Developing a Predictive Model for Hospital-Acquired Catheter-Associated Urinary Tract Infections Using Electronic Health Records and Nurse Staffing Data

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Jung In Park

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Bonnie L. Westra, Advisor

August, 2016

Acknowledgements

The author acknowledges the support of the Beatrice L. Witt Endowment Fund, Katharine Densford Dreves Scholarship, Lucille Casas Paradela International Fellowship Fund, Connie W. Delaney Fellowship in Nursing Innovation, Violet A. Shea Nursing Scholarship, Nursing Research Fellowship Fund, and the University of Minnesota, School of Nursing for completion of academic work and writing of this dissertation.

Dedication

I am grateful to the members of my dissertation committee for reviewing my work and sharing their thoughts and suggestions:

Bonnie Westra

Connie Delaney

Donna Bliss

Chih-Lin Chi

Special gratitude and appreciation go to Drs. Westra and Delaney for their devoted help in preparing this paper. I am thankful for their immeasurable time and patience in route to publication. I have been fortunate to have learned from the great pioneers of this era. Their passion, excellence, and wisdom as scholars have inspired me greatly.

I would like to dedicate this dissertation to my dear parents, my brother, and my beloved husband who supported and encouraged me along the way. Thanks for trusting me, loving me, and being there for me. I am also thankful to my colleagues and friends. I could not have done this without the help of my people.

Abstract

There are a number of clinical guidelines and studies about hospital-acquired catheter-associated urinary tract infections (CAUTIs), but the rate of CAUTI occurrence is still rising. Hospitals are focusing on preventing hospital-acquired CAUTI, as the Centers for Medicare and Medicaid Services (CMS) does not provide payment for hospital-acquired infections anymore. There is a need to explore additional factors associated with hospital-acquired CAUTI and develop a predictive model to detect patients at high risk.

This study developed a predictive model for hospital-acquired CAUTIs using electronic health records (EHRs) and nurse staffing data from multiple data sources. Research using large amounts of data could provide additional knowledge about hospital-acquired CAUTI. The first aim of the study was to create a quality, de-identified dataset combining multiple data sources for machine learning tasks. To address the first aim of the study, three datasets were combined into a single dataset. After integrating the datasets, data were cleaned and prepared for analysis. The second aim of the study was to develop and evaluate predictive models to find the best predictive model for hospital-acquired CAUTI. For the second aim of the study, three predictive models were created using the following data mining method: decision trees (DT), logistic regression (LR), and support vector machine (SVM). The models were evaluated and DT model was determined as the best predictive model for hospital-acquired CAUTI.

The findings from this study have presented factors associated with hospital-acquired CAUTI. The study results demonstrated that female gender, old adult (≥56),

Charlson comorbidity index score ≥ 3, longer length of stay, glucose lab result > 200 mg/dl, present of rationale for continued use of catheter, higher percent of direct care RNs with associate's degree in nursing, less total nursing hours per patient day, and lower percent of direct care RNs with specialty nursing certification was related to CAUTI occurrence.

Implications for future research include the use of different analytic software to investigate detailed results for LR model, adding more factors associated with CAUTI in modeling, using a larger sample with more patients with CAUTI, and patient outcomes research using nursing-sensitive indicators. This study has important implications for nursing practice. According to the study results, nurse specialty certification, nurse's education at the baccalaureate level or higher, and more nursing hours per patient day were associated with better patient outcomes. Therefore, considerable efforts are needed to promote possession of nurse specialty certification and higher level of nursing education, as well as enough supply of nursing workforce.

Table of Contents

List of Tables

List of Figures

CHAPTER I

INTRODUCTION

In the United States, health care-associated infections (HAIs) are a major threat to patient safety (Kohn, Corrigan, & Donaldson, 2010). HAIs are defined as infections that patients acquire in a health care facility while receiving health care treatment for other conditions (CDPH, 2016). Specifically, hospital-acquired HAIs are called *nosocomial infections*. Nosocomial infections are closely related to patient outcomes such as morbidity and mortality, and often result in higher costs of treatment. Approximately 2 million patients in the United States have HAIs, and about 90,000 are estimated to die from an HAI annually (Stone, 2009). There were approximately 722,000 HAIs in U.S. hospitals in 2011, and about 75,000 patients with HAIs died during hospitalizations (Magill et al., 2014; Shang, Stone, & Larson, 2015). There is also a substantial economic impact by nosocomial infection with financial burden estimated $28 - $45 billion per year (Scott, 2009).

The Office of Disease Prevention and Health Promotion (ODPHP) reported that the four major nosocomial infections are urinary tract infections (UTIs), surgical site infections, bloodstream infections, and pneumonia (ODPHP, 2016). According to a recent study, nosocomial UTIs are the most common type of HAIs reported to the National Healthcare Safety Network (NHSN) and account for up to 40% of nosocomial infections and 23% of infections in intensive care units (ICUs) (Chenoweth & Saint, 2013; Lo et al., 2008). In the United States, UTIs are responsible for 7 million clinic visits annually, with costs exceeding $1.6 billion (Sheerin, 2011). UTIs are developed by such microbes as fungi, viruses, and bacteria (NIH, 2012). The term *UTI* is applied to any infection in the urinary tract and presence of bacteria in the urine (Lee & Neild, 2007). The most common cause of infection is Escherichia coli.

A large portion of hospital-acquired UTIs are related to the use of indwelling urinary catheters (Chenoweth & Saint, 2013; Saint et al., 2008). At some point during hospitalization, 12- 16% percent of adult hospitalized patients have an indwelling urinary catheter (Lo et al., 2014). The Centers for Disease Control and Prevention (CDC) estimated the number of catheter-associated UTIs (CAUTIs) is 561,677 and the number of deaths caused by CAUTIs is 8,205 annually (Stone, 2009). In 2007, 139,000 CAUTIs occurred in U.S. hospitals and resulted in about $ 131 million in direct medical costs nationwide annually (CDC/NHSN, 2016; Chenoweth & Saint, 2013; Gokula, Smith, & Hickner, 2007). Some researchers have argued that CAUTI is the most common type of HAI, accounting for more than 30% of nosocomial infections (Alexaitis & Broome, 2014; Klevens et al., 2007).

However, a 2015 report from the CDC showed that CAUTI incidents have been increased by 6% between 2009 and 2013 (CDC, 2015a). Although many studies during the past decades identified risk factors and strategies to lower CAUTI incidents, it is still a prevalent nosocomial infection (Tambyah, 2015). A significant number of nosocomial infections are preventable, including CAUTIs (Krein, Kowalski, Hofer, & Saint, 2012; Stone, 2009). The National Quality Forum (NQF) emphasized preventing HAIs as a key focus for patient safety (NQF, 2011).

A recent incentive program driven by the Centers for Medicare and Medicaid

Services (CMS) for adoption of certified electronic health records (EHRs) and meaningful use has led to wide implementation of EHRs. Starting in 2008, the CMS no longer reimburses hospitals for hospital-acquired conditions such as CAUTI, which has resulted in hospitals being more cautious about HAIs and focusing on prevention. Implementation and use of EHRs enabled collecting vast amounts of data in a standardized format; the amount of data has allowed conducting large-scale research for prevention recommendations, predictive studies, and clinical reports.

Significance of CAUTI

The CDC estimated that 75% of UTIs are attributable to the use of indwelling urinary catheters and 95% of UTIs occurred in ICUs (CDC/NHSN, 2016; Chenoweth & Saint, 2013; Nicolle, 2012). The CDC defines an *indwelling urinary catheter* as a drainage tube that is inserted into the urinary bladder through the urethra, left in place, and connected to a closed collection system (CDC/NHSN, 2016). The use of an indwelling urinary catheter is often essential for inpatient care. However, the use of an indwelling urinary catheter leads to a critical risk of CAUTIs. It has been reported that CAUTIs are associated with increased morbidity, mortality, health care costs, and length of hospital stay (Burton, Edwards, Srinivasan, Fridkin, & Gould, 2011). A CAUTI occurs when bacteria enter the urinary tract through the urinary catheter and cause an infection. Most bacteria causing CAUTI enter the bladder by ascending the urethra from the perineum (Chenoweth & Saint, 2011). Most CAUTIs are asymptomatic; but if they are symptomatic, the infection can lead to bladder discomfort, dysuria, and fever (Palmer, Lee, Dutta-Linn, Wroe, & Hartmann, 2013).

Clinical guidelines for CAUTI recommend that catheters should be used only when needed and removed as soon as possible; when placed, sterile equipment and proper aseptic technique should be used by trained personnel, and a sterile, closed, unobstructed drainage system should be always maintained (CDC, 2015a). Some studies reported that nurse-driven catheter interventions and education of staff on CAUTI can reduce the duration of urinary catheterization or incidence of CAUTI (Fuchs, Sexton, Thornlow, & Champagne, 2011; Oman et al., 2012; Topal et al., 2005).

The duration of the catheterization period is the most dominant related factor for CAUTI. The duration of catheterization should be minimized especially for those who are at high risk, such as women, elderly persons, and patients with impaired immunity (Gould et al., 2010). If the catheteriation period is too long, CAUTIs are inevitable because biofilm formations from microbes occur and these are less susceptible to antibiotics (Lo et al., 2008). Bacteriuria develops quickly at an average daily rate of 5% per day of catheterization (Chenoweth & Saint, 2013). After 17 days of catheterization, 90% of patients have bacteriuria (Lee & Neild, 2007). If an indwelling urinary catheter has been in place for longer than 2 weeks at the onset of the UTI and is still indicated, the catheter should be replaced to reduce the risk of subsequent catheter-associated infection (Hooton et al., 2010).

Other risk factors associated with CAUTI are female gender, severe underlying illness, nonsurgical disease, being over 50 years old, diabetes mellitus, serum creatinine level greater than 2 mg/dL, duration of catheterization, adherence to aseptic catheter care, catheter insertion after the sixth day of hospitalization, and catheter insertion outside the operating room (Chenoweth & Saint, 2013).

Appropriate indications for indwelling urinary catheter use include acute anatomic or functional urinary retention or obstruction, urinary incontinence in the setting of open perineal or sacral wounds, perioperative use for selected surgical procedures, accurate monitoring of urine output of unstable patients, improving comfort for end-of-life care, or patient preference (Gould et al., 2010; Saint & Chenoweth, 2003). The 2009 Infectious Diseases Society of America (IDSA) guidelines for CAUTIs stated that an indwelling catheter may be used at the patient's request in exceptional cases and when other approaches to incontinence management have been ineffective (Hooton et al., 2010).

Hospital-acquired CAUTIs has been targeted as a 'never event' for complete elimination (Cardo et al., 2010). If a CAUTI diagnosis code does not have a "present-on-admission" indicator, which means the patient did not have a CAUTI when admitted but acquired it during hospitalization, then CMS does not provide payment for that. Therefore, hospitals face the challenge of implementing evidence-based guidelines to prevent hospital-acquired CAUTIs. Prevention has become a priority for most hospitals, as 65 - 70% of CAUTIs are estimated to be preventable (Umscheid et al., 2011).

Nursing-Sensitive Indicators and Nurse-Staffing Data

The American Nurses Association (ANA) stated that nurses are the largest direct health care workforce affecting patient outcomes (Montalvo, 2008). A CAUTI is one of the nursing-sensitive outcomes defined by the ANA. *Nursing-sensitive outcomes* are defined as changes in health status when nursing care has had a direct influence; nursing-sensitive outcome indicators show correlations between nurses' intervention patients have received and how these affect changing health statuses (International Council of Nurses, 2009). Nursing-sensitive indicators are strongly influenced by the care that nurses provide and can effectively demonstrate that nurses make critical, cost-effecting differences in patient's health outcome.

As nurses are frontline bedside care providers and have become the largest segment of the U.S. health care workforce, patient outcomes are directly influenced by nurse staffing levels and nurses' capabilities (Shang et al., 2015; Stone, Pogorzelska, Kunches, & Hirschhorn, 2008). The nursing profession plays an important role in preventing and controlling nosocomial infections including CAUTIs. The quality of nursing care is related to nurse staffing and better nurse staffing is recommended to improve patient outcomes (Kane, Shamliyan, Mueller, Duval, & Wilt, 2007). Several studies have reported that a higher nurse staffing level is related to lower rates of nosocomial infections (Hugonnet, Villaveces, & Pittet, 2007; Jackson, Chiarello, Gaynes, & Gerberding, 2002; Kovner & Gergen, 1998; Lichtig, Knauf, & Milholland, 1999; Manojlovich, Sidani, Covell, & Antonakos, 2011). Additionally, Aiken, Smith, and Lake (1994) found that quality of nursing care contributes to reduction of mortality rate of 4.6%.

In 2004, the NQF endorsed nursing-sensitive care performance measures (Kurtzman & Corrigan, 2007), which include:

- Death Among Surgical Inpatients with Treatable Serious Complications
- Pressure Ulcer Prevalence (Hospital-Acquired)
- Restraint Prevalence (Vest and Limb)

- Patient Falls
- Falls with Injury
- CAUTI Rate for ICU Patients
- Central Line Catheter-Associated Bloodstream Infection (CLABSI) Rate for ICU and Neonatal ICU (NICU) Patients
- Ventilator-Associated Pneumonia (VAP) Rate for ICU and NICU Patients
- Skill Mix (registered nurse [RN], licensed practical nurse [LPN], unlicensed assistive personnel [UAP], and contract)
- Nursing Care Hours per Patient Day
- Voluntary Turnover
- Practice Environment Scale-Nursing Work Index (Joint Commission, 2009)

The ANA established the National Database of Nursing Quality Indicators (NDNQI) to develop and support the implementation of new indicators and methodologies for nursing research (Montalvo, 2007a). The NDNQI is a portable, expandable, and standardized repository (Gallagher & Rowell, 2003). The database has comparative nursing care data with a nationally standardized format and relationships to patient outcomes. Data are collected locally at hospitals and submitted on a quarterly basis to the NDNQI which includes data on NQF-endorsed nursing-sensitive measures, nurse staffing measures, and evaluations of RN job satisfaction and the nursing work environment (Duncan, Montalvo, & Dunton, 2011; NQF, 2004).

Nursing-sensitive indicators in the NDNQI enable researchers to evaluate the structures of care and care processes' both influence on nursing care outcomes. The

NDNQI publishes its annual reporting of structure, process, and outcome indicators on a unit level (Montalvo, 2007). Quality of care is measured using Donabedian's structure-process-outcome framework (ANA, 1999). *Structure* is the setting in which care is provided and includes the number and qualifications of staff, the availability of equipment and supplies, nursing work environment, retention and turnover rate, and staff specialty certification. *Process* refers to the activities of nurses in providing care and encompasses assessment, care planning, medication administration, preventive interventions, evidence-based procedures, education, and nurse-patient interactions. *Outcomes* means changes in patients' health that are attributable to the care they received and include hospital-acquired pressure ulcers, falls, device-associated infections, and peripheral intravenous infiltrations. The use of the NDNQI enables the collection of valid and reliable quality indicators that are nursing-sensitive on a large scale (Weaver, Delaney, Weber, & Carr, 2010).

Nurse staffing factors and RN job satisfaction are essential in preventing CAUTIs because they affect HAIs (Cimiotti, Aiken, Sloane, & Wu, 2012). Unique NDNQI measures other than NQF-endorsed measures include: hospital-acquired pressure ulcer prevalence, RN satisfaction, RN education/certification, completeness of the pediatric pain assessment-intervention-reassessment (AIR) cycle, pediatric peripheral intravenous infiltration rate, and psychiatric physical/sexual assault rate (Montalvo, 2007a). These measures are supported by several studies. RN education levels have an inverse effect on patient outcomes (Aiken, Clarke, Sloane, Lake, & Cheney, 2008; Aiken, Clarke, Cheung, Sloane, & Silber, 2003). Specialty certification refines nursing care by validating consistent standards of care to enhance patient safety and quality of care (Boyle, Cramer, Potter, Gatua, & Stobinski, 2014; Kendall-Gallagher & Blegen, 2009).

Catheter-associated urinary tract infections are one of the NQF-endorsed nursing-sensitive measures because nurses are most often responsible for the management of indwelling urinary catheters. Some studies have found that nurse staffing levels and skill mixes were significantly related to UTIs (Needleman, Buerhaus, Mattke, Stewart, & Zelevinsky, 2002; Unruh, 2003). Therefore, investigating nurse-related factors that influence CAUTI occurrence is necessary. This study uses nurse-staffing measures in the NDNQI to find factors related to CAUTI occurrence in ICUs.

Big Data Research Using Electronic Health Records

According to a systematic review from the Agency for Healthcare Research and Quality (AHRQ), health information technology (HIT) systems, especially EHRs, can enhance guideline-adherent care delivery, improve quality of care through clinical monitoring, and reduce medical error rates in a cost-effective and viable way (Chaudhry et al., 2006; Linder, Ma, Bates, Middleton, & Stafford, 2007). The Institute of Medicine (IOM) 2001 report, “Crossing the Quality Chasm,” emphasized the use of electronic solutions to improve the quality of care delivery (IOM, 2001).

EHRs are defined as “longitudinal electronic records of patient health information generated by one or more encounters in any care delivery setting. Included in this information are patient demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data, and radiology reports” (HIMSS, 2016). Most EHRs include quantitative data (e.g., vital signs), qualitative data (e.g., text data as nursing narratives), and transactional data (e.g., claims) (Murdoch & Detsky, 2013). Nursing documentation also plays an important role for EHRs because nurses continuously document the direct care they provide to patients and use their clinical expertise to assess, plan, intervene, and evaluate their patients (Kelley, Brandon, & Docherty, 2011).

In 2009, the Health Information Technology for Economic and Clinical Health (HITECH) act authorized incentive payments up to $27 billion over 10 years to increase hospital and physician adoption of EHR systems. The intention was to accelerate the adoption of HIT and facilitate the use of qualified EHRs (CMS, 2016). Physicians and hospitals were incentivized if they "meaningfully used" EHRs to achieve significant improvements of quality in care delivery (Blumenthal & Tavenner, 2010; Hsiao, Hing, Socey, & Cai, 2011; Jones, Adams, Schneider, Ringel, & McGlynn, 2010). Most physicians have reported EHR systems improved diagnosis, infection control, and patient care (Shang et al., 2015).

EHRs deliver the functions of traditional health records and other advantages. The EHRs provide standardized data format and terminology, information accessible from multiple sites, efficient communication among health care providers, and alerts and reminders for care quality improvement (Watzlaf, Zeng, Jarymowycz, & Firouzan, 2004). Manyika et al. (2011) described five 'big data levers' that affect providers, payers, and patients in clinical care. The five big data levers include:

- Outcomes-based research including comparative effectiveness research. Outcomes-based research can identify the most effective treatments for patients through the use of big data which include patients' medical information and outcomes of treatments.

