selected AHRQ patient safety indicators based on VA National Surgical Quality Improvement program data. *Health Services Research, 44*(1), 182-204.

Shever, L. L. (2007). *Predictors of failure to rescue and failure to rescue from an adverse incident with closer examination of nursing surveillance.* (Ph.D.), University of Iowa. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=rzh&AN=2010060648&site=ehost-live Available from EBSCOhost rzh database.

Shojania, K. G., Duncan, B., W., McDonald, K. M., Wachter, R. M., & Markowitz, A. J. (2001). Making health care safer: a critical analysis of patient safety practices (Vol. 43, pp. 1-668): Evidence Report Technology Assessment.

Silber, J. H. (1998). Modifying the failure-to-rescue rate when detailed complication data is not available: a US-Canadian cross national study. *Health Services Research, 15*, 102-103.

Silber, J. H., Romano, P. S., Rosen, A. K., Wang, Y., Even-Shoshan, O., & Volpp, K. G. (2007). Failure-to-rescue: comparing definitions to measure quality of care. *Medical Care, 45*(10), 918-925.

Silber, J. H., Rosenbaum, P. R., Schwartz, S., Ross, R. N., & Williams, S. V. (1995). Evaluation of the complication rates as a measure of quality of care in coronary artery bypass surgery. *Journal of the American Medical Association*, *274*, 317-322.

Silber, J. H., Williams, S. V., Krakauer, H., & Schwartz, J. S. (1992). Hospitals and patient characteristics associated with death after surgery: A study of adverse occurences and failure-to-rescue. *Medical Care, 30*, 615-622.

Talsma, A., Bahl, V., & Campbell, D. A. (2008). Exploratory analyses of the "failure to rescue" measure: evaluation through medical record review. *Journal of Nursing Care Quality, 23*(3), 202-210.

Talsma, A., Bahl, V., & Campbell, D.A. (2008). Exploratory analyses of the "failure to rescue" measure evaluation through medical record review. *Journal of Nursing Care Quality, 23*(3), 202-210.

Trivedi, A. N., Sequist, T. D., & Ayanian, J. Z. (2006). Impact of hospital volume on racial disparities in cardiovascular procedure mortality. *Journal of the American College of Cardiology*, *47*, 417-424.

Yen, M., & Lo, L. H. (2004). A model for testing the results of nursing care and patient outcomes. *Nursing Economic$, 22*(3), 75-80.

APPENDIX A

AHRQ INCLUSION AND EXCLUSION ICD-9 CM CODES

Table A.1 AHRQ Inclusion and Exclusion ICD-9 CM Codes

	Inclusion criteria	**Exclusion criteria**
General Structure	All discharges with a disposition of "deceased"	Exclusion noted for each complication of care as specified in each row below
1. shock and cardiac arrest	**Secondary diagnosis codes:** 4275 6395 7855, 50, 51, 52, 59 7991 9950, 4 9980 9994 **Secondary procedure codes:** 9393, 9960, 9963	*Hemorrhage Diagnosis Codes* 4590, 9582, 99811 2800, 2851, 291, 303, 425 *GI Hemorrhage Diagnosis Codes* 4560, 45620, 5307,53082 53100, 01, 20, 21, 40, 41, 60, 61 53200,01, 53220,21,40,41,60, 61 53300, 01, 20, 21,40,41, 60, 61 53400, 01, 20, 21, 40, 41, 60, 61 53501, 11, 21, 30, 31, 41, 51, 61; 53783, 84 56202, 03, 12, 13; 5693, 85, 86 5710,1,2, 3; 5780, 1, 9 9800, 09 *Trauma Diagnosis Codes* 800 - 825, 827-833, 835- 839 850, 11, 12; 851-854 860- 884, 887, 890- 892, 894, 896, 897 900 – 904, 925- 929, 940- 949, 952-953, 958 *Trauma DRGs* 002, 027- 033, 072, 083, 084, 235-237, 440- 446, 456 – 457, 459, 484 – 487, 491, 504- 511

	Inclusion criteria	Exclusion criteria
2. Pneumonia	**Secondary diagnosis codes:** 4820, 1, 2, 3, 30, 31, 32, 39, 4, 40 41, 49, 8, 81, 82, 83, 84, 89, 9 485, 486, 5070, 514	**Any diagnosis codes:** *1) Viral Pneumonia* 480.0, 480.1, 480.2, 480.8, 480.9, 483, 483.0, 483.1, 483.8, 484.1, 484.3, 484.5, 484.6, 484.7, 484.8, 487.0, 487.1, 487.8 *2) Immuncompromised State* 042, 136.3, 279.00, 279.01, 279.02, 279.03, 279.04, 279.05, 279.06, 279.09, 279.10, 279.11, 279.12, 279.13, 279.19, 279.2, 279.3, 279.4, 279.8, 279.9
3. Deep Vein Thrombosis/ Pulmonary Embolism	**Secondary diagnosis codes:** 4151, 11, 19 45111, 19, 2, 81, 9 45340, 41, 42, 4538, 4539	