- Clinical decision support tools. Clinical decision support tool enables accurate and efficient process, which leads to lower rates of adverse events and medical errors.
- Transparency in medical data. Transparency about medical data and process is important to both providers and patients, because it gives quality and performance data of providers to optimize the caring process, which can also help patients make informed health care decisions.
- Remote patient monitoring. Remote patient monitoring is useful to collect data from long-distance or chronically ill patients, and monitor patients' adherence. This monitoring system can reduce the inpatient hospital days, emergency department visits, and long-term care complications.
- Advanced analytics. Advanced data analytics such as predictive modeling can identify patients at high risk of certain disease using patient profiles.

The adoption of EHRs has enabled easier access and aggregation of clinical data. The amount of collected data through EHR is enormous and storage of data is expanding quickly (Murdoch & Detsky, 2013). Gartner (2016) defines *big data* as "high volume, high velocity, and high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization". Research methods for big data are advancing to convert the vast amount of data into useable information to help generate knowledge. Improvements in data analytic techniques, especially machine learning from computer sciences, have facilitated this new flow dealing with large datasets. Machine learning develops a model for automatic procedure to learn patterns and to predict outcomes based on data (Witten, Frank, & Hall, 2011).

Many techniques support the analysis of big data. Data mining is a set of techniques that discover complex patterns from large datasets. Examples of these techniques include association rule learning, cluster analysis, classification, and regression (Manyika et al., 2011). Natural Language Processing (NLP) is used to analyze natural human language from social media or nursing narratives (Bakken, Hyun, Friedman, & Johnson, 2005). Visualization techniques are used for presentation of images or graphs to understand and interpret the new patterns found from big data analysis (Keim, Qu, & Ma, 2013).

Several studies have indicated a growing need for conducting research using big data collected from clinical care (Hersh, 2007; Safran et al., 2007). The National Institutes of Health (NIH) called for the reuse of EHR data for clinical research (Weiskopf & Weng, 2013). Big data research using EHRs has several advantages: (a) it is useful when generating new insights from large amounts of data; (b) it is helpful to disseminate knowledge through data-driven decision support tools; (c) it enables personalized medical care through collaboration with systems biology such as genomics; and, (d) it allows direct delivery of clinical information to patients and integration of personalized data (e.g., education, diet habits, and exercise regimens) into health records for targeted health care (Murdoch & Detsky, 2013; Raghupathi & Raghupathi, 2014). The fact that research using EHR is retrospective is another strength of big data research.

It uses secondary data from a repository, and thus the research does not require patient recruitment or further data collection, which can speed up the rate of clinical knowledge discoveries (Weiskopf & Weng, 2013).

The variables related to CAUTIs are quite diverse, including factors related to the patient, environment, and staff. Sometimes this diversity makes the impact and import of such variables inconsistent among studies. Traditional statistical methods have limitations when it comes to investigating several interactions among many variables and providing in-depth knowledge from copious data (Murdoch & Detsky, 2013; Zhao & Luan, 2006). Big data research using EHRs can assist in discovering new insights and knowledge for CAUTIs from a large amount of data.

This study uses existing data from an EHR stored in the University of Minnesota (UMN) Academic Health Center-Information Exchange (AHC-IE) Clinical Data Repository (CDR). The AHC-IE includes more than 2 .4 million patients from seven hospitals and 40 clinics. Also, University of Minnesota Medical Center (UMMC) reports NDNQI data that are electronically available. Therefore, multiple electronic data sources can be linked to an integrated dataset. The study uses the knowledge discovery in databases (KDD) approach for data analysis. This big data research analytic method includes a data mining process that consists of selecting, exploring, and modeling large amounts of data to discover new patterns or relationships (Bellazzi & Zupan, 2008). The data mining process involves considering various models and choosing the best one based on its predictive performance.

Scientific Importance of This Study

Studies identifying risk factors and evidence-based prevention guidelines for CAUTI have existed for years, but a recent report from the CDC showed that CAUTI occurrence still increased by 6% in U.S. hospitals from 2009 – 2013, with an especially significant increase in CAUTIs from 2012 – 2013 (CDC, 2015b). There is a need for further understanding of additional factors associated with hospital-acquired CAUTIs to reduce the occurrence rates. Big data research using EHRs and nurse staffing data has a merit for such new knowledge discovery from large-scale data.

Catheter-associated urinary tract infection is one of the nursing-sensitive outcome indicators strongly influenced by nurses' activities and abilities (Montalvo, 2007). The quality of nursing care is related to nurse staffing, and better nurse staffing has been recommended as a means to improve patient outcomes (Kane et al., 2007). The IOM's 2010 report, "The Future of Nursing," recommended higher levels of education and training for nurses to meet the patient needs and provide safe, high-quality, and patient-centered care (IOM, 2010). Previous studies have found that nurse staffing-related factors, including a higher percentage of RNs in the staffing mix and the educational level of nurses, are inversely associated with UTIs (Kovner & Gergen, 1998; Needleman et al., 2002; Seago, 2001). However, these studies are dated, and only a few studies have been conducted recently on this topic.

This study is scientifically important because there is as yet no research on a predictive model of hospital-acquired CAUTI using a combination of data from EHRs and nurse staffing factors collected as part of the NDNQI submission. Thus, this study can provide new insights about hospital-acquired CAUTIs.

Purpose and Specific Aims of Study

Purpose of Study

The overall purpose of this study was to develop a predictive model for hospital-acquired CAUTIs using multiple data sources. This goal is achievable by creating an integrated dataset from multiple data sources, generating models to classify patients who acquired CAUTI during their hospitalization using data mining methods, and then evaluating the results to find the best predictive model. This predictive model can be implemented in real settings as part of the evidence-based prevention recommendation for hospital-acquired CAUTIs.

Specific Aims of Study

The specific aims of this study are:

Aim 1: Create a quality, de-identified dataset combining multiple data sources for machine learning tasks.

Aim 2: Develop and evaluate predictive models to find the best predictive model for hospital-acquired CAUTI.

Conceptual Framework

Developing the Conceptual Framework

A conceptual framework is the overall theoretical underpinnings to organize a study's ideas. Researchers can integrate observations and knowledge into a structured schema using the conceptual framework. Linking findings into a robust theoretical structure helps evidence be more accessible and useful (Polit & Beck, 2008). Conceptual frameworks guide researchers to investigate the phenomena in a more structured and coherent way.

This study uses systems theory as a conceptual framework. Systems theory was first introduced by Ludwig von Bertalanffy in 1950 and developed by Daniel Katz and Robert Louis Kahn in 1978 (Katz & Kahn, 1978). In systems theory, a system can be viewed as a dynamic entity and a complicated set of elements that interrelate and interact. The interaction among the elements is the key in systems theory. The elements include input, throughput, output, feedback, and environment. According to Katz and Kahn (1978), *inputs* are defined as the influx of information from the external environment; *throughputs* are energies inside the system, transformed by reorganizing the inputs; and *outputs* are products that are exported to the external environment (Katz & Kahn, 1978).

A system takes inputs from the environment, inputs are transformed via throughput and exported in the form of output to the environment (Katz & Kahn, 1978). The environment supports a system as it continuously provides inputs for sustainable inflow. Then a system processes the inputs through interactions and activities for production of outputs (Meyer & O'Brien-Pallas, 2010). In systems theory of nursing care delivery, input is the patients' condition and nurse staffing, throughput means the process such as interventions, environment refers to the hospital system or the element of outer system that can affect the system, and output is the patient outcome. This conceptual framework is the theoretical baseline of this study.

Description of the Conceptual Framework

Figure 1.1 illustrates a conceptual framework for understanding the associated factors of hospital-acquired CAUTI. Known associated factors of hospital-acquired CAUTIs can be categorized into four parts: patient factors, environmental factors, staff factors, and interventions. These known factors can be applied to systems theory in nursing. In systems theory, input is transformed via throughput and results as an output; therefore, in this framework, patient and nurse staff factors can be considered as the input, hospital environment as the environment, interventions as the throughput, and hospital-acquired CAUTIs as the output.

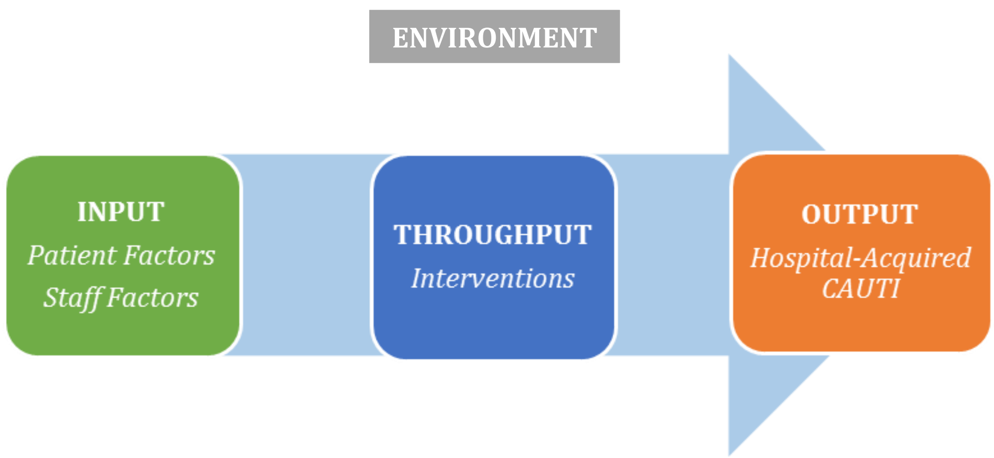


Figure 1.1. Conceptual framework for understanding the hospital-acquired CAUTI.

Conclusion

Nosocomial infections are a prevalent health care problem and are associated with substantial morbidity and mortality. UTIs are one of the most dominant nosocomial infections, and most UTIs are attributable to indwelling urinary catheter use. Studies and guidelines for CAUTI prevention have existed for several decades, but the occurrence rate has increased. Thus, there is a need to discover additional factors that may be related to CAUTI.

The nursing-sensitive quality indicators are strongly influenced by the care nurses provide. CAUTI is a nursing-sensitive quality indicator because nurses take responsibility for managing indwelling urinary catheters. There is a need to look at nurse staffing data to find factors related to CAUTI. In some studies, nurse-staffing related factors have influence on HAIs including CAUTI.

With the implementation of EHRs, researchers can use large amounts of data for knowledge discovery. EHRs contain a wide variety of patient information, including medical history, demographics, encounters, diagnoses, medications, lab results, procedures, allergies, and flowsheets. Big data research using these EHRs is emerging as analytic methods for big data have been advanced and quickly adopted in clinical fields. Data mining, a big data research method, uses large amounts of data to find relationships, associated factors and patterns to help predict outcomes. This research method is especially powerful when there are many variables to analyze.

In this study, the researcher developed a predictive model for hospital-acquired CAUTI using EHR and nurse staffing data from three different data sources. NDNQI reported nurse staffing data are electronically available, which makes the data linkable to EHR. To date, no studies have been found examining predicting hospital-acquired CAUTI using EHRs and nurse staffing data through a big data research method. There is a merit to investigating the factors related to hospital-acquired CAUTIs in a fairly new way.

The literature review in the following chapter will focus on the different organizations' definitions of *CAUTI* and factors related to CAUTI.

CHAPTER II

LITERATURE REVIEW

Organization of the Literature Review

The aim of this study was to develop a predictive model for hospital-acquired catheter-associated urinary tract infections (CAUTIs) using electronic health records (EHR) and nurse staffing data from three data sources. Therefore, the purpose of this literature review is to provide a critical review of the literature for factors and their definitions associated with or predictive of hospital-acquired CAUTI in the adult population. There were few studies about hospital-acquired CAUTI alone, so studies about hospital-acquired UTIs were also included.

Because numerous definitions for hospital-acquired UTIs and CAUTIs have been used, it is essential to find an appropriate and consistent definition. This chapter begins with an examination of various UTI and CAUTI definitions. Then, factors associated with hospital-acquired UTIs and CAUTIs found from literature review are described that may be useful for the development of a model of hospital-acquired CAUTIs.

Methodology for Literature Search

A systematic search for studies and guidelines published between 1999 and 2014 was undertaken using the following biomedical databases: CINAHL, PubMed, OVID Medline, The Cochrane Library, and Google Scholar. The rationale for choosing a 15-year span was that the National Nosocomial Infections Surveillance (NNIS) summary report that came out in 2000 updated previously published data (NNIS, 2000). The search focused on the literature after that report, although hand-searched articles included some earlier works. The key words used were “urinary tract infection,” “catheter-associated urinary tract infection,” “catheter-related infection,” “urinary catheterization,” and “cross infection.” In addition, citations and the list of references were retrieved and systematically reviewed. Unpublished studies, abstracts, dissertations, and theses were not reviewed.

Inclusion and Exclusion Selection Criteria

Studies were included using the following criteria: (a) contained a definition of UTI or CAUTI for the adult population, (b) presented factors associated with UTIs or CAUTIs, (c) conducted in hospital setting, (d) available in printed or downloadable form, and (e) written in English. Exclusion criteria were the following: (a) failure to include a definition of a UTI or a CAUTI and a description of factors associated with UTIs or CAUTIs, (b) contained only microbiological factors for UTIs or CAUTIs, and (c) conducted in a home care setting. The researcher manually screened the title and abstract of each citation based on the inclusion and exclusion criteria. Articles deemed potentially eligible underwent full-text screening.

Selection and Abstraction of Studies

The main outcome of the review was the extraction of definitions and factors associated with hospital-acquired UTIs and CAUTIs. The researcher compared the definitions of UTI and CAUTI by concept in chronological order when searching for pertinent terms. The factors associated with hospital-acquired UTIs and CAUTIs were categorized into four factors, using the conceptual framework introduced in Chapter 1: patient factors, environmental factors, staff factors, and interventions. Patient factors are defined as those related to patients such as a patient's demographics, history of illness, and current health condition. Environmental factors are the external factors affecting development of a hospital-acquired UTI or CAUTI, including duration of catheterization and the location and type of catheter. Staff factors are those related to health care providers such as educating staff on UTIs or CAUTIs and RN staffing levels. Interventions are treatments and actions performed to enhance patient outcome, such as nurse-driven interventions. The variables in each category will be described in detail. Findings are presented about the most prevalent definitions and factors associated with hospital-acquired UTIs and CAUTIs that need to be examined to create a clinical data model for future research.

A total of 350 studies and guidelines were found from the database search and the references cited in the papers. After excluding 42 studies that were inaccessible, the abstracts of 308 studies and guidelines were assessed for review. Of those 308 studies and guidelines, 147 were excluded from the assessment: 21 were not relevant to the topic, 120 did not meet the inclusion criteria, and 6 were not about the adult population. Therefore, 161 studies and guidelines were subject to full-text review and included in the synthesis.

Definitions of UTIs

According to the review, UTI has two major types: symptomatic UTI (S-UTI) and asymptomatic bacteremic UTI (AB-UTI). If the UTI was catheter associated, which means catheter was in place more than 2 days prior to the date of the UTI, it is considered to be a CAUTI. The definitions of CAUTI are premised on the presence of a urinary catheter, and the count of bacteria in a urine culture varied from $\geq 10^2$ to $\geq 10^8$ colony-forming units (CFU)/mL, with no more than two species of microorganisms. The definitions from the 1988 and 2008 Centers for Disease Control and Prevention (CDC) guidelines were mostly used for S-UTI and catheter-associated S-UTI.

Definition of UTI

The definition of a UTI varied in the literature in the count of bacteria in a urine culture. According to Parker et al. (2009), a UTI is an inflammatory response of the epithelium of the urinary tract to invasion and colonization by a pathogen, usually a bacterial species. Patients are diagnosed with a UTI when a urine culture contains $\geq 10^2$ CFU/mL urine (Lundeberg, 1986; Saint, 2000), $\geq 10^3$ CFU/mL urine (Lee et al., 2004; Saint et al., 2006; Stark & Dennis , 1984), and $\geq 10^5$ CFU/mL urine (Bouza et al., 2001; Crouzet et al., 2007; Hakvoort, Elberink, Vollebregt, Ploeg, & Emanuel, 2004; Huang et al., 2004; Ksycki & Namias, 2009; Liedberg & Lundeberg, 1990; Liedberg, Lundeberg, & Ekman, 1990; Lundeberg, 1986; Mnatzaganian et al., 2005; Puri et al., 2002; Schumm & Lam, 2008; Scott, 2010; Sujijantararat, Booth, & Davis, 2005; Topal et al., 2005) with ≤ 2 species of microorganisms.

The most prevalent definition for S-UTI was Garner et al.'s (1988) CDC definition. Patients are diagnosed with an S-UTI when they meet one of the following criteria:

- At least one of the following symptoms with no other recognized cause:
    - Fever >38 °C; urgency; frequency; dysuria; or suprapubic tenderness;
- A positive urine culture of $\geq 10^5$ CFU/mL with ≤ 2 species of microorganisms; and

- Any of the following:
    - Positive dipstick urinalysis (leukocyte esterase or nitrite test); pyuria (10 WBCs/mm^3 or at least 3 WBCs/high-power field); positive urine Gram stain; two urine cultures with 10^2 CFU/mL of the same uropathogen; one urine culture with $\leq 10^5$ CFU/mL of a single pathogen in a patient being treated with antimicrobials; or physician's diagnosis or appropriate therapy for UTI (Andreessen, Wilde, & Herendeen, 2012; Bagshaw & Laupland, 2006; Balkhy et al., 2006; Beveridge, Davey, Phillips, & McMurdo, 2011; CDC/NHSN, 2016; Chen et al., 2013; Crouzet et al., 2007; Garner, William, Emori, Horan, & Hughes, 1988; Gould et al., 2010; Horan & Gaynes, 2004; Horan, Andrus, & Dudeck, 2008; Huang et al., 2004; Nicolle, 2012; NNIS, 2000; NNIS, 2001; NNIS, 2003; Owen, Perez, Bornstein, & Sweeney, 2012; Robert et al., 2000; Rosenthal, Guzman, & Safdar, 2004; Schumm & Lam, 2008; Srinivasan, Karchmer, Richards, Song, & Perl, 2006).

Wilde and Carrigan (2003) and White and Ragland (1995) defined S-UTI when there is a new pain in the back over the kidney region or pain/tenderness over the bladder region, a change in urine characteristics, a new laboratory test result showing a new urine infection or blood in the urine (previously negative), or a clinical diagnosis of UTI. Other symptoms include fever >38 °C, chills, or worsening mental or functional status.

Definition of CAUTI

In the literature, the definition of CAUTI varied in the count of bacteria in a urine culture. A patient is diagnosed with a CAUTI with the presence of a urinary indwelling catheter, urine culture contains $\geq 10^2$ CFU/mL urine (Falkiner, 1993; Saint & Lipsky, 1999; Stamm, 1975; Stamm & Hooton, 1993; Stamm, 1991), $\geq 10^3$ CFU/mL urine (Chenoweth & Saint, 2013; Chenoweth & Saint, 2011; Clec'h et al., 2007; Horan & Gaynes, 2004; Huth, Burke, Larsen, Classen, & Stevens, 1992; Lee et al., 2004; Maki & Tambyah, 2001; Mnatzaganian et al., 2005; Nicolle et al., 2005; Riley, Classen, Stevens, & Burke, 1995; Schumm & Lam, 2008; Stark & Dennis, 1984; Tambyah & Maki, 2000; Tambyah, Halvorson, & Maki, 1999; Tambyah & Maki, 2000; Tambyah, Knasinski, & Maki, 2002; Trautner & Darouiche, 2004), $\geq 10^4$ CFU/mL urine (Bonadio et al., 2005), $\geq 10^5$ CFU/mL urine (Bagshaw & Laupland, 2006; Chant, Smith, Marshall, & Friedrich, 2011; Huang et al., 2004; Laupland et al., 2005; Laupland et al., 2002; Lundeberg, 1986; Platt, Polk, Murdock, & Rosner, 1986; Tambyah & Maki, 2000), $\geq 10^6$ CFU/mL urine (Maki, 1997), and $\geq 10^8$ CFU/mL urine (Al-Habdan et al., 2003; Liedberg et al., 1990; Thibon, Le Coutour, Leroyer, & Fabry, 2000; Verleyen, De Ridder, Van Poppel, & Baert, 1999).

Parker et al. (2009) diagnosed CAUTI if patients have bacteriuria and an elevated white blood cell (WBC) count on a urinalysis examination or an elevated WBC count and two or more of the following signs/symptoms: (a) pain or burning in the region of the bladder, urethra, or flank; (b) fever (>38 °C) or chills; (c) malaise; (d) offensive urine odor; (e) change in color or character of urine; (f) hematuria; (g) bladder spasms/leakage;

(h) catheter obstruction; (i) increased weakness or spasticity; (j) change in mental status; or (k) bacteremia.

In 2016, the CDC/NHSN defined CAUTI as "[a] UTI where an indwelling urinary catheter was in place for > 2 calendar days on the date of UTI event, with day of device placement being day 1, and an indwelling urinary catheter was in place on the date of event or the day before. If an indwelling urinary catheter was in place for > 2 calendar days and then removed, the date of event for the UTI must be the day of discontinuation or the next day for the UTI to be catheter-associated" (CDC/NHSN, 2016).

The most commonly used definition of catheter-associated S-UTI in the literature was from the CDC's 1988 guidelines (Appelgren et al., 2001; Askarian, Hosseini, Kheirandish, & Memish, 2003; Bochicchio et al., 2003; Bruminhent, Keegan, Lakhani, Roberts, & Passalacqua, 2010; Burton et al., 2011; Chan, Semenov, & Gourin, 2013; Chant et al., 2011; Chen et al., 2013; Chenoweth & Saint, 2011; Cope et al., 2009; Crouzet et al., 2007; Elpern et al., 2009; Esposito et al., 2006; Finkelstein, Rabino, Kassis, & Mahamid, 2000; Garner et al., 1988; Gentry & Cope, 2005; Ghezzi et al., 2007; Hashmi, Edward, Selwyn, & Jonathan, 2003; Hazelett, Tsai, Gareri, & Allen, 2006; Horan et al., 2008; Karchmer, Giannetta, Muto, Strain, & Farr, 2000; Keerasuntonpong et al., 2003; Krein et al., 2012; Kubler et al., 2012; Leone, Garnier, Avidan, & Martin, 2004; Leone et al., 2003; Lo et al., 2008; Marra et al., 2011; Merle et al., 2002; Morgan et al., 2012; Peschers, Kempf, Jundt, Autenrieth, & Dimpfl, 2001; Pickard, Lam, MacLennan et al., 2012; Pickard, Lam, Maclennan et al., 2012; Rasslan et al., 2012; Rebmann & Greene, 2010; Rosenthal, Guzman, & Orellano, 2003; Shuman &

Chenoweth, 2010; Stovall et al., 2013; Tissot, Limat, Cornette, & Capellier, 2001; Topal et al., 2005; van der Kooi et al., 2007). According to the CDC's 1998 definition, patients with a urinary indwelling catheter are diagnosed with catheter-associated S-UTI when they have one of the two following conditions:

a) One or more of the symptoms (fever ≥38 °C, urgency, or suprapubic tenderness) with no other causes and a urine culture of $\geq 10^5$ CFU/mL and with no more than 2 species of microorganisms;
b) At least two of the following symptoms with no other causes: positive dipstick urinalysis (leukocyte esterase or nitrite test), pyuria ≥3 WBCs/high-power field, positive urine Gram stain, two urine cultures with at least 10^2 CFU/mL of the same pathogen, one urine culture with $\leq 10^5$ CFU/mL of a single pathogen, or a physician's diagnosis or treatment of a UTI.

The CDC updated the CAUTI guidelines and definition for catheter-associated S-UTI in 2009, and several studies adopted the modified definition (CDC/NHSN, 2016; Gould et al., 2010; Meddings, Rogers, Macy, & Saint, 2010; Meddings, Saint, & McMahon, 2010; Meddings et al., 2012; Nicolle, 2012). According to the CDC's 2009 definition, patients with a urinary indwelling catheter are diagnosed with a catheter-associated S-UTI when they have one or more of these symptoms—fever >38 °C, suprapubic tenderness, costovertebral angle pain, or tenderness—with no other cause and meet one of the two following criteria:

a) Had a positive urine culture of $\geq 10^5$ CFU/mL with no more than 2 species of microorganisms;

b) Had at least one of the following signs and a positive urine culture of $\geq 10^3$ and $<10^5$ CFU/mL with no more than 2 species of microorganisms: positive dipstick urinalysis (leukocyte esterase or nitrite test), pyuria (urine specimen with ≥ 10 WBCs/mm^3 or >5 WBCs/high-power field), or microorganisms seen on Gram stain of urine.

In the Infectious Diseases Society of America (IDSA) guidelines, Hooton et al. (2010) defined catheter-associated S-UTI as the presence of UTI signs with no other recognized source of infection with $\geq 10^3$ CFU/mL of more than 1 bacterial species in a single catheter urine specimen from a patient. Microbiologically confirmed catheter-associated S-UTI has the following conditions: a positive urine culture, bacteriuria up to 3 days after urinary catheter removal, changes in quality of health within 6 weeks, and urethral discomfort related to catheterization.

In 2016, the CDC/NHSN defined catheter-associated S-UTI in the following manner:

Patient must meet 1, 2, and 3 below:

1. “Patient had an indwelling urinary catheter that had been in place for >2 days on the date of event (day of device placement = day 1) AND was either:
   a. Present for any portion of the calendar day on the date of event
   b. Removed the day before the date of event
2. Patient has at least one of the following signs or symptoms with no other identified cause:

   • Fever (>38.0 °C)

• Suprapubic tenderness with no other identified cause

• Costovertebral angle pain or tenderness

3. Patient has a urine culture with no more than 2 species of organisms identified, and at least 1 of which is a bacterium of e w^5 CFU/mL. All elements of the UTI criterion must occur during the infection window period (CDC/NHSN, 2016)".

Definition of Asymptomatic Bacteremic UTI

According to the CDC's 2014 guidelines, an AB-UTI is diagnosed when a patient without a urinary indwelling catheter has (a) no symptoms, (b) a urine culture of $\geq 10^5$ CFU/mL with no more than 2 species of microorganisms, and (c) a positive blood culture with at least one matching microorganism to the urine culture, or at least two matching blood cultures drawn on separate occasions if the matching pathogen is a common skin commensal (CDC/NHSN, 2016).

Factors Associated with Hospital-Acquired CAUTI

The factors associated with hospital-acquired CAUTI were investigated during the literature review. The factors associated with CAUTI in the literature were appraised using four categories from the conceptual framework: patient factors, environmental factors, staff factors, and interventions. All factors associated with UTI and CAUTI found in the literature are summarized in detail in Table 2.1.

Table 2.1.

Summary of factors associated with hospital-acquired UTI or CAUTI

Category	Factors associated with UTI or CAUTI
Patient Factors	▪ Female ▪ Older age >80, >60, and >50 ▪ Previous conditions • Antibiotic usage within the last 3 months • Immunosuppressant therapy within the last 14 days • Previous stroke • Bacteriuria before catheterization • Hospitalization within the last 6 months ▪ Current conditions • Diabetes mellitus • Glucose level >200 mg/dl • Severe underlying illness • Renal function impairment • Incontinence • Major surgical procedures • Urinary tract obstruction • Predisposing medical conditions for UTI

	• Nonsurgical disease ▪ Pregnancy
Environmental Factors	▪ Devices • Prolonged catheterization > 7 and > 6 days • Aseptic placement techniques • Use of closed drainage systems • Antimicrobial silver alloy catheter placement • Urinary catheter reminder system • Drainage tube and collection bag remain below bladder level • Frequency of replacement • Cater insertion after the sixth day of hospitalization • Colonization of drainage bag ▪ Hospitalization • Length of hospital stay • Urgent/ emergent admission ▪ Medical/surgical patients ▪ Medical ICU patients ▪ Medicaid payer status
Staff Factors	▪ Education/training of staff on UTI or CAUTI ▪ RN staffing

Interventions	▪ Nurse-driven interventions

Patient Factors

Female gender was the most common patient factor related to the occurrence of CAUTIs (Bagshaw & Laupland, 2006; Bouza et al., 2001; Bruminhent et al., 2010; Chan et al., 2013; Chenoweth & Saint, 2013; Chenoweth & Saint, 2011; Crouzet et al., 2007; Gentry & Cope, 2005; Gould et al., 2010; Hampton, 2004; Hashmi, Edward, Selwyn, & Jonathan, 2003; Hazelett et al., 2006; Laupland et al., 2005; Leone et al., 2004; Leone, Garnier, Dubuc, Bimar, & Martin, 2001; Leone et al., 2003; Lo et al., 2008; Maki & Tambyah, 2001; Marra et al., 2011; Nicolle, 2012; Puri et al., 2002; Saint & Lipsky, 1999; Shuman & Chenoweth, 2010; Stamm, 1975; Tambyah & Maki, 2000; Tambyah et al., 1999; Tambyah & Maki, 2000; Tambyah et al., 2002; Tambyah & Oon, 2012; Tenke et al., 2008; Tissot et al., 2001; Topal et al., 2005; van der Kooi et al., 2007; Wang et al., 2013).

Older age was reported as a predictor of CAUTI (Bagshaw & Laupland, 2006; Crouzet et al., 2007; Hampton, 2004; Lee et al., 2004; Lo et al., 2008; Muzzi-Bjornson & Macera, 2011; Puri et al., 2002; Saint & Lipsky, 1999; Tambyah et al., 1999; Tambyah et al., 2002; van der Kooi et al., 2007), especially when the patient was over the age of 80 (Chan et al., 2013), over the age of 60 (Pickard, Lam, Maclennan et al., 2012; Tissot et al., 2001; Wang et al., 2013), and over the age of 50 (Chenoweth & Saint, 2013; Chenoweth & Saint, 2011; Graves et al., 2007; Saint & Lipsky, 1999; Shuman & Chenoweth, 2010).

A patient's previous health experience was associated with CAUTI. Previous antibiotic usage within the last 3 months (Bagshaw & Laupland, 2006; Crouzet et al., 2007; Guggenbichler, Assadian, Boeswald, & Kramer, 2011; Johansen et al., 2006; Leone et al., 2004; Leone et al., 2003; Pickard, Lam, Maclennan et al., 2012; Saint & Lipsky, 1999; Tambyah & Oon, 2012; Tenke et al., 2008; Tissot et al., 2001; Topal et al., 2005; van der Kooi et al., 2007; Wagenlehner, Cek, Naber, Kiyota, & Bjerklund-Johansen, 2012) and immunosuppressant therapy within 14 days (Tambyah et al., 1999; Tambyah et al., 2002; Tissot et al., 2001) were associated with CAUTI because an immunosuppressive status can make a patient susceptible to infection. Other factors associated with CAUTI are previous stroke (Saint & Lipsky, 1999; Wagenlehner et al., 2012), bacteriuria before catheterization (Pickard, Lam, MacLennan et al., 2012), and hospitalization within the last 6 months (Johansen et al., 2006; Wagenlehner et al., 2012).

Current health conditions also have influence on CAUTI occurrence. Diabetes mellitus has been reported as a patient-related factor for CAUTI (Chenoweth & Saint, 2013; Chenoweth & Saint, 2011; Foxman, 2002; Gould et al., 2010; Greene et al., 2012; Hampton, 2004; Hooton et al., 2010; Lo et al., 2008; Maki & Tambyah, 2001; Marra et al., 2011; Saint & Lipsky, 1999; Shuman & Chenoweth, 2010; Tambyah et al., 1999; Tambyah & Oon, 2012; Tissot et al., 2001; Topal et al., 2005; Wang et al., 2013). Glucose level >200 mg/dL was associated with CAUTI (Hagerty et al., 2015). Other factors associated with CAUTI include severe underlying illness (Bagshaw & Laupland, 2006; Chan et al., 2013; Chenoweth & Saint, 2013; Chenoweth & Saint, 2011; Gentry & Cope, 2005; Gould et al., 2010; Graves et al., 2007; Leone et al., 2003; Puri et al., 2002;

Saint & Lipsky, 1999; Tambyah et al., 1999; Tambyah et al., 2002; Tissot et al., 2001; Wagenlehner et al., 2012), renal function impairment (Chenoweth & Saint, 2013; Chenoweth & Saint, 2011; Gould et al., 2010; Hooton et al., 2010; Lo et al., 2008; Maki & Tambyah, 2001; Marra et al., 2011; Saint & Lipsky, 1999; Shuman & Chenoweth, 2010; Tambyah et al., 2002; Tambyah & Oon, 2012; Tenke et al., 2008), incontinence (Bliss, Westra, Savik, & Hou, 2013; Foxman, 2002; Gould et al., 2010; Tsuchida et al., 2008), major surgical procedures (Chan et al., 2013; Tambyah et al., 1999; Tambyah et al., 2002), urinary tract obstruction (Johansen et al., 2006; Wagenlehner et al., 2012; Wang et al., 2013), predisposing medical conditions for UTI (Chan et al., 2013; Wagenlehner et al., 2012), nonsurgical disease (Chenoweth & Saint, 2013; Chenoweth & Saint, 2011; Shuman & Chenoweth, 2010), and pregnancy (Bouza et al., 2001; Foxman, 2002).

Environmental Factors

Many environmental factors were related to indwelling urinary catheter use. Prolonged duration of catheterization was the most common environmental factor related CAUTI occurrence (Bagshaw & Laupland, 2006; Bernard, Hunter, & Moore, 2012; Bruminhent et al., 2010; Chan et al., 2013; Chenoweth & Saint, 2013; Chenoweth & Saint, 2011; Clarke et al., 2013; Crouzet et al., 2007; Esposito et al., 2006; Foxman, 2002; Gould et al., 2010; Hampton, 2004; Hart, 2008; Hashmi, Edward, Selwyn, & Jonathan, 2003; Hooton et al., 2010; Keerasuntonpong et al., 2003; Ksycki & Namias, 2009; Laupland et al., 2005; Lee et al., 2004; Leone et al., 2004; Leone et al., 2001; Leone, et al., 2003; Lo et al., 2008; Macfarlane, 1985; Marra et al., 2011; Meddings et al., 2010; Merle et al., 2002; Nicolle, 2012; Palmer et al., 2013; Parker et al., 2009; Pellowe & Pratt, 2004; Plowman, Graves, Esquivel, & Roberts, 2001; Pratt et al., 2001; Pratt et al., 2007; Puri et al., 2002; Saint et al., 2008; Saint, Meddings, Calfee, Kowalski, & Krein, 2009; Saint & Lipsky, 1999; Schwartz & Barone, 2006; Scott, 2010; Shuman & Chenoweth, 2010; Tambyah & Oon, 2012; Tenke et al., 2008; Tissot et al., 2001; Trautner, Hull, & Darouiche, 2005; Warren, 2001; Willson et al., 2009). Some studies specified catheterization greater than seven days (Huang et al., 2004; Nicolle, 2012; Shapiro, Simchen, Izraeli, & Sacks, 1984; Wang et al., 2013) while others required more than six days (Chen et al., 2013; Gentry & Cope, 2005; Maki & Tambyah, 2001). Aseptic catheter placement techniques were reported to help reduce the number of CAUTI incidents (Chan et al., 2013; Chenoweth & Saint, 2013; Chenoweth & Saint, 2011; Conway & Larson, 2012; Conway, Pogorzelska, Larson, & Stone, 2012; Falkiner, 1993; Gould et al., 2010; Hooton et al., 2010; Keerasuntonpong et al., 2003; Lo et al., 2008; Macfarlane, 1985; Maki & Tambyah, 2001; Marra et al., 2011; Nicolle, 2012; Parker et al., 2009;Pratt et al., 2001; Pratt et al., 2007; Saint et al., 2009; Saint & Lipsky, 1999; Shuman & Chenoweth, 2010; Tambyah & Oon, 2012; Tenke et al., 2008; Wagenlehner et al., 2012; Willson et al., 2009). The use of closed drainage systems was recommended to lower the occurrence of CAUTIs (Chan et al., 2013; Conway & Larson, 2012; Conway et al., 2012; Gould et al., 2010; Hart, 2008; Hooton et al., 2010; Ksycki & Namias, 2009; Lo et al., 2008; Macfarlane, 1985; Marra et al., 2011; Nicolle, 2012; Parker 2009; Pellowe & Pratt, 2004; Pratt et al., 2001; Pratt et al., 2007; Puri et al., 2002; Tambyah & Oon, 2012; Tenke et al., 2008; Trautner et al., 2005; Willson et al., 2009). Antimicrobial silver alloy catheter placement was also recommended as a means to reduce the occurrence of CAUTIs (Brosnahan, Jull, & Tracy, 2004; Chan et al., 2013; Gould et al., 2010; Hooton et al., 2010; Lo et al., 2008; Niel-Weise, Arend, & van den Broek, 2002; Parker et al., 2009;Pratt et al., 2007; Seymour, 2006; Tenke et al., 2008; Wagenlehner et al., 2012; Willson et al., 2009). The use of a urinary catheter reminder system was reported to reduce CAUTI incidents (Chen et al., 2013; Gould et al., 2010; Guggenbichler et al., 2011; Huang et al., 2004; Maki & Tambyah, 2001; Meddings et al., 2010; Nicolle, 2012; Palmer et al., 2013; Parker et al., 2009; Pratt et al., 2001; Pratt et al., 2007; Saint, Kaufman, Thompson, Rogers, & Chenoweth, 2005; Scott, 2010; Willson et al., 2009). Drainage tube and collection bag remaining below the level of the bladder prevented CAUTI (Conway & Larson, 2012; Hooton et al., 2010; Lo et al., 2008; Macfarlane, 1985; Maki & Tambyah, 2001; Marra et al., 2011; Nicolle, 2012; Tambyah & Oon, 2012). Relative factors for CAUTIs included frequency of replacement (Parker et al., 2009; Pratt et al., 2001; Pratt et al., 2007; Willson et al., 2009), catheter insertion after the sixth day of hospitalization (Chenoweth & Saint, 2013; Chenoweth & Saint, 2011; Shuman & Chenoweth, 2010), and colonization of drainage bag (Guggenbichler et al., 2011; Leone et al., 2003). Factors related to hospitalization included prolonged length of hospital stay (Leone et al., 2004; Leone et al., 2001; Leone et al., 2003) and urgent/emergent admission (Chan et al., 2013; Leone et al., 2003). Medical/surgical and medical ICU patients (Bagshaw & Laupland, 2006) and Medicaid payer status (Chan et al., 2013) were also related to CAUTIs.