	Inclusion criteria	**Exclusion criteria**
4. Sepsis	**Secondary diagnosis codes:** 0380, 1, 10, 11, 19 0382, 0383 03840, 41, 42, 43, 49 0388, 0389 78552, 59 99592, 9980	*2) Immuncompromised State* 042, 136.3, 279.00, 279.01, 279.02, 279.03, 279.04, 279.05, 279.06, 279.09, 279.10, 279.11, 279.12, 279.13, 279.19, 279.2, 279.3, 279.4, 279.8, 279.9 *Infection Diagnosis Codes:* 0010 , 11, 19 0020, 1, 2, 3, 9 0030, 1, 20-24, 29 0038 -39 0040-0043, 0048, 0049 0050-0054, 00581, 0059 00800, 01-04, 09, 0081, 0082 *Infection DRGs* 020, 068, 069, 070, 079, 080, 081, 089, 090, 091, 126, 238, 242, 277, 278, 279, 320 321, 322, 368

	Inclusion criteria	*Exclusion criteria*
5. GI hemorrhage/ Acute Ulcer	**Secondary diagnosis codes:** 4560,20 5307, 82 53100, 01, 10, 11, 20, 21, 30, 31, 90, 91 53200, 01, 10, 11, 20, 21, 30, 31, 90, 91 53300, 01, 10, 11, 20, 21, 30, 31, 90, 91 53400, 01, 10, 11, 20, 21, 30, 31, 90, 91 53501, 11,21,31,41, **Secondary procedure codes:** 4995	*1) Trauma as defined by principal diagnoses* 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 817, 818, 819, 820, 821, 822, 823, 824, 825, 827, 828, 829, 830, 831, 832, 833, 835, 836, 837, 838, 839, 850, 851, 852, 853, 854, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 884, 887, 890, 891, 892, 894, 896, 897, 900, 901, 902, 903, 904, 925, 926, 927, 928, 929, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 952, 953, 958 *2) Trauma DRGs* 002, 027, 028, 029, 031, 032, 072, 083, 084, 235, 236, 237, 440, 441, 442, 443, 444, 445, 446, 456, 457, 458, 459, 460, 484, 485, 486, 487, 491, 504, 505, 506, 507, 508, 509, 510, 511 *3) History of alcoholism defined as secondary diagnosis* 2910-5, 29181, 29189, 2919, 30300-3, 30390-2, 30500-2 **Principal procedure codes:** 444, 4440-2 if secondary diagnoses 5780-1, 9

	Inclusion criteria	*Exclusion criteria*
6. Acute Renal Failure	**Secondary diagnosis codes:** 5845-9, 6393 66930-34	Comorbidity of renal failure defined as any of the following diagnoses 4275 *Hemorrhage Diagnosis Codes* 4590 9582, 99811 2800. 2851, 291 303, 425 *GI Hemorrhage Diagnosis Codes* 4560, 45620 5307,53082 53100, 01, 20, 21, 40, 41, 60, 61 53200,01 53220,21,40,41,60, 61 53300, 01, 20, 21,40,41, 60, 61 53400, 01, 20, 21, 40, 41, 60, 61 53501, 11, 21, 30, 31, 41, 51, 61 53783, 84 56202, 03, 12, 13 5693, 85, 86 5710,1,2, 3 5780, 1, 9 9800, 09 *Shock diagnosis codes* 63450, 51, 52 63550, 51, 52 63650, 51, 52

		63750, 51, 52 6385, 6395 66910, 11, 12, 13, 14 7855, 50, 51, 52, 59 9950, 4 9980, 9994 *Trauma Diagnosis Codes* 800 - 825, 827-833, 835- 839 850, 11, 12 851-854 860- 884, 887, 890- 892+, 894, 896, 897 900 – 904, 925- 929, 940- 949, 952-953, 958 *Trauma DRGs* 002, 027- 033, 072, 083, 084, 235-237, 440- 446, 456 – 457, 459, 484 – 487, 491, 504- 511 *Acute Myocardial Infarction Codes:* 41000, 01, 10, 11, 20, 21, 30, 31, 40, 41, 50, 51, 60, 61, 70, 71, 80, 81, 90, 91 *Cardiac Arrhythmia Codes:* 4260, 4270,1,2, 42731, 32, 41, 42, 9