Staff Factors

Education/training of staff about UTI or CAUTI prevention decreased the occurrence of CAUTIs (Andreessen et al., 2012; Conway & Larson, 2012; Gould et al., 2010; Hooton et al., 2010; Lo et al., 2008; Macfarlane, 1985; Marra et al., 2011; Parker et al., 2009; Pellowe & Pratt, 2004; Pratt et al., 2001; Pratt et al., 2007; Tambyah & Oon, 2012; Wagenlehner et al., 2012; Willson et al., 2009).

RN staffing was a factor related to CAUTIs. Several studies have found that higher RN staffing is related to lower incidents of CAUTIs (Blegen, Goode, & Reed, 1998; Kane et al., 2007; Kovner & Gergen, 1998; Lankshear, Sheldon, & Maynard, 2005; Needleman et al., 2002; Unruh, 2003).

Interventions

The advantages of nurse-driven interventions for CAUTIs were reported in many studies (Andreessen et al., 2012; Blodgett, 2009; Chenoweth & Saint, 2013; Conway et al., 2012; Dinc & Erdil, 2000; Fakih et al., 2008; Fuchs et al., 2011; Gould et al., 2010; Lo et al., 2008; Marra et al., 2011; Oman et al., 2012; Palmer et al., 2013; Parker et al., 2009; Pratt et al., 2001; Pratt et al., 2007; Saint et al., 2009; Tambyah & Oon, 2012; Topal et al., 2005; Willson et al., 2009). Supported by the CDC guidelines, nurse-driven interventions included nurse-directed catheter removal for CAUTI prevention.

Overall Summary and Conclusion

This literature review identified the definitions and factors associated with or predictive of hospital-acquired UTIs or CAUTI in the adult population. Most studies used definitions from the CDC and IDSA guidelines.

The review indicated that the most widely used definitions for UTIs and CAUTIs were from the 1988 and 2008 CDC guidelines. These two guidelines shared most criteria. The 2008 guideline added some items. The definition of CAUTI varies and is often used interchangeably with UTI or bacteriuria, as early studies used bacteriuria to define catheter-associated infection (Chenoweth & Saint, 2013).

The studies reviewed have indicated that female gender is the most important variable among patient factors for hospital-acquired UTIs and CAUTIs. It is followed by older age, a diagnosis of diabetes mellitus, the presence of a severe underlying illness, previous antibiotic usage within the last 3 months, and renal function impairment. Females have a higher risk of bacteriuria than males, probably because the shorter female urethra provides perineal flora with easier access to the bladder. Previous/current health conditions were also significant variables for developing hospital-acquired UTIs and CAUTIs.

Among the environmental factors, prolonged catheterization was the most important one. It was the most frequently referenced variable for hospital-acquired CAUTIs. The risk of CAUTIs increases by an estimated 5%–10% for each day the catheter remains in place (Wagenlehner et al., 2012). Aseptic placement techniques, use of closed drainage systems, antimicrobial silver alloy catheter placement, and urinary catheter reminder systems were also considered important. In patients with short-term indwelling catheterization, silver alloy catheters were considered to reduce or delay the onset of CAUTI. After catheter insertion, maintenance of closed drainage was the highest priority. For up to 2 weeks of catheterization, closed drainage can keep the overall risk of

CAUTI below 25% (Kunin, 2006). Most environmental factors were related to catheter use. While patient factors are generally unchangeable, environmental and staff factors are modifiable. Therefore, many articles focused on modifiable factors such as silver alloy–coated catheter use, aseptic placement techniques, and real-time alert systems for health care providers.

The fewest factors were related to staff factors and intervention. They included staff education on UTIs and CAUTIs, RN staffing, and nurse-driven intervention.

The study results have shown that the majority of factors fell into either patient or environmental categories rather than the staff category. Because UTIs and CAUTIs make up a significant portion of nosocomial infections, nurses have an important role in preventing hospital-acquired CAUTIs. CAUTI is one of the nursing-sensitive quality indicators defined by the American Nurses Association (ANA), nosocomial infection is influenced by nursing care. The review results have demonstrated that staff factors included what staff should do to reduce hospital-acquired UTIs or CAUTI rather than the staff members' characteristics. Staff members' characteristics are important, especially because nurses' characteristics are directly linked to nursing care and patient outcomes (Blegen et al., 1998; Curtin, 2003).

Limited research exists linking nurse staff characteristics with patient and environmental characteristics. Although studies that directly investigated the relationship between nurse staffing and hospital-acquired UTIs or CAUTIs were insufficient, it was found that RN staffing, including education level, is an important factor for patient mortality and failure to rescue. The fact that CAUTIs have been associated with increased mortality implies that RN staffing has the potential to affect hospital-acquired CAUTIs (Aiken et al., 2014; Givens & Wenzel, 1980; Platt, Polk, Murdock, & Rosner, 1982; Saint, 2000; Tambyah et al., 2002). A recent Institute of Medicine (IOM; 2010) report, *The Future of Nursing*, indicated the need for highly educated nurses from BSN-educated nurses to graduate level. Investigators and policymakers often focus on nurse-to-patient ratios rather than characteristics of the nursing staff. There is a need for a future study to address the impact of nurses' characteristics on hospital-acquired CAUTIs.

The conceptual framework contributed to the understanding of the factors associated with hospital-acquired UTIs and CAUTIs. Analyzing the literature using this structured framework provided scientific evidence for the study. This review contributes to the theory of hospital-acquired UTI and CAUTI, and through critical review of the literature, provides structured knowledge about the factors associated with hospital-acquired CAUTI to address the gap in preventing hospital-acquired CAUTI. It is important for research to identify a knowledge gap where there is insufficient evidence for practice. There is a need for further research to investigate the relationship between nurse staffing and hospital-acquired CAUTIs. The review is also significant because it identified factors associated with CAUTI that can be used in a practical setting.

Ethically, nurses are responsible for promoting patient advocacy, protecting the art and science of nursing, and encouraging excellence in nursing practice. To provide high-quality nursing care, it is crucial that not only quantitative considerations of RN staffing are required but also qualitative considerations such as the education status of staff and level of experience.

Operational Definition of Hospital-Acquired CAUTI

The most widely used definition for hospital-acquired CAUTI was the CDC's 1998 definition. However, the University of Minnesota Medical Center (UMMC) uses the CDC/NHSN's up-to-date definition of CAUTI. Therefore, this study uses the CDC/NHSN's 2016 definition for CAUTI.

Important patient-related factors for hospital-acquired CAUTI were female gender, old age, severe underlying illness, immunosuppressive status, and severe underlying disease. Important environmental factors for hospital-acquired CAUTI were duration of catheterization, aseptic placement technique, use of closed drainage systems, and urinary catheter reminder system. Important staff-related factors and interventions for hospital-acquired CAUTI were educating staff on UTIs and CAUTIs, and nurse-driven intervention. The results show a need for a future study that identifies the characteristic factors of staff that may have an influence on hospital-acquired CAUTI.

In the following chapter, each variable described in the literature review will be operationalized, and the study's methodology will be described in detail. Of the factors associated with hospital-acquired CAUTI found from the literature, available/attainable variables will be used in the data analysis. The final selected variables for data analysis will be described in detail in Chapter 3. Chapter 3 will explain an analytic plan to address research questions posed in Chapter 1, using the definition and variables introduced in Chapter 2.

CHAPTER III

METHODOLOGY

This chapter focuses on the methodology used in this study. The first section describes the research design and the setting. Detailed descriptions of datasets, outcomes, and factors associated with CAUTI follow, along with supporting rationale for their use. The analytic steps and rationale are presented in detail. Data analysis steps include: data selection, data preparation, data transformation, data mining, and model evaluation.

Research Design and Settings

Knowledge Discovery and Data Mining

This study is designed as a retrospective observational study to develop a predictive model for hospital-acquired catheter-associated urinary tract infection (CAUTI). The study employs a knowledge discovery and data mining (KDDM) approach to unearth new knowledge from large datasets. The knowledge discovery and data mining approach encompasses a variety of statistical analysis, pattern recognition, and machine-learning techniques (Freitas, 2002). This approach is often used for prediction, which is, using some variables in the database to predict the unknown values of other variables of interest. This research method is useful because it can reveal information hidden in large quantities of data and provide a framework for understanding this new information (Pal & Jain, 2007). The KDDM process can be described as an exhaustive search to find information that was previously unnoticed in a database (Giudici, 2003). It is an iterative process of identifying valid and ultimately useful patterns from large collections of data (Cios & Kurgan, 2005).

Large sample sizes allow researchers to discover new insights and knowledge and to create evidence-based guidelines for the clinical field. The use of electronic health records (EHRs) increases the potential for efficient access to comprehensive and standardized data, as EHRs contain a large quantity of patient information and provide massive amounts of structured and unstructured data.

With the wide implementation of EHRs, the data mining approach is emerging as a useful tool for large-scale data analysis. The process of data mining consists of selecting, exploring, and modeling large amounts of data to uncover unknown patterns or relationships (Bellazzi & Zupan, 2008). The factors related to hospital-acquired CAUTIs are quite diverse, comprising factors related to the patient, environment, and staff. Because of this variety, studies of the impact or importance of such variables tend to yield inconsistent results. Traditional analytic methods have limitations investigating multiple interactions among many variables and providing in-depth knowledge from a large amount of data (Zhao & Luan, 2006).

However, large datasets require complicated technological approaches for storage, management, extraction, and analysis. Also, data quality can present a problem, such as when important attributes are missing or when an overabundance of unnecessary data causes noisy data. Poor data quality lowers predictive performance and may obscure the real relationships among data. Hence, datasets require additional preprocessing prior to use. Another issue is that of overfitting. This usually occurs when a model is highly complex, such as when there are too many instances. An overly fitted model generally has poor predictive performance. Therefore, additional techniques, such as cross-validation or pruning, should be used to improve performance (Kohavi & Sommerfield, 1995).

The overall purpose of this study was to develop a predictive model for hospital-acquired CAUTI using an EHR and nurse staffing data from multiple data sources. The study has two specific aims: first, to create a quality, de-identified dataset combining multiple data sources for machine learning tasks, and second, to develop and evaluate predictive models to find the best predictive model for hospital-acquired CAUTI. This predictive model can support the implementation of new knowledge discovered from the study related to hospital-acquired CAUTI. The findings from the study needs to be examined before implanting in real clinical settings.

Setting

The setting for this study includes three intensive care units (ICUs) for adults, specifically, a medical ICU (MICU), a surgical ICU (SICU), and a cardiovascular ICU (CVICU), all located at the University of Minnesota Medical Center (UMMC). The UMMC is a division of Fairview Health Services, a statewide network of hospitals, community clinics, and specialty clinics. Catheter-associated urinary tract infections are the most prevalent type of health care-associated infection in ICUs (Shuman & Chenoweth, 2010), making them an appropriate setting for the CAUTI. Indeed, according to an expert nurse at UMMC who reports CAUTI incidents, very few or no CAUTIs occur outside of ICUs (Mary Fran Tracy, personal communication, October 14, 2014).

Description of Datasets

Three different datasets were used for this study. Dataset 1 from the University of Minnesota Academic Health Center- Information Exchange (hereafter referred to as the UMN AHC-IE dataset) includes electronic health records (EHRs) of ICU patients in the UMN data warehouse. Dataset 2, the National Database of Nursing Quality Indicators (NDNQI) for the UMMC (the UMMC-NDNQI dataset), includes nurse staffing data. Dataset 3 (the UMMC-ICU CAUTI dataset) includes a list of patients who acquired CAUTIs in the ICUs at UMMC. An electronic health records extracted from the UMN AHC-IE clinical data repositories (CDRs) were combined with NDNQI data and the list of CAUTI patients in ICUs.

The UMN AHC-IE dataset is a subset of existing EHR data from the UMN AHC-IE, a CDR that includes more than 2.4 million patients from seven hospitals and over 40 clinics. At the UMN AHC-IE, the following data categories are available: allergies, claims, demographics, diagnoses and problems, encounters and visits, flowsheets, histories, immunizations and vaccinations, institutions and locations, labs, medications, procedures, providers, and vitals. The University of Minnesota collaborated with Fairview Health Services and its Epic EHR system to create the CDR. The CDR has a consented production zone where production-ready clinical data are available for patients who did not opt out to have their data used for research. The UMN AHC-IE is funded through a Clinical Translational Science Award (CTSA) to accelerate research within the community and across CTSA sites. The selected subset of the EHR data for this study contains data on adult patients (aged 18 or over) admitted between January 1, 2012, and June 30, 2015 to any of three ICUs (MICU, SICU, and CVICU) at the UMMC. Patients who did not have indwelling urinary catheters were excluded. The available data for research was collected from January 1, 2012. The total number of patients in three ICUs was 8,496. The total number of unique ICU admissions for final analysis was 11,226, since a patient could have several ICU admissions during the relevant time period.

The UMMC-NDNQI dataset consists of nurse staffing data. The UMMC submits unit-level nurse staffing data to the NDNQI on a quarterly basis. The NDNQI data are collected to measure nursing care quality and nursing work environment at a national level. More than 2,000 hospitals in U.S. and 98% of Magnet recognized clinical facilities participate in the NDNQI program (Press Ganey Associates, 2015). The nurse staffing data submitted to NDNQI between January 1, 2012, and June 30, 2015 in three ICUs were extracted, then included in the UMN data shelter for mapping with other datasets. The NDNQI-submitted data are collected based on nursing hours per patient day, skill mix of nursing staff, registered nurse (RN) education and certification, nurse turnover, nursing work environment, pressure ulcers/prevention, patient falls/prevention, pediatric pain assessment, healthcare-associated infections, restraint prevalence, psychiatric physical/sexual assault rate, and pediatric peripheral IV infiltration.

The UMMC-ICU CAUTI dataset includes a list of patients who acquired CAUTIs during hospitalization in any of the three UMMC ICUs. The list is maintained separately in the UMMC ICU database to track patients with CAUTIs in the ICUs. The list includes the patient medical record number (MRN), department name, and date of CAUTI diagnosis for each patient. The list of CAUTI patients was used to determine which patients acquired CAUTIs during hospitalization in the three ICUs. The list is the gold standard for the outcome measure in this study, as it was created after manual chart reviews using the Centers for Disease Control and Prevention (CDC) definition for CAUTI.

Sample for the Study

The UMN AHC-IE dataset and UMMC-NDNQI dataset were linked using department name and NDNQI reported date. The UMMC-ICU CAUTI dataset was matched to the UMN AHC-IE and UMMC-NDNQI dataset using patient MRN and date of CAUTI diagnosis. The relationship among the three datasets is illustrated in Figure 3.1.

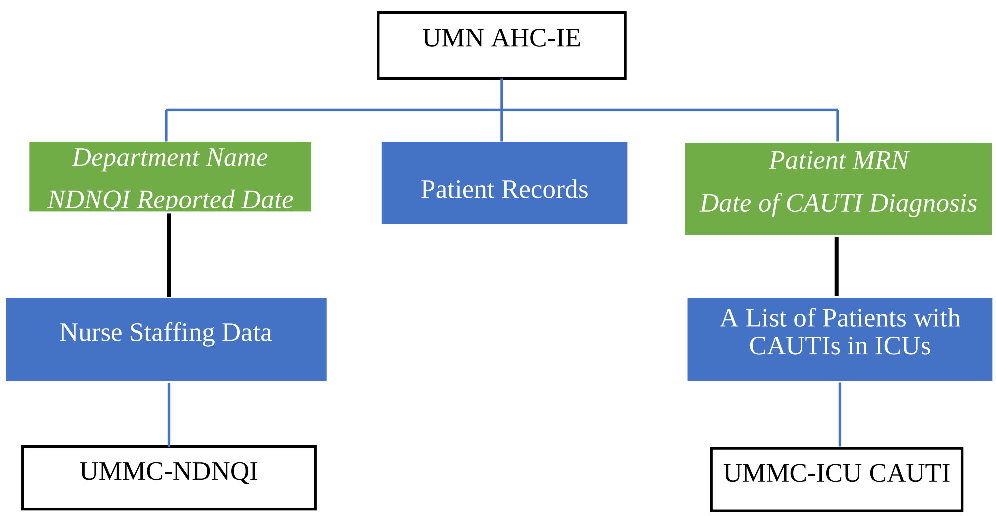


Figure 3.1. Relationship of the three datasets of the study.

Variables

Outcome Variable

According to the CDC definition, a CAUTI is a UTI where an indwelling urinary catheter was in place for more than 2 days on the date of UTI diagnosis, with the day of the device placement being day 1, and an indwelling urinary catheter was in place on the date of the UTI diagnosis or the day before (CDC/NHSN, 2016; Gould et al., 2010). An indwelling catheter is a drainage tube inserted into the urinary bladder through the urethra, left in place, and connected to a drainage bag (including leg bags).

At the UMMC, patients are diagnosed as having CAUTIs based either on the CDC's recent definition or a physician's diagnosis. The ICUs maintain a list of patients who have been diagnosed with CAUTIs to provide better nursing care and follow-up. Because the determination of a CAUTI diagnosis is based on a combination of lab results, observation of symptoms, and clinical judgment, it is reasonable to use the list of patients with CAUTIs for the sake of accuracy and time. This study employed the list in order to determine whether a patient acquired a CAUTI in an ICU. Therefore, the outcome variable in this study refers to patients who acquired a CAUTI during hospitalization between January 1, 2012, and June 30, 2015, at one of three ICUs at UMMC; whose patient MRN appeared on the list of patients having CAUTIs in the ICUs (the UMMC-ICU CAUTI dataset); and who did not have a CAUTI/ UTI when admitted. A detailed description of the conceptual and operational definitions of the outcome variable can be found in Table 3.1.

Table 3.1.

Conceptual and Operational Definitions of Outcome Variable

Variable	Conceptual Definition	Operational Definition and Variable Type
Patient with CAUTI (CDC/NHSN, 2016)	• A UTI where an indwelling urinary catheter was in place for >2 calendar days on the date of event, with day of device placement being Day 1, AND an indwelling urinary catheter was in place on the date of event or the day before. If an indwelling urinary catheter was in place for > 2 calendar days and then removed, the date of UTI event must be the day of discontinuation or the next day for the UTI to be considered catheter-associated. • *Indwelling catheter:* A drainage tube that is inserted into the urinary bladder through the urethra, left in place, and connected to a drainage bag (including leg bags). These devices are also called Foley	▪ A patient who meets the following criteria: 1) A patient whose patient MRN exists in the list of patients who acquired CAUTIs in the ICUs during the period of 01/01/12–06/30/15 2) A patient who did not have a CAUTI/ UTI when admitted ▪ Binary

	catheters. Condom or straight in-and-out catheters are not included nor are nephrostomy tubes, ileoconduits, or suprapubic catheters unless a Foley catheter is also present. Indwelling urethral catheters that are used for intermittent or continuous irrigation are included in CAUTI surveillance.	0 = patient without CAUTI [does not appear on the list] 1 = patient with CAUTI [appears on the list]

Factors Associated with CAUTI

Data selection is the process of sorting out variables relevant to a study (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). The presence of too many unrelated variables can lower the efficiency and accuracy of data mining performance. Removing unnecessary variables and concentrating on essential data samples improves both the speed and quality of the data analysis.

There are two data sources for the factors associated with CAUTI: the UMN AHC-IE and the UMMC-NDNQI. The researcher reviewed the data in the UMN AHC-IE and UMMC-NDNQI datasets, and selected or added variables potentially useful for this study based on the literature review and domain expertise. The results from the literature have shown that factors associated with hospital-acquired CAUTI are categorized into four parts: patient, environment, staff factors, and intervention. However, specific interventions were not measured in this study. The table 3.2 shows factors associated with CAUTI in three categories.

The aim of this study was to create a predictive model for hospital-acquired CAUTIs in an effort to reduce the rate of such infections in the future. Thus, this study examined factors related to hospital-acquired CAUTIs; all of the patient variables collected at admission, during hospitalization, and at discharge were subject to analysis.

Table 3.2.

Factors associated with CAUTI in three categories

	Variables	Data Sources
Patient Factors	Age	UMN AHC-IE
	Gender	
	Race	
	Use of Immunosuppressive Agents	
	Charlson Comorbidity Index Score prior to Admission	
	Lab Results- WBC	
	Lab- Results- Glucose	
	Surgical Procedure History	
	Pre-existing Indwelling Catheter	
Environmental Factors	Prior Hospitalization within 6 Months	UMN AHC-IE
	Length of Stay	
	Rationale for Continued Use of Catheter	
Staff Factors	Total Nursing Hours per Patient Day	UMMC-NDNQI
	Total RN Hours per Patient Day	
	Skill Mix	
	- Percentage of Direct Care RNs with Associate's Degrees in Nursing	
	- Percentage of Direct care RNs with Bachelor	

	of Science in Nursing (BSN), Master of Science in Nursing (MSN), or Doctor of Philosophy (PhD) - Percent of Direct care RNs with Specialty Nursing Certification

Selected Variables from UMN AHC-IE, UMMC-ICU CAUTI, and UMMC-NDNQI Datasets

The conceptual definitions of selected variables from the UMN AHC-IE, UMMC-ICU CAUTI, and UMMC-NDNQI datasets are described in the following.

Selected Variables for Data Extraction and Mapping

Department name. This variable indicates where service was provided. The name of the department is used to limit the UMN AHC-IE data to the three ICUs at UMMC. There are three types of departments in this data— MICU, SICU, and CVICU. All three datasets include department name; therefore, this variable is one of the key variables linking the three datasets.

Admission datetime and discharge datetime. These variables indicate when a patient was admitted and discharged from the hospital; therefore, the dates define a hospitalization period. The patient admission date must be within the range of January 1, 2012, to June 30, 2015. Discharge datetime is used to calculate the number of days stay in hospital. Once used, these variables are not included in the analysis.

Medical Record Number (MRN). The patient MRN is one of the key variables for linking the UMN AHC-IE and UMMC-ICU CAUTI datasets.

Master service ID. Each master service represents a single hospitalization for a given patient, and may include several services performed during that one hospitalization. When developing a predictive model, a patient's hospitalization with unique ICU admission was analyzed to see the consequences of a hospitalization at a ICU related to CAUTI occurrence. Since a patient could have several hospitalizations between January 1, 2012 and June 30, 2015, it is logical to investigate the results of each hospitalization that may have resulted in a CAUTI.

Lab result date. The date on which a lab sample was tested must be between January 1, 2012, and June 30, 2015. Because the conditions of immunosuppression or blood sugar level > 200 mg/dl are associated factors of CAUTI, it is important to capture the appropriate lab results for each patient.

Diagnosis date of CAUTI. The date a patient was diagnosed as having a CAUTI was used to link the diagnosis to a patient's specific hospitalization.

Selected Variables for Data Analysis

Age. Age is defined as a patient's age at admission. Older age is a predictor of CAUTI.

Gender. A patient's gender was either male or female. This variable is selected because female gender is a predictor of CAUTI.

Race. A patient's race was included in the analysis. Race includes three categories: White, non-White, and unknown. Due to the prevalence of Whites in Minnesota's population distribution, non-Whites were combined together for the data analysis.

Use of immunosuppressive agents. A patient was considered as using immunosuppressive agents when "immunosuppressive agents" category existed in the patient's medication list. Use of immunosuppressive agents is associated with CAUTI.

Charlson Comorbidity Index Score. Charlson comorbidity index score is a commonly used tool to quantify comorbidity of a patient (Southern, Quan, & Ghali, 2004). Charlson Comorbidity Index Score prior to admission was used to find illness burden at admission that might be associated with patient outcome. Charlson et al. (1987) defined 17 comorbidities and designated weights to measure burdens of clinical conditions (Charlson, Pompei, Ales, & MacKenzie, 1987). Quan et al. (2005) developed the International Classification of Disease, 9th Revision, Clinical Modification (ICD-9-CM) coding algorithms for the Charlson comorbidities (Quan et al., 2005). This study used Quan et al.'s algorithms to calculate the severity of a patient's illness prior to admission. Only the diagnosis codes bearing the "Present-on-Admission (POA)" indicator were used in these calculations. The Centers for Medicare and Medicaid Services (CMS) do not provide payment for hospital-acquired conditions. Thus, CMS collects a POA indicator for all claims involving inpatient admission. The POA indicator was used to determine whether or not the diagnosis code was present at admission. The comorbidities, related weights, and ICD-9-CM are presented in Table 3.3. The total weights were discretized using three categories: 0, 1–2, and ≥ 3.

Table 3.3.

Weighted Index of Comorbidities

Weight	Comorbidities	ICD–9–CM
1	Myocardial infarction	410.x, 412.x
	Congestive heart failure	398.91, 402.01, 402.11, 402.91, 404.01, 404.03, 404.11, 404.13, 404.91, 404.93, 425.4–425.9, 428.x
	Peripheral vascular disease	093.0, 437.3, 440.x, 441.x, 443.1–443.9, 447.1, 557.1, 557.9, V43.4
	Cerebrovascular disease	362.34, 430.x–438.x
	Dementia	290.x, 294.1, 331.2
	Chronic pulmonary disease	416.8, 416.9, 490.x-505.x, 506.4, 508.1, 508.8
	Peptic ulcer disease	531.x–534.x
	Mild liver disease	070.22, 070.23, 070.32, 070.33, 070.44, 070.54, 070.6, 070.9, 570.x, 571.x, 573.3, 573.4, 573.8, 573.9, V42.7
	Diabetes without chronic complication	250.0–250.3, 250.8, 250.9
2	Diabetes with chronic	250.4–250.7

	complication	
	Hemiplegia or paraplegia	334.1, 342.x, 343.x, 344.0–344.6, 344.9
	Renal disease	403.01, 403.11, 403.91, 404.02, 404.03, 404.12, 404.13, 404.92, 404.93, 582.x, 583.0–583.7, 585.x, 586.x, 588.0, V42.0, V45.1, V56.x
	Any malignancy, including lymphoma and leukemia, except malignant neoplasm of skin	140.x–172.x, 174.x–195.8, 200.x–208.x, 238.6
3	Moderate or severe liver disease	456.0–456.2, 572.2–572.8
6	Metastatic solid tumor	196.x–199.x
	AIDS/HIV	42.x–044.x

Prior hospitalization within 6 months. Prior hospitalization within 6 months refers to a patient's experience of hospitalization within 6 previous months. Hospitalization within the previous 6 months is associated with CAUTI.

Length of stay. Length of stay refers to the number of a patient's hospitalization days. Excess length of stay is related to high rates of nosocomial infections in ICUs (Rosenthal et al., 2003).

Lab result— WBC. White blood cell (WBC) lab result indicates whether a patient is in immunosuppression or not. The condition of immunosuppression is associated with

CAUTI. A patient is considered to be immunosuppressed when the lab result for his or her WBC count (normal: 4,500–10,000 cells/mcL) is under 4,500 (CDC, 2007).

Lab result— glucose. Glucose lab result indicates a patient's blood sugar level. Glucose level >200 mg/dl is an associated factor of CAUTI (Hagerty et al., 2015).

Surgical procedure history. Surgical procedure history is defined as having surgical procedure codes in the EHR. History of surgery is a predictor of CAUTI.

Pre-existing indwelling urinary catheter. This variable shows if the patient had an indwelling urinary catheter at admission. This variable was used to see if presence of a urinary catheter at admission had any influence on CAUTI occurrence.

Rationale for continued use of catheter. According to CDC recommendation, patients are at risk if an indwelling urinary catheter remains in place for more than 48 hours. Rationale for continued use of catheter in this study refers to having any of the rationale following for continued use of catheter: anesthesia, incontinence, deep sedation/paralysis, end of life, epidural/intrathecal catheter, genitourinary/gastrointestinal/gynecology pelvic procedure, gross hematuria, insertion difficulty, neurogenic bladder, obstruction, retention, strict Input & Output, transplant, or wound healing. This variable shows a binary answer for whether there was a reason for leaving the indwelling catheter in place for more than 48 hours. This variable was recorded in flowsheet data where nurses document their assessment, intervention, or patient care provided.

CAUTI diagnosis. This variable represents whether a patient was diagnosed with a CAUTI during the hospitalization. If the answer was "yes", then the patient's current hospitalization resulted in CAUTI. At the UMMC, the determination of a CAUTI diagnosis is based on a combination of lab results, observation of symptoms, and clinical judgment.

Total nursing hours per patient day. This variable refers to the number of productive hours worked by RNs, LVN/LPNs, and UAPs per patient day.

*Total RN hours per patient day***.** This variable represents the number of productive hours worked by RNs per patient day. Research has shown that a higher proportion of hours of care per day provided by RNs is associated with a shorter length of stay and lower rates of UTIs and other adverse patient outcomes (Needleman et al., 2002). The skill mix of the nursing staff has an influence on the quality of care and patient outcomes (Joint Commission, 2009). If the hours supplied by RNs is insufficient, less- skilled staff may perform tasks for which they are not trained, thus increasing the risk of adverse patient outcomes (NDNQI, 2012).

Percentages of direct care RNs with associate degrees in nursing. This variable measures direct care RNs' educational level with associate degrees.

Percentages of direct care RNs with BSN, MSN, or PhD. This variable represents the percent of RNs with BSNs or higher nursing degrees, which is related to better patient outcomes (Aiken et al., 2003). The Institute of Medicine (IOM) report "The Future of Nursing" recommended higher levels of education and training for nurses in order to meet the patient needs and provide safe, high-quality, and patient-centered care (IOM, 2010).

Percentage of direct care RNs with specialty nursing certification. This variable measures the percentage of certified nurses who provide direct care. A higher percentage of certified nurses is related to better patient outcomes (Westra et al., 2013; Bliss et al., 2013), and certified nurses perform better than non-certified nurses when testing specialty knowledge (Kendall-Gallagher & Blegen, 2009).

Between the NDNQI data is collected quarterly, the date range between January 1, 2012, and June 30, 2015 included several quarters. The reporting periods are presented in Table 3.4.

Table 3.4.

The quarterly reporting period for NDNQI between January 1, 2012, and June 30, 2015

2012_Q1	01/01/12– 03/31/12	2013_Q4	10/01/13– 12/31/13
2012_Q2	04/01/12– 06/30/12	2014_Q1	01/01/14– 03/31/14
2012_Q3	07/01/12– 09/30/12	2014_Q2	04/01/14– 06/31/14
2012_Q4	10/01/12– 12/31/12	2014_Q3	07/01/14– 09/30/14
2013_Q1	01/01/13– 03/31/13	2014_Q4	10/01/14– 12/31/14
2013_Q2	04/01/13– 06/31/13	2015_Q1	01/01/15– 03/31/15
2013_Q3	07/01/13– 09/30/13	2015_Q2	04/01/15– 06/30/15

The summary of selected variables is presented in Table 3.5. Selected variables are associated factors of CAUTI or work as keys used to link or limit the data tables using structured query language (SQL) queries; after their initial use, these linking keys are not included in the data analysis.

Table 3.5.

Summary of selected variables from UMN AHC-IE, UMMC-ICU CAUTI, and UMMC-NDNQI datasets

Selected Variables for Data Extraction and Mapping					
Variable	Description	Values	Data Type	Rationale	Source
Department Name	Name of the department	MICU, SICU, and CVICU	Not included in data analysis	Limits data to patients at the three ICUs at UMMC	All Three Datasets
Admission Datetime	Date of admission to the hospital	MM/DD/YY (Between 01/01/12 and 06/30/15)		The admission date should be between 01/01/12 and 06/30/15	UMN AHC-IE
Discharge Datetime	Date of discharge from the hospital	MM/DD/YY (Between 01/01/12 and 06/30/15)		To calculate the length in days of the hospital stay	UMN AHC-IE

Patient MRN	Patient’s MRN	######	A key variable used to link the data tables	UMN AHC-IE / UMMC-ICU CAUTI
Master Service ID	Each master service represents a single hospitalization of a given patient	######	Each master service comprises several services during one hospitalization	UMN AHC-IE
Lab Result Date	The date on which the sample was tested at the lab	MM/DD/YY (Between 01/01/12 and 06/30/15)	The date the lab sample was tested at the lab should be between 01/01/12 and 06/30/15	UMN AHC-IE

Diagnosis Date of CAUTI	The date a patient was diagnosed as having a CAUTI	MM/DD/YY (Between 01/01/12 and 06/30/15)		To limit the diagnosis to a patient's certain hospitalization	UMMC-ICU CAUTI
		Selected Variables for Data Analysis			
Age	Patient age at the time of hospitalization	Young Adult (18 ≤ and ≤ 35) Middle-aged Adult (36 ≤ and ≤ 55) Young-old Adult (56 ≤ and ≤ 74) Old-old Adult (75 ≤)	Category	Older age is a predictor of CAUTI	UMN AHC-IE
Gender	Gender of patient	Male Female	Binary	Female gender is a predictor of CAUTI	UMN AHC-IE
Race	Race of a patient	White Non-White Unknown	Category		UMN AHC-IE

Use of Immunosuppressive Agents	If a patient used immunosuppressive agents	Yes No		Binary	Immunosuppression is associated with CAUTI	UMN AHC-IE
Charlson Comorbidity Index Score	Charlson Comorbidity Index Score using diagnoses at admission	Weight	Comorbidities	Continuous	To find potential illness burdens present on admission that might be associated with patient outcome. “Present on Admission” is an indicator of which diagnosis code was used.	UMN AHC-IE
		1	Myocardial infarction			
			Congestive heart failure			
			Peripheral vascular disease			
			Cerebrovascular disease			
			Dementia			
			Chronic pulmonary disease			
			Peptic ulcer disease			
			Mild liver disease			
			Diabetes without chronic complication			

		2	Diabetes with chronic complication			
			Hemiplegia or paraplegia			
			Renal disease			
			Any malignancy			
		3	Moderate or severe liver disease			
		6	Metastatic solid tumor			
			AIDS/HIV			
Prior Hospitalization within 6 Months	Patient is readmitted within 6 months	Yes No		Binary	Readmission within 6 months is an associated factor of CAUTI	UMN AHC-IE

Length of Stay	Number of days the patient was an inpatient	###### (1—169 days)	Continu ous	Length of stay has an influence on nosocomial infections	UMN AHC-IE
Lab Result — WBC	Description of WBC lab result	WBC count < 4,500 cells/mcL N/A	Binary	Immunosuppression is associated with CAUTI	UMN AHC-IE
Lab Result — Glucose	Description of glucose lab result	Glucose level > 200 mg/dl N/A	Binary	Glucose level >200 mg/dl is associated with CAUTI	UMN AHC-IE
Surgical Procedure History	Surgical codes (ICD-9)	Yes No	Binary	History of surgery is associated with CAUTI	UMN AHC-IE
Pre-existing Indwelling Urinary	Patient with indwelling urinary	Yes No	Binary	To investigate if pre-existing catheters have	UMN AHC-IE

Catheter	catheter at admission				influence on CAUTI occurrence	
Rationale for Continued Use of Catheter	Reason for leaving indwelling catheter more than 48 hours	Yes No		Binary	According to CDC recommendation, patients are at risk if an indwelling urinary catheter is in place more than 48 hours. Thus, there should be a reason why an indwelling catheter is in place.	UMN AHC-IE
CAUTI Diagnosis	A patient's diagnosis of	Yes No		Binary		UMMC-ICU

	CAUTI				
Total Nursing Hours per Patient Day	The number of productive hours worked by RN, LVN/LPN, and UAP per patient day	0–24 hours	Continuous	The relationship among nurse staffing, quality of care and patient outcomes has been widely studied.	UMMC-NDNQI
Total RN Hours per Patient Day					
Percentage of Direct Care RNs with Associate Degrees in Nursing	Percentage of RNs in each educational level	0–100%	Continuous	A higher percentage of RNs with BSN or higher nursing degrees is related to better patient outcomes	UMMC-NDNQI
Percentage of Direct Care RNs with BSN, MSN, or					

PhD					
Percentage of Direct Care RNs with Specialty Nursing Certification	Percentage of RNs with specialty nursing certification	0–100%	Continu ous	A higher percentage of certified nurses is related to better patient outcomes	UMMC-NDNQI

Steps in the Analysis

Fayyad et al. (1996) described the KDDM in their work as an interactive and iterative process. The steps of the KDDM process are illustrated in Figure 3.2.