APPENDIX B

DEMOGRAPHIC PREDICTOR DEFINITIONS

Table A.2 Demographic Predictor Definitions

Predictor	Definition	Source of Definition
Age	The age classification is based on the age of the person in complete years derived from their date of birth information.	Census 2000. Summary File 3 prepared by the U.S. Census Bureau, 2002
Sex/Gender	Either ‘‘male’’ or ‘‘female’’	Census 2000. Summary File 3 prepared by the U.S. Census Bureau, 2002
Race/Ethnicity	The five minimum race categories are American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White.	http://www.whiteh ouse.gov/omb/bul letins/b00-02.html, March, 2007
Length of Stay	Same-day stays are therefore coded as 0.	ahrq.gov/db/va rs/sasddistnote.js p?los_x
Body Mass Index	Body Mass Index (BMI) is a number calculated from a person's weight and height. BMI is a fairly reliable indicator of body fatness for most people.	http://www.cdc.gov/healthyw eight/assessing/bmi/adult_bmi /#Definition
Number of transfers during the total LOS	Number of times a patient is transferred from one patient care unit to another during the total length of stay. Transfers from the Emergency Department to an inpatient care unit will be counted as one inter-hospital transfer.	(Shever, 2007)

APPENDIX C

SENSITIVITY AND SPECIFICITY OF CLINICAL PREDICTORS

Table A.3 Sensitivity and Specificity of Clinical Predictors

Sensitivity and Specificity of Clinical Predictors at the 95% CI				
Clinical Predictor	Sensitivity	95% CI	Specificity	95% CI
Positive blood cultures	17.7%	0.14-0.21	81.5%	0.78-0.85
Cultures positive for infection	59.2%	0.55-0.64	26.3%	0.22-0.30
PTT > 100	30.9%	0.26-0.35	80.2%	0.76-0.84
Glucose < 50 mg/ml	17.2%	0.14-0.21	82.7%	0.79-0.86
BUN > 2X baseline	60.6%	0.56-0.65	33.3%	0.29-0.38
Creatinine > 2X baseline	49.3%	0.44-0.54	49.5%	0.45-0.55
25% drop in hemoglobin/ hematocrit	50.6%	0.46-0.55	70.2%	0.66-0.75
BNP >100	30.8%	0.27-0.35	65.1%	0.61-0.70
Transfusion of blood products	63.0%	0.59-0.67	50.7%	0.46-0.55
Negative pathology report	36.4%	0.28-0.42	80%	0.73-0.87
Unit level procedure	87.7%	0.84-0.90	25.3%	0.21-0.29
Any complication with procedure	65.4%	0.61-0.70	83.2%	0.80-0.87
x-ray for PE or DVT	16.5%	0.13-0.20	82.2%	0.34-0.40
Chest x-ray positive for pneumonia	64.6%	0.60-0.69	38.5%	0.34-0.40
Vitamin K administration	16.3%	0.13-0.20	85.7%	0.82-0.89

Sensitivity and Specificity of Clinical Predictors at the 95% CI				
Narcan administration	10%	0.07-0.13	96.3%	0.95-0.98
Promethazine adminstration	10%	0.07-0.13	92.4%	0.90-0.95
Abrupt stop in medication	11.3%	0.08-0.14	97.2%	0.96-0.98
Dialysis	13.8%	0.11-0.17	88.5%	0.85-0.91
Fall	8.8%	0.06-0.11	94.9%	0.93-0.97
Restraint use	30.0%	0.26-0.34	82.3%	0.79-0.86
Stroke	6.3%	0.04-0.09	98.0%	0.97-0.99
Transfer to a higher level of care	51.3%	0.47-0.56	63.2%	0.59-0.67
Readmission to the ICU	22%	0.18-0.26	91.9%	0.89-0.95
Oversedation/ hypotension	56.3%	0.52-0.61	52.5%	0.48-0.57
Return to surgery	12.5%	0.09-0.16	98.9%	0.98-0.99
Change in procedure	6.3%	0.04-0.09	100%	1.00
Mechanical ventilation > 24 hours	10%	0.07-0.13	98.0%	0.97-0.99
OR time > 6 hours	3.8%	0.02-0.06	100%	1.00
Organ removal or repair	10%	0.07-0.13	97.2%	0.96-0.98
Consult in the PACU	2.5%	0.01-0.04	100%	1.00
Intraoperative epinephrine	3.8%	0.02-0.06	99.7%	0.99-1.00