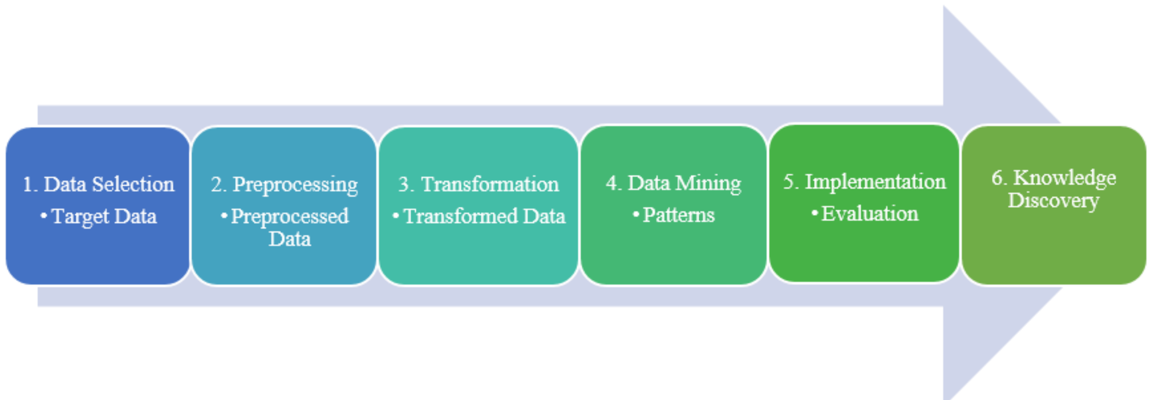


Figure 3.2. An overview of the steps of the KDDM process

The first step of KDDM (previously described) is selecting a target dataset. It is essential to identify a subset of data on which data mining is to be performed. The next step of KDDM is preprocessing the data, which involves data cleaning, such as removing noisy data, collecting the necessary information for model construction, and compensating for missing data fields. Noisy data is meaningless; it cannot be computed by machines. The third step of KDDM is the transformation of data, including reducing data, discretizing numeric attributes, putting data in a unified format, selecting attributes, and transforming multiple classes in an attribute into binary ones. Dimensionality reduction, or the transformation of data, reduces the useful number of attributes under consideration. Data preparation and transformation are important for actual data mining in the following step. The pre-data mining process accounts for roughly 70% of the KDDM process (Ganas, 2009). Therefore, considerable time and effort should be spent on preparing data. The fourth step of KDDM is data mining. Patterns of interest are investigated using the data mining algorithms. This process consists of applying data analysis and discovery algorithms that produce a particular list of patterns or models from the data (Fayyad et al., 1996). This is an important step, best performed by an expert in the field. The model itself is meaningless if experts do not make appropriate interpretations. The fifth step entails interpreting mined patterns; at this stage, one or more of the previous steps may be repeated for further iteration. The sixth step is using the knowledge discovered from the process, which involves incorporating the knowledge into another system for further action or reporting it to other relevant parties.

Analytic Software

Data preparation was carried out using two software; SQL and Python. SQL was used for data extraction and combination after data tables were delivered to the researcher. The Python was used for preprocessing and transformation of data. The data were then collapsed into an integrated dataset, because many data mining techniques require the data to be in this format (Breault, Goodall, & Fos, 2005).

The analysis was carried out using the Weka program, commonly used data mining software developed by a machine learning group at the University of Waikato, New Zealand. Weka supports data preprocessing, attribute selection, visualization of data and analyses such as clustering, classification, and regression.

Data Analysis

Data Preprocessing

Collecting structured data using EHR systems does not always guarantee high-quality data. Common reasons for compromised data quality include missing information, misspelled data, incorrectly categorized data, use of different variables to signify the same concepts, use of retired variables, and use of higher-concept definitions when more detailed sub-concept definitions are available (Brown & Kros, 2003).

There are several ways to handle missing data: a) eliminate all missing values; b) treat missing values as special values "unknown"; c) substitute a feature's mean value by calculating from available data to fill in missing data values; d) choose the most common feature values; e) use the closest fit algorithm, which replaces a missing value with an existing value of the same attribute from another case that resembles it as closely as possible; f) imputing missing values using *k*-nearest neighbors; and g) impute missing values using neural networks (Kaiser, 2014; Kotsiantis, Kanellopoulos, & Pintelas, 2006). Misspelled data, incorrectly categorized data, different variables for the same concept, and retired variable are considered as separate variable from correctly documented variable, which have different meaning. However, the data actually indicate the same concept. These data can mislead the analysis and may result in wrong conclusion. Use of higher-concept definitions rather than detailed sub-concept definitions provide insufficient information about the variable.

Therefore, before conducting the analysis, data were cleaned and organized to enhance the quality of the final product. In this study, the data in the UMN AHC-IE dataset required preprocessing to ensure best results. Because NDNQI-submitted data are reportable and publishable data, their quality was better than that of UMN AHC-IE data. The data in the UMMC-ICU CAUTI dataset were clean as a gold standard for this study.

Data Transformation

A large number of possible variables makes the data mining process slow and ineffective (Kotsiantis et al., 2006). Therefore, data were transformed into a simpler format to reduce the number of possible values to enhance the data mining performance. Data transformation consisting of data discretization, binarization, and data format unification followed the preprocessing. *Data discretization* is defined as a process of converting continuous variable into a finite set of intervals without loss of information (Jin, Breitbart, & Muoh, 2009). Data discretization is needed when the task involves numerical attributes but the chosen data mining algorithm better/only computes categorical ones. During data discretization, numerical values such as patient age were converted into categorical ones. *Data binarization* refers to converting data with multiple classes into binary format (FernáNdez, LóPez, Galar, Del Jesus, & Herrera, 2013). *Data format unification* cleans scattered format of data such as date and time, to improve data processing. Table 3.6 represents the data transformation methods for each variable selected.

Table 3.6.

Data Transformation Methods for Each Variable

Variables	Values	Data Transformation
Age	Young Adult (18 ≤ and ≤ 35) Middle-aged Adult (36 ≤ and ≤ 55) Young-old Adult (56 ≤ and ≤ 74) Old-old Adult (75 ≤)	Continuous → Categorical
Race	White Non-White Unknown	Simplified format
Use of Immunosuppressive Agents	Yes No	Nominal → Binary
Charlson Comorbidity Index Score	0 1–2 3 ≤	Continuous → Categorical
Lab Result—WBC	Yes N/A	Continuous → Binary
Lab Result—Glucose	Yes N/A	Continuous → Binary
Surgical Procedure History	Yes	Nominal →

	No	Binary
Rationale for Continued Use of Catheter	Yes No	Nominal → Binary

Data Mining Analysis

Linking aims to the analytic plan. For the first aim of this study— creating a quality, de-identified dataset combining multiple data sources for machine learning tasks—three data sources were merged into an integrated data file, and data were preprocessed to be compatible with the analytic software. The merging process comprised two steps. First, the UMN AHC-IE and UMMC-ICU CAUTI datasets were linked using patient MRN, department ID, and CAUTI diagnosis date; these datasets were merged using SQL queries. Then this merged dataset was mapped with the UMMC-NDNQI dataset using department ID and NDNQI reporting period. If a single patient had multiple hospitalizations with different departments or reporting periods, each hospitalization was counted separately because nurse staffing data were different for each department or reporting period. The Python program was used to preprocess the data and create the final integrated data file for analytic software.

For the second aim of this study— developing and evaluating predictive models to find the best predictive model for hospital-acquired CAUTIs —the researcher employed classification modeling. This process entails learning a function that classifies data into one of several predefined classes (Kohavi & Quinlan, 2002). This study uses two predefined classes as outcomes: patients with a CAUTI and patients without a

CAUTI. Classification modeling comprises several data mining methods, but not all of them were suitable for this study; different data mining methods require different data types, so it is important to consider the data types of the selected variables when choosing data mining methods. Selected variables from the UMN AHC-IE are mostly categorical and include nominal and binary values. Selected variables from the UMMC-NDNQI and the UMMC-ICU CAUTI datasets are mostly numeric, which means they are continuous variables. Therefore, the data mining methods for this study must be capable of processing both categorical and numerical data. The prediction models from data mining process were compared using discrimination ability and clinical interpretability (discussed later in this chapter). The data mining process can create robust prediction models that have high performance, but sometimes the results are not understandable or clinically valid. Therefore, it is important to take clinical interpretability into consideration.

Unbalanced classes may lower the predictive power of a model. This is important because there is a critical gap between the number of patients with CAUTI (n=10,365) and patients without CAUTIs (n=55). In this study, a cost-sensitive classification method was used to counter class imbalance. Cost-sensitive classification is popularly used in cases of highly imbalanced class distributions (Japkowicz & Stephen, 2002). Cost-sensitive classification refers to the classification method that uses misclassification costs as a penalty to correct class imbalance; the goal is to minimize the total cost (Ling & Sheng, 2011). Most classification algorithms are programmed to minimize the error rate regardless of misclassification errors. That is, whether the error is false positive or false negative, misclassification error are treated equally. However, in medical diagnosis, different types of misclassification errors have different consequences (Turney, 1995). For example, if a patient with a CAUTI is the positive class and a patient without a CAUTI is negative, then misclassifying a CAUTI (e.g., the patient is positive but classified as negative—a false negative) is far more serious, and therefore expensive, than a false positive error. The delay in proper diagnosis and treatment can even endanger the patient. Therefore, cost-sensitive classification gives much higher cost (penalty) to false negatives than to false positives. Mostly, the rarer class is regarded as the positive class; misclassifying an actual positive as a negative (i.e., a false negative) entails a much higher cost than misclassifying an actual negative as a positive. In cost-sensitive classification, any given instance should be classified into the class that has the minimum expected cost. This principle prevents an instance from being classified as a false negative.

When constructing and evaluating predictive models, it is important that they not be built and tested on the same dataset for accuracy (Dreiseitl & Ohno-Machado, 2002). Therefore, a dataset is usually divided into two sets: a training set, which is a set of examples used to construct the prediction model, and a test set, which is used to estimate the prediction performance of the model (Penny & Chesney, 2006). Generally, the larger the training sample, the better the model; by the same logic, the larger the test sample, the more reliable the error estimate. The standard way to validate a model’s performance is to use stratified tenfold cross-validation (Witten et al., 2011). In this procedure, the data are divided randomly into 10 approximately equal partitions; each portion is then used for testing, while the remainder are used for training. That is, nine-tenths of the data are used for training and one-tenth is used for testing, and the procedure is repeated 10 times so that each instance has been used once for testing. This study used tenfold cross-validation for model development and validation.

Data mining techniques for data analysis. What follows is a description of the classification data mining techniques that were used in this study—specifically, decision trees, logistic regression, and support vector machines.

Decision trees. Decision trees (DTs) are tree-like structures that start from root nodes and end with leaf nodes. Decision trees have several branches consisting of different attributes, and the leaf node on each branch represents a class or a kind of class distribution. The model is shown in Figure 3.3.

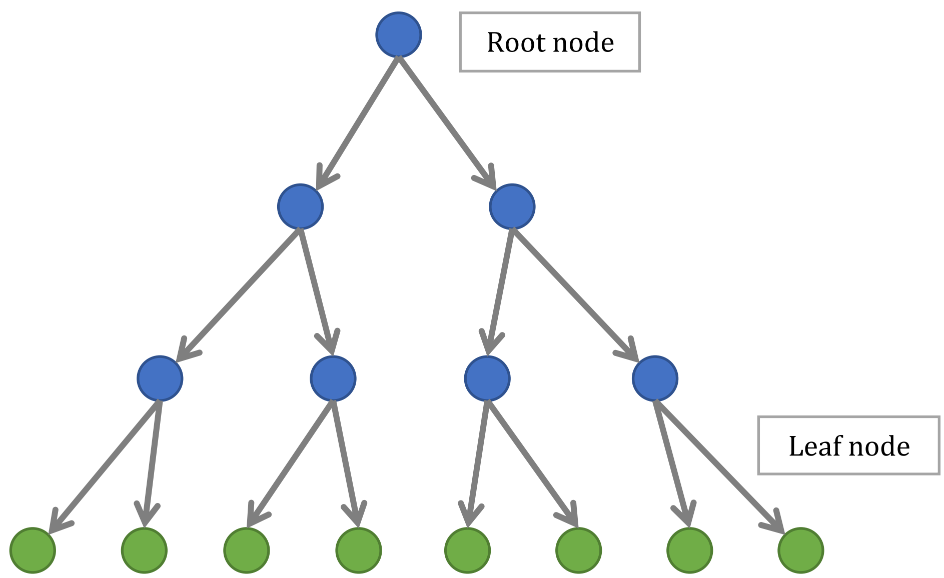


Figure 3.3. Graphical representation of decision trees.

Decision trees describe the relationship among attributes and the relative importance of attributes. This method uses recursive data separation to construct a tree for the purpose of improving predictive accuracy. Mathematical algorithms are used to identify an attribute and corresponding threshold that splits the input observation into two or more subgroups. The threshold maximizes the homogeneity of the resulting subgroups. This step is repeated at each leaf node until the complete tree is constructed. The most commonly used mathematical algorithm for splitting is *entropy-based information gain*, which is used at each split. When this split is maximized, entropy decreases (Dreiseitl & Ohno-Machado, 2002). Decision trees are easy to interpret. However, they are not sensitive to outliers and it has overfitting problem. Due to overfitting, DTs may yield results that are too complex and do not generalize well from the training data. To avoid this, appropriate pruning is necessary. Pruning is a technique that reduces the DT size to identify smaller tree with the lowest error rate (Bradford, Kunz, Kohavi, Brunk, & Brodley, 1998). In Weka, the *RepTree* function was used.

Logistic regression. Logistic regression (LR) is an extension of traditional regression wherein a set of independent attributes is usually used to model a binary outcome (Bellazzi & Zupan, 2008). Logistic regression is an appropriate method for this study to model the dichotomous variable of patients with CAUTIs and patients without CAUTIs. Logistic regression builds the model to predict the odds of an event's occurrence using weights to maximize the log-likelihood. An odds ratio (OR) is a measure of association between a variable and an outcome. When OR>1, a variable is associated with higher odds of outcome; when OR<1, a variable is associated with lower odds of outcome; when OR=1, a variable does not have influence the odds of outcome (Scotia, 2010). Although LR is vulnerable to overfitting, it has proven to be robust in a number of domains and has proven to be an effective method of estimating probabilities from dichotomous variables (Long, Griffith, Selker, & D'agostino, 1993). In Weka, the *Logistic* function was used.

Support vector machines. Support vector machines (SVMs) are among the most powerful classification algorithms in terms of predictive accuracy (Cristianini & Shawe-Taylor, 2000). Support vector machines are becoming popular in the bioinformatics field because of its robust performance and strong mathematical/statistical foundations (Bellazzi & Zupan, 2008). Support vector machines work well with high-dimensional data, and overfitting is unlikely to occur. However, they require extensive computing memory and time to deal with the large amounts of data. Also, the decision during the modeling process are not easy to interpret (Barbella et al., 2009). The core of the SVM method is a process that finds a hyperplane, which separates the examples into different outcomes. An example is illustrated in Figure 3.4. Being separated for two-class problems, SVMs find an optimal hyperplane with a maximum distance to the closest point of each of the two classes. When separated, similarity of the points to each other is important. The similarity is computed by *kernel function* (Leslie, Eskin, & Noble, 2002). The *linear kernel function* that leads to a linear decision hyperplane was used in this study. In Weka, the *SMO* method was used.

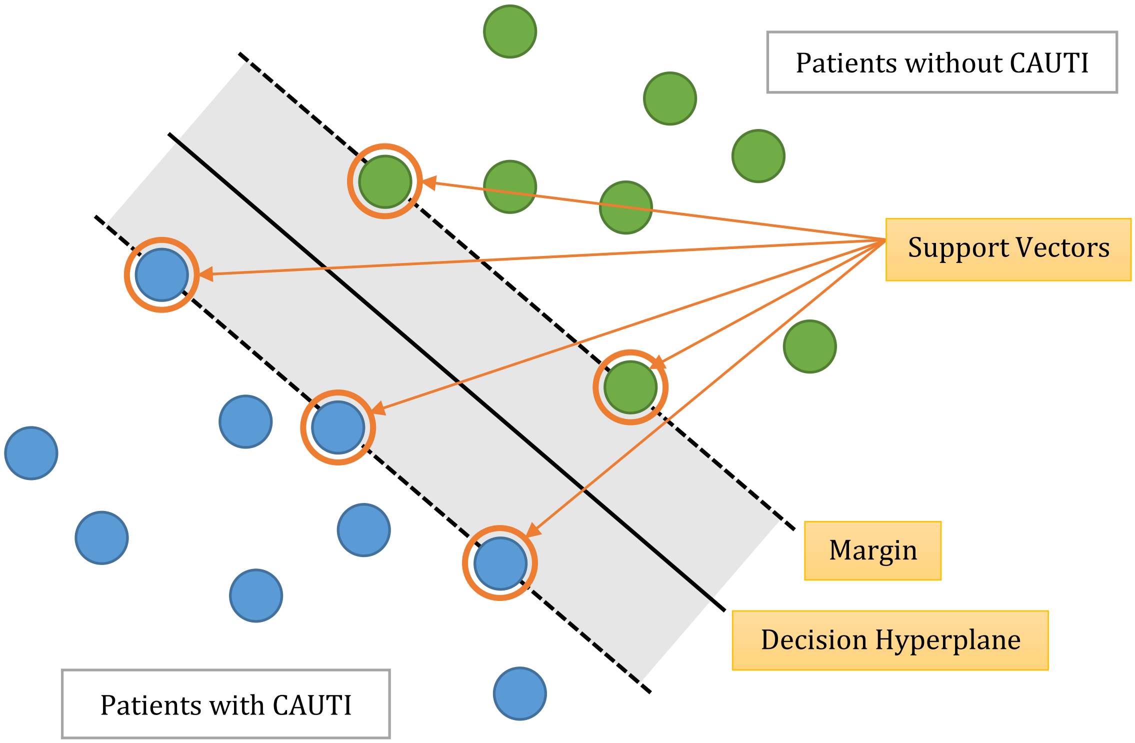


Figure 3.4. Scatterplot of attributes with maximum margin hyperplanes found by an SVM.

Model Evaluation

The main criterion for assessing the quality of a classification model is discrimination ability (Witten et al., 2011). *Discrimination* is a measure of how well two classes in a dataset are separated. Commonly used measures of discrimination are sensitivity, specificity, accuracy, and the area under the receiver operating characteristic (ROC) curve (Beck & Shultz, 1986; Lavrač, 1999). Weka provides the percentages of sensitivity, specificity, and accuracy as well as the area under the ROC curve for each trained model.

In evaluating the models, the researcher used the number of false negatives, accuracy, sensitivity, precision, specificity, and area under the ROC curve to compare the results of several classifiers. Also, the models should be clinically interpretable. Accuracy refers to the overall performance of a model. Sensitivity, precision, and specificity are indicated by true positive (TP), true negative (TN), false positive (FP), and false negative (FN). In this study, TPs refer to the number of patients with CAUTI who are correctly sorted into the patient-with-CAUTI group; TNs refer to the number of patients without CAUTI who are correctly sorted into the patient-without-CAUTI group; FPs refer to the number of patients without CAUTI who are incorrectly sorted into the patient-with-CAUTI group; and FNs refer to the number of patients with CAUTI who are incorrectly sorted into the patients-without-CAUTI group.

Sensitivity (or *TP rate*, as Weka calls it) refers to the ability of a test to be positive when the condition is actually present, it is calculated by measuring TPs among all positive assessments, including TPs and FNs (Zhu, Zeng, & Wang, 2010): $TP/(TP+FN)$. In this study, sensitivity refers to the number of patients with CAUTIs whom the model classifies correctly into the patient-with-CAUTI group (TP) divided by the total number of patients who actually have CAUTIs ($TP + FN$). Sensitivity ranges from 0 to 1, and it is considered the higher the better.

Precision (positive predict value) is the ability when the model predicts positive, how often it is correct. Therefore, precision is the proportion of TPs among the all positively classified positives (TP + FP): $TP/(TP+FP)$. In this study, precision refers to the number of patients with CAUTIs whom the model classifies correctly into the patient-with-CAUTI group (*TP*) divided by the total number of patients that the model classified as positive (*TP*+ *FP*). Precision and sensitivity both focuses on TPs; however, precision finds TPs among all positively classified patients (*TP* + *FP*), whereas sensitivity finds TPs among all patients who actually have positive conditions (*TP* + *FN*).

Specificity (or *TN rate*) measures the proportion of TNs that are correctly identified as negative. In this study, specificity refers to the number of patients without CAUTIs whom the model classifies correctly into the patient-without-CAUTI group (*TN*) divided by the total number of patients who actually do not have CAUTIs (*TN* + *FP*): $TN/(TN + FP)$.

The *area under the ROC curve* is used to evaluate and compare overall model performance (Witten et al., 2011). The discrimination of the model —that is, its ability to distinguish between patients with CAUTIs and patients without CAUTIs—was evaluated by calculating the area under the ROC curve. The discrimination of a predictive model is accurate if the area is 1; it is considered acceptable if it is more than 0.7 (Redon et al., 2010).

Ethical issues

This study was approved by the UMN institutional review board (IRB) for medical chart review of the study (IRB study number: 1506E72421). The data was de-identified except for the patient's MRN and dates associated with events. A limited dataset that includes the patient MRN and dates associated with patient (e.g. labs, birth date, admission, discharge, and transfer) were needed to do the mapping of the datasets. The data for this study was available in the UMN AHC-IE data secure data shelter, accessible only by virtual private network (VPN). Since the data was only available to the researcher and dissertation committee assisting with analysis in the secure workbench at the individual patient-identifiable level, there were no additional risks to subjects. Only aggregated data resulting from analyses, were removed from the secure data shelter.

Conclusion

This study used KDDM approach for prediction, to uncover new knowledge using large amount of data. The KDDM process consist of data selection, preprocessing, transformation, data mining, implementation/evaluation, and knowledge discovery. The setting for the study includes three ICUs for adults located at the UMMC. Three different datasets were used for this study; Dataset 1 (the UMN AHC-IE dataset) including EHRs of ICU patients, dataset 2 (the UMMC-NDNQI dataset) including quarterly reported nurse staffing data, and dataset 3 (the UMMC-ICU CAUTI dataset) including a list of patients who acquired CAUTIs in the ICUs at UMMC. The UMN AHC-IE and the UMMC-NDNQI datasets included factors associated with CAUTI, whereas the UMMC-ICU CAUTI included outcome variable. The study employed the list in the UMMC-ICU CAUTI dataset as outcome variable. Selected factors associated with CAUTI in the UMN AHC-IE dataset include age, gender, race, use of immunosuppressive agents, Charlson Comorbidity Index Score, prior hospitalization within 6 months, length of stay, lab result—WBC and glucose, surgical procedure history, presence of pre-existing indwelling urinary catheter, and rationale for continued use of catheter. Selected factors associated with CAUTI in the UMMC-NDNQI dataset include total nursing hours per patient day; total RN hours per patient day; skill mix—percentage of direct care RNs with an associate degree in nursing; percentage of direct care RNs with bachelor of science in nursing (BSN), master of science in nursing (MSN), or Doctor of Philosophy (PhD); and percent of direct care RNs with specialty nursing certification.

The three datasets were merged into an integrated dataset using key variables including patient MRN, department ID, CAUTI diagnosis date, and NDNQI reported quarter. The data in the integrated dataset were preprocessed and transformed using two software; SQL and Python. The classification modeling was carried out using the Weka program. A cost-sensitive classifier was used to correct unbalanced data. Predictive models using DT, LR, and SVM were developed and compared to find the best predictive model for CAUTI using discrimination ability and clinical interpretability.

In the following chapter, the process of creating an integrated dataset from multiple data sources for machine learning tasks will be described. Then, the study results from each aim and the best predictive model will be explained.

CHAPTER IV

RESULTS

Introduction

The overall purpose of this study was to develop a predictive model for hospital-acquired CAUTIs using an EHR and nurse staffing data from three data sources. The specific aims of this study were to 1) Create a quality, de-identified dataset combining multiple data sources for machine learning tasks, and 2) Develop and evaluate predictive models to find the best predictive model for hospital-acquired CAUTIs. Chapter 3 described the data mining methods to address each of the study aims and selected attributes for analysis. This chapter describes the data analysis findings for each aim.

To address the first aim of the study, the steps in creating an integrated dataset are described. Following is a description of the data preparation process. Then, the characteristics of the data and final attributes for data analysis are presented. The figures present each step of the data preprocessing, creating a flat file, and final attributes for analysis.

To address the second study aim, predictive modeling results were used. The data mining methods of decision trees (DTs), logistic regression (LR), and support vector machines (SVMs) were used for the modeling; then the results were compared to determine the best predictive model for the hospital-acquired CAUTIs. The evaluation criteria were predictive performance – number of false negatives, accuracy, sensitivity, specificity, precision, and area under the Receiver operating characteristic (ROC) curve – and clinical interpretability.

Aim One: Create a Quality, De-identified Dataset Combining Multiple Data Sources for Machine Learning Tasks

The first aim of this study was to create a quality, de-identified dataset combining multiple data sources for machine learning tasks. Dataset 1, extracted from the University of Minnesota Academic Health Center Information Exchange (the UMN AHC-IE dataset), includes Intensive Care Unit (ICU) patient data from the UMN data warehouse. Dataset 2, the National Database of Nursing Quality Indicators (NDNQI) from the University of Minnesota Medical Center (the UMMC-NDNQI dataset), includes nurse staffing data. Dataset 3 includes a list of patients who acquired CAUTIs in the ICUs at UMMC (the UMMC- ICU CAUTI dataset).

Creating an Integrated Dataset

To create an integrated dataset from the three different data sources, data were extracted and then mapped using shared key components. First, selected attributes in the UMN AHC-IE dataset were linked with each other using patient medical record numbers (MRN) and master service IDs. Because data in the UMN AHC-IE dataset were delivered in multiple tables, tying the attributes in the UMN AHC-IE dataset was the first step before mapping this dataset to other datasets. A master service ID represents a unique type of encounter such as hospitalization or clinic. In this dataset, a master service ID was used as a unique hospitalization with admission and discharge dates. Once those data from the UMN AHC-IE dataset were prepared, the dataset was mapped to the UMMC-ICU CAUTI dataset, which contained patient MRNs, department IDs, and CAUTI diagnosis dates. Patient MRN and department IDs were used because they were the shared key attributes. Because a patient could have multiple hospitalizations, – which means one patient MRN could be linked to multiple hospitalizations –, the CAUTI diagnosis date in the UMMC-ICU dataset was used to determine when the CAUTI occurred among the multiple hospitalizations.

For example, if a patient had multiple hospitalizations A, B, and C in different periods –01/01/2012 - 02/01/2012 [A], 08/01/2013 - 09/01/2013 [B], and 11/01/2014 - 12/01/2014 [C] –and the patient was listed on the UMMC-ICU CAUTI dataset which indicated that the patient had been diagnosed as CAUTI in ICU. The patient's CAUTI diagnosis date (e.g. 08/23/2013) in the UMMC-ICU CAUTI dataset was within hospitalization B period. In this case, only hospitalization B would be mapped to UMMC-ICU CAUTI dataset as CAUTI diagnosis of Yes, and other hospitalizations A and C would be mapped as CAUTI diagnosis of No. Table 4.1 visually shows how UMMC AHC-IE and UMMC-ICU CAUTI datasets were mapped.

Table 4.1.

Mapping UMN AHC-IE and UMMC-ICU CAUTI datasets

Data Source	UMN AHC-IE	UMN AHC-IE	UMN AHC-IE	UMN AHC-IE		
		UMMC-ICU CAUTI		UMMC-ICU CAUTI	UMMC-ICU CAUTI	
Period of Hospitalization (Admission- Discharge)		Patient MRN	Patient Data (Gender, age, etc.)	Department ID	CAUTI Diagnosis Date within the Period of Hospitalization (→ CAUTI Diagnosis Y/N)	
H-1	DD/MM/YY	P-1	Patient Data	Department A	DD/MM/YY	No
H-2	DD/MM/YY	P-1	Patient Data	Department A	DD/MM/YY	Yes
H-3	DD/MM/YY	P-1	Patient Data	Department B	DD/MM/YY	No
H-4	DD/MM/YY	P-2	Patient Data	Department C	DD/MM/YY	Yes
H-5	DD/MM/YY	P-3	Patient Data	Department B	DD/MM/YY	Yes
H-6	DD/MM/YY	P-3	Patient Data	Department B	DD/MM/YY	No

Periods of hospitalization are shown using admission and discharge date. Since a patient (e.g. P-1) may have several hospitalizations with different periods, each hospitalization was labeled differently (e.g. H-1, H-2, and H-3). The CAUTI diagnosis date in the UMMC-ICU CAUTI dataset was used to see if the diagnosis date was within the period of hospitalization. If the date was within the period of hospitalization, it was considered as "CAUTI Diagnosis – Yes". If the date was not within the period of hospitalization, or the patient was not listed in the UMMC-ICU CAUTI dataset, it was considered as "CAUTI Diagnosis – No". Figure 4.1 shows all linkages of attributes in UMN AHC-IE using unified modeling language (UML) diagram. The UML diagram is a modeling language, provides a general way to visualize the design of a system.

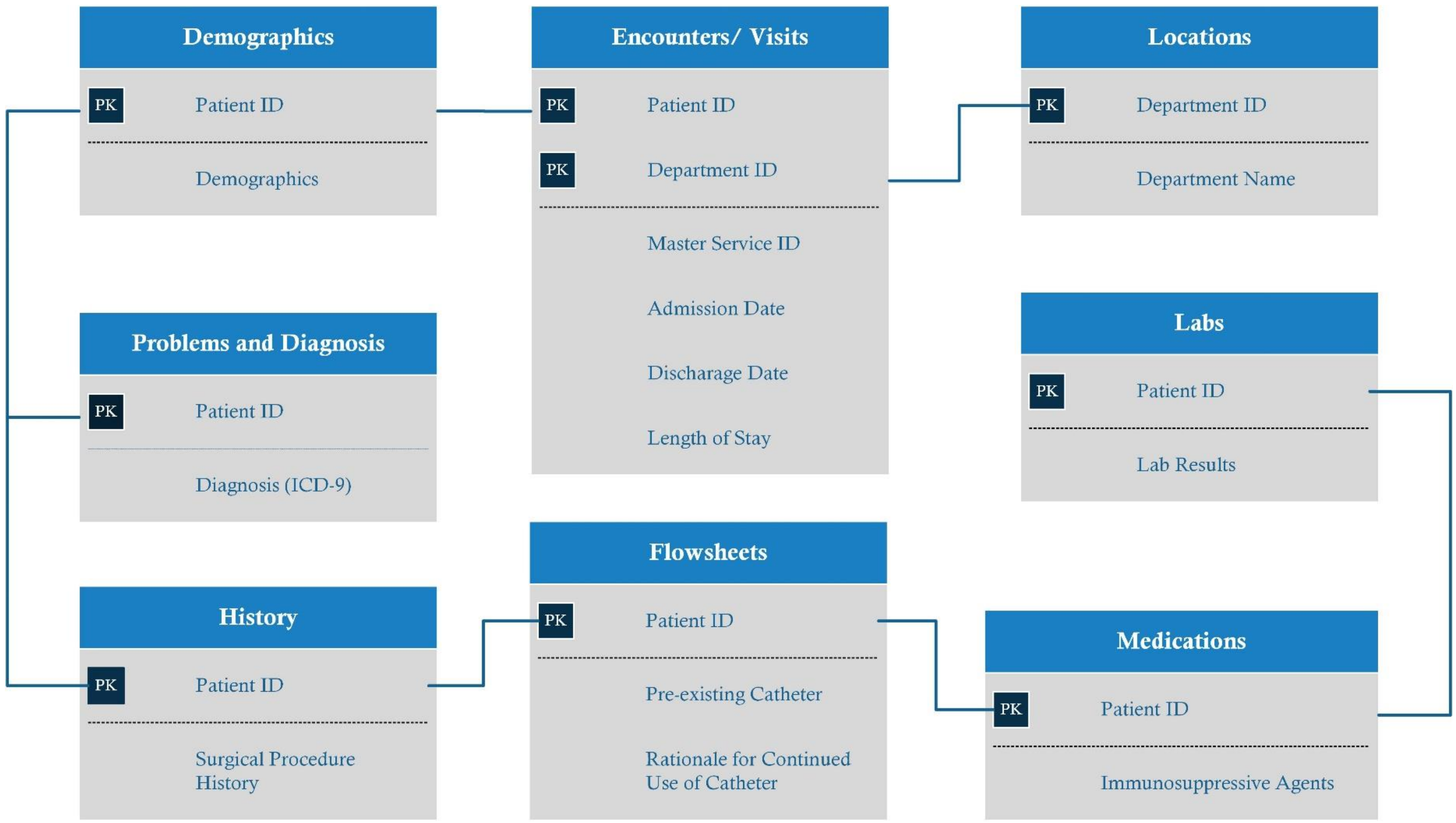


Figure 4.1. The UML Diagram of the UMN AHC-IE dataset.

The quarterly reported NDNQI data from the UMMC-NDNQI dataset was linked to the previous two datasets – UMN AHC- IE and UMMC-ICU CAUTI datasets – using the quarterly reported date and department IDs. For example, patient hospitalizations that occurred from January 1, 2013, to February 20, 2013, were linked to the NDNQI data for the 1st quarter of 2013. Mapping the two datasets with the UMMC-NDNQI set resulted in a larger sample (n = 11,226); this was because some people had stayed in different ICUs during one hospitalization. Therefore, in the final integrated dataset, there was a unique number of total admissions (n = 11,226), and a unique ICU admission was the unit of analysis.

Two software products were used for to create the dataset. SQL was used to export and map the UMN AHC-IE and UMMC-ICU CAUTI datasets and Python software to map those datasets with the UMMC-NDNQI dataset and preprocess the data. After all the datasets were mapped, Python produced an integrated dataset that was tailored to the Weka software used for Aim 2 of the study. Table 4.2 shows the structure of the final integrated dataset.

Table 4.2.

Mapping UMN AHC-IE, UMMC-ICU CAUTI, and UMMC-NDNQI datasets – The final integrated dataset

Data Source	UMN AHC-IE	UMN AHC-IE	UMN AHC-IE		UMN AHC-IE		
			UMMC-ICU CAUTI	UMMC-ICU CAUTI	UMMC-ICU CAUTI		
					UMMC-NDNQI	UMMC-NDNQI	UMMC-NDNQI
Unique ICU Admission	Hospitalization	Patient Data	Patient MRN	CAUTI Diagnosis Y/N	Department ID	NDNQI Reported Quarter	NDNQI Data
1	H-1	Patient Data	P-1	No	Department A	2013_1Q	NDNQI Data_A
2	H-2	Patient Data	P-1	Yes	Department A	2013_3Q	NDNQI Data_A
3	H-2	Patient Data	P-1	No	Department B	2013_3Q	NDNQI Data_B

4	H-3	Patient Data	P-1	No	Department B	2014_2Q	NDNQI Data_B
5	H-4	Patient Data	P-2	Yes	Department C	2013_2Q	NDNQI Data_C
6	H-5	Patient Data	P-3	Yes	Department B	2014_1Q	NDNQI Data_B
7	H-6	Patient Data	P-3	No	Department B	2015_1Q	NDNQI Data_B

A unique ICU admission was the unit for analysis. There were some patients who had multiple hospitalizations. For example, a patient (e.g. P-1) had multiple hospitalizations (e.g. H-1, H-2, and H-3) in the record. During the hospitalization H-2, the patient had stayed in two different departments A and B. In this case, these two admissions were considered as separate admissions (unique ICU admission 2 and 3) because nursing care the patient provided and nurse staffing in each unit were different (NDNQI Data_A and NDNQI Data_B). Unique admission 2 and 3 indicated that hospitalization 2 in department A was related to CAUTI event (CAUTI Diagnosis Yes), whereas hospitalization 2 in department B did not have influence on CAUTI event (CAUTI Diagnosis No). Individual-level patient data were mapped with department-level nurse staffing data. That is, if a patient was hospitalized in department A during the 1st quarter of 2013, the patient data would be mapped with nurse staffing data in department A during 1st quarter of 2013. Figure 4.2 shows all linkages of attributes in the integrated dataset using UML diagram.

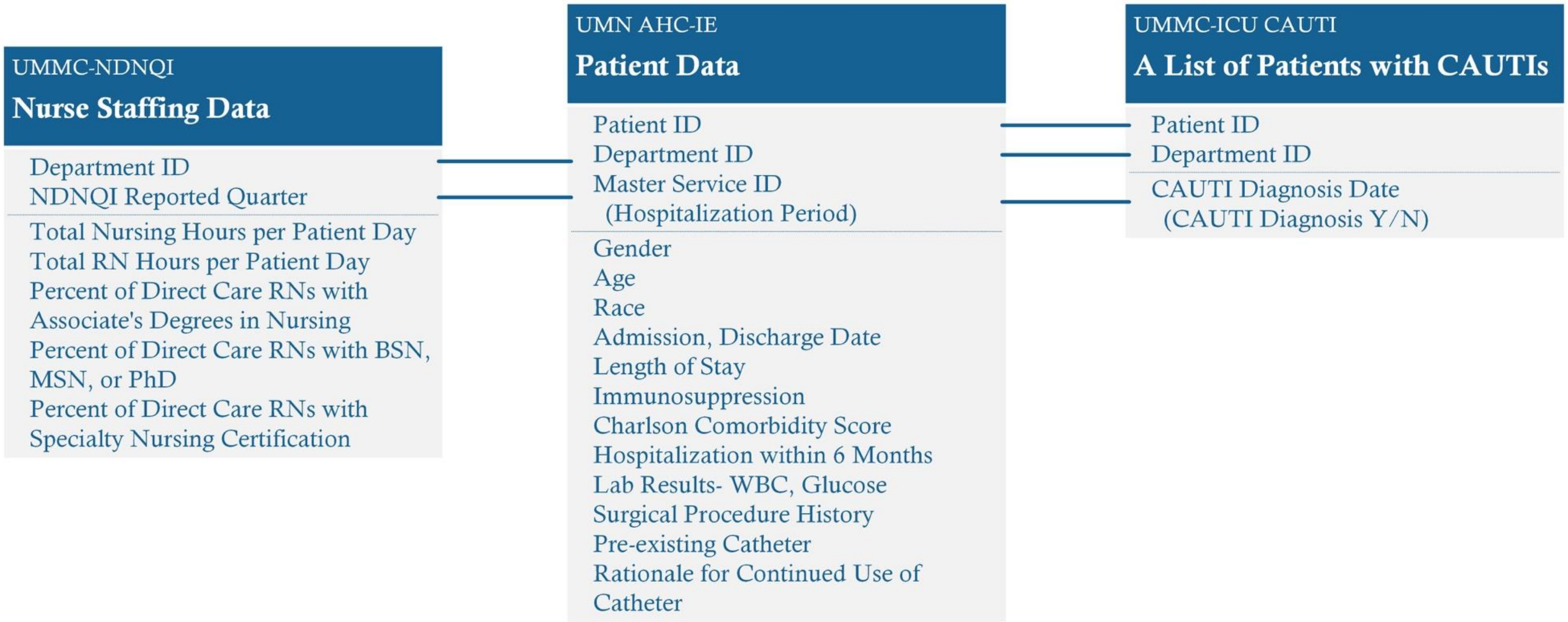


Figure 4.2. The UML diagram of the integrated dataset.

Data Preprocessing and Transformation

While preprocessing the data, missing values were found in the following attributes: race, immunosuppressive agents, lab results, pre-existing urinary catheter, and rationale for continued use of catheter. Missing values in the "race" variable were considered as unknown and moved to the unknown category. Missing values in "immunosuppressive agents" variable were investigated in detail, and the names of any medications administered to the patient were reviewed. If the detailed medication names include known immunosuppressive agents such as Cyclosporine and Mycophenolate, then the value was considered as "yes." Otherwise, it was considered as "no." Missing values in lab results—WBC and glucose—were considered as "no." Because these lab results were critical, it could be assumed that a "yes" result should have been recorded. Missing values in "pre-existing indwelling urinary catheter" were imputed using the *k-nearest neighbors* (k-NN) algorithm which is widely used to classify new cases based on a similarity of available cases (Seidl & Kriegel, 1998). In this algorithm, *k* is the number of nearest neighbors that the algorithm uses to find similarity for prediction of a new case. Therefore, missing values in pre-existing indwelling urinary catheter were re-classified based on the characteristics of nearest neighbors (k=5). Missing data values in "rationale for continued use of catheter" were considered as "no," because otherwise the reason would have been stated. There were no missing values in other variables.

All categorical data were transformed into binary format because logistic regression (LR) in Weka outperforms when the data format is either binary or numeric. Therefore, age was divided into four different groups: young adult (18 ≤ and ≤ 35), middle-aged adult (36 ≤ and ≤ 55), young-old adult (56 ≤ and ≤ 74), and old-old adult (75 ≤). The Charlson comorbidity index score was divided into three groups: score 0, score 1–2, and score ≥ 3.

Characteristics of Data

The data for adult patients admitted between January 1, 2012, and June 30, 2015, to any of three ICUs – Medical ICU (MICU), Surgical ICU (SICU), and Cardiovascular ICU (CVICU) – at the UMMC were extracted from the data shelter. The total number of patients in the three ICUs during the time period was 8,496, and total unique hospitalizations were 10,420. There were 1,292 patients with multiple hospitalizations. Therefore, demographic characteristics are reported for hospitalizations (n= 10,420).

The outcome variable in this study refers to patients who acquired a CAUTI during hospitalization between January 1, 2012, and June 30, 2015 at one of three UMMC ICUs; whose patient MRN appeared on the list of patients having CAUTIs in the ICUs; and, who did not have a CAUTI/ UTI when admitted. A CAUTI was present for a small proportion (0.5%, n= 55) of the hospitalizations.

Using the integrated dataset that was created, the characteristics of each attribute are described. The summary of the characteristics for each attribute is displayed in Table 4.3.

Table 4.3.

Summary of the characteristics for each attribute

Attributes	# (%)	CAUTI Diagnosis Yes (%)
Age		
Young Adult (18 ≤ and ≤ 35)	1,220 (11.7)	2 (3.6)
Middle-aged Adult (36 ≤ and ≤ 55)	3,081 (29.6)	12 (21.8)
Young-old Adult (56 ≤ and ≤ 74)	4,569 (43.8)	32 (58.2)
Old-old Adult (75 ≤)	1,550 (14.9)	9 (16.4)
Race		
White	8,568 (82.2)	48 (87.3)
Non-White	1,224 (11.8)	7 (12.7)
Unknown	628 (6.0)	0
Gender		
Male	4,592 (44.1)	21 (38.2)
Female	5,828 (55.9)	34 (61.8)
Immunosuppression	Y- 967 (9.3)	4 (7.3)
	N- 9,453 (90.7)	51 (92.7)
Surgical Procedure History	Y- 9,331 (89.5)	55 (100%)
	N- 1,089 (10.5)	0 (0)
Lab Result – White Blood Cell	Y- 13 (0.1)	0 (0)

(WBC < 4,500 – Y /N)		N- 10,407 (99.9)	55 (100%)
Lab Result – Glucose		Y- 3,828 (36.7)	32 (58.2)
(Glucose > 200 – Y /N)		N- 6,592 (63.3)	23 (41.8)
Length of Stay	1-10	6,909	6 (10.9)
	11-20	(0.7)	17 (30.9)
	21-30	2,117	13 (23.6)
	31-40	(0.2)	7 (12.7)
	41-50	736 (0.1)	3 (5.5)
	51-60	311 (0)	0 (0)
	61-70	138 (0)	2 (3.6)
	71-80	89 (0)	2 (3.6)
	81-90	57 (0)	1 (1.8)
	91-100	20 (0)	2 (3.6)
	101-110	15 (0)	1 (1.8)
	111-120	6 (0)	0 (0)
	121+	9 (0)	1 (1.8)
		5 (0)	
		8 (0)	
Charlson Comorbidity Score			
Charlson Score = 0		1,859 (17.8)	6 (10.9)
Charlson Score = 1–2		3,433 (33.0)	18 (32.7)

Charlson Score ≥ 3	5,128 (49.2)	31 (56.4)
Hospitalization within 6 Previous Months	Y- 8,649 (83.0) N- 1,771 (17.0)	39 (70.9) 16 (29.1)
Pre-existing Catheter	Y- 774 (7.4) N- 9,646 (92.6)	11 (20.0) 44 (80.0)
Rationale for Continued Use of Catheter	Y- 4,202 (40.3) N- 6,218 (59.7)	51 (92.7) 4 (7.3)
Total Nursing Hours per Patient Day	MICU- 20.61 SICU- 18.95 CVICU- 20.49	
Total RN hours per Patient Day	MICU- 17.39 SICU- 16.27 CVICU- 18.30	
Percent of Direct Care RNs with Associate's Degree Nursing	MICU- 18.67 SICU- 20.53 CVICU- 21.68	
Percent of Direct Care RNs with BSN, MSN, or PhD	MICU- 74.92 SICU- 74.28 CVICU- 71.85	
Percent of Direct Care RNs with	MICU- 18.29	

Specialty Nursing Certification	SICU- 21.81 CVICU- 37.88

Age. Catheter-associated urinary tract infections in young-old adult group showed the highest frequency (58.2%, n=32), whereas CAUTI in young adult group was least frequent (3.6%, n=2).

Race. The majority of the population was white (82.2%), and 87.3% of the CAUTIs (n = 48) occurred in the white population.

Gender. The results showed that out of 10,420 hospitalizations, 44% of patients were female and 56% were male. Despite a lower percent of females, 61.8% of the CAUTIs occurred among the females (n = 34).

Immunosuppression. Most of the population (90.7%) was not immunosuppressed, and most of the CAUTIs (92.7%, n = 51) occurred in the population who was not immunosuppressed.

Surgical Procedure History. The majority of the population, 89.5%, had a history of a surgical procedure, and CAUTIs occurred only in patients with a surgical procedure history.

Lab Result – WBC. Nearly all participants, 99.9%, had WBC counts higher than 4,500 cells/mcL, and 100% of these patients experienced a CAUTI.

Lab Result – Glucose. Most of the population (63.3%) either had no glucose level reported or glucose level lower than 200mg/dl. Over half (58.2%, n=32) of the CAUTIs occurred in the population where patients had glucose level higher than 200mg/dl.

Length of Stay. Nearly two-thirds, 66.3%, of the population had hospital stays of 10 days or fewer, and 86.6% stayed in the hospital for fewer than 20 days. Almost half, 41.8%, of CAUTIs (n = 23) occurred for hospital stays fewer than 20 days.

Charlson Comorbidity Index Score. A total of 17.8% of the population had Charlson comorbidity index scores equal to 0, 33% of the population had the scores 1–2, and 49.2% of the population had the scores greater than or equal to 3. Most CAUTIs (56.4%) occurred in the population with Charlson Comorbidity Index scores greater than or equal to 3.

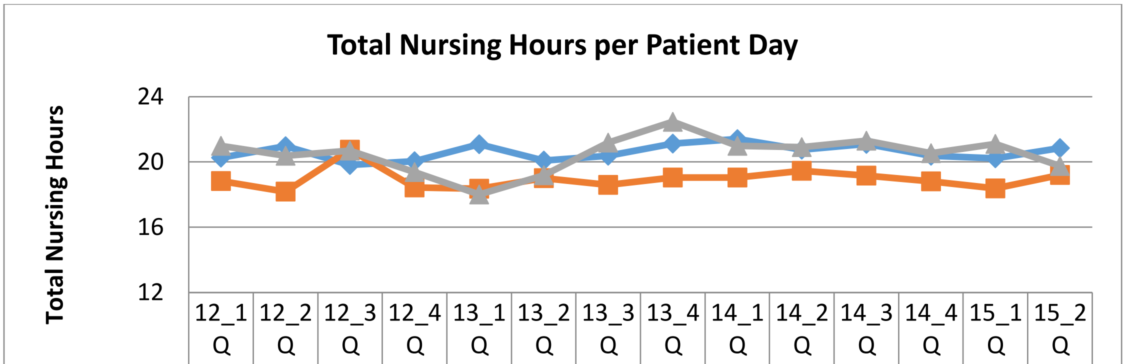
Hospitalization within 6 Previous Months. The great majority (83%) of the population had previous hospitalization within 6 months, and most CAUTIs (70.9%, n = 39) occurred in this population.

Pre-existing Catheter. The majority of the population (92.6%) did not have a pre-existing indwelling catheter when they were admitted. A total of 80.0% of CAUTIs occurred in the population where patients did not have pre-existing indwelling catheter.

Rationale for Continued Use of Catheter. Over half of the population, 59.7%, either had no data reported or did not have a rationale for continued use of an indwelling catheter. The majority, 92.7%, of CAUTIs occurred in the population who did have a rationale for using a catheter.

The characteristics of each unit's quarterly nurse staffing data are presented in the following.

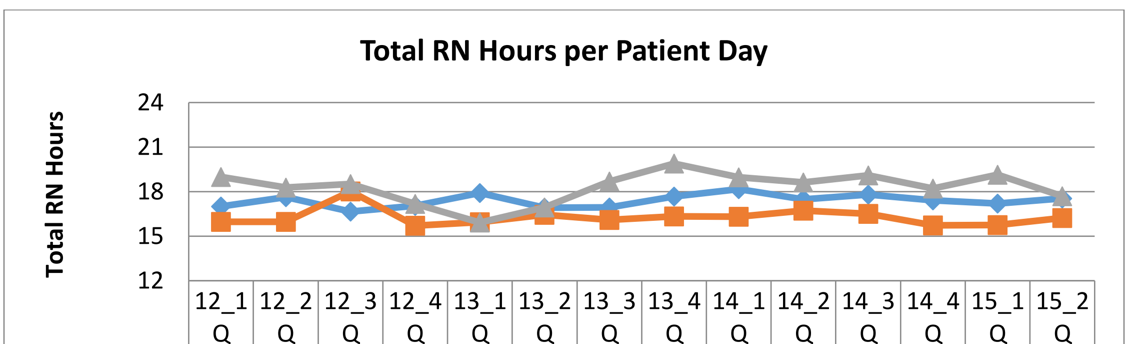
Total Nursing Hours per Patient Day. The average total nursing hours per patient day during the 1st quarter of 2012 – 2nd quarter of 2015 in MICU was 20.61, SICU was 18.95, and CVICU was 20.49 (See Figure 4.3).



	12_1 Q	12_2 Q	12_3 Q	12_4 Q	13_1 Q	13_2 Q	13_3 Q	13_4 Q	14_1 Q	14_2 Q	14_3 Q	14_4 Q	15_1 Q	15_2 Q
MICU	20.26	20.98	19.82	20.04	21.08	20.08	20.39	21.13	21.41	20.76	21.11	20.39	20.23	20.86
SICU	18.84	18.18	20.76	18.45	18.36	19.00	18.59	19.05	19.04	19.46	19.16	18.82	18.38	19.21
CVICU	20.99	20.39	20.69	19.38	18.01	19.18	21.18	22.47	21.00	20.92	21.31	20.54	21.1	19.75

Figure 4.3. Total nursing hours per patient day.

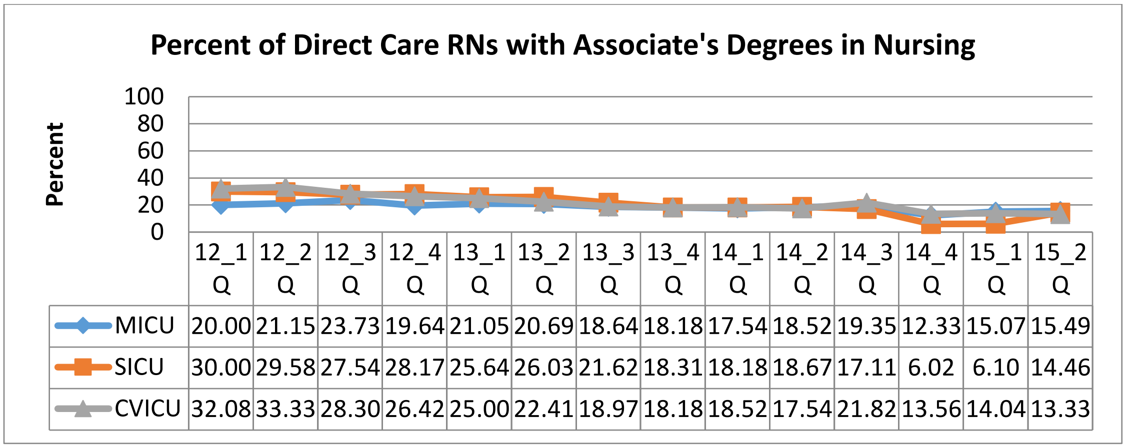
Total RN Hours per Patient Day. The average total RN hours per patient day during the 1st quarter of 2012 – 2nd quarter of 2015 in MICU was 17.39, SICU was 16.27, and CVICU was 18.30 (See Figure 4.4).



	12_1 Q	12_2 Q	12_3 Q	12_4 Q	13_1 Q	13_2 Q	13_3 Q	13_4 Q	14_1 Q	14_2 Q	14_3 Q	14_4 Q	15_1 Q	15_2 Q
MICU	17.01	17.63	16.66	17.04	17.92	16.93	16.94	17.67	18.17	17.49	17.81	17.42	17.21	17.54
SICU	15.98	15.97	18.02	15.71	15.94	16.44	16.10	16.33	16.31	16.73	16.51	15.73	15.75	16.23
CVICU	19.00	18.28	18.52	17.18	15.92	16.95	18.69	19.89	18.97	18.63	19.10	18.21	19.14	17.69

Figure 4.4. Total RN hours per patient day.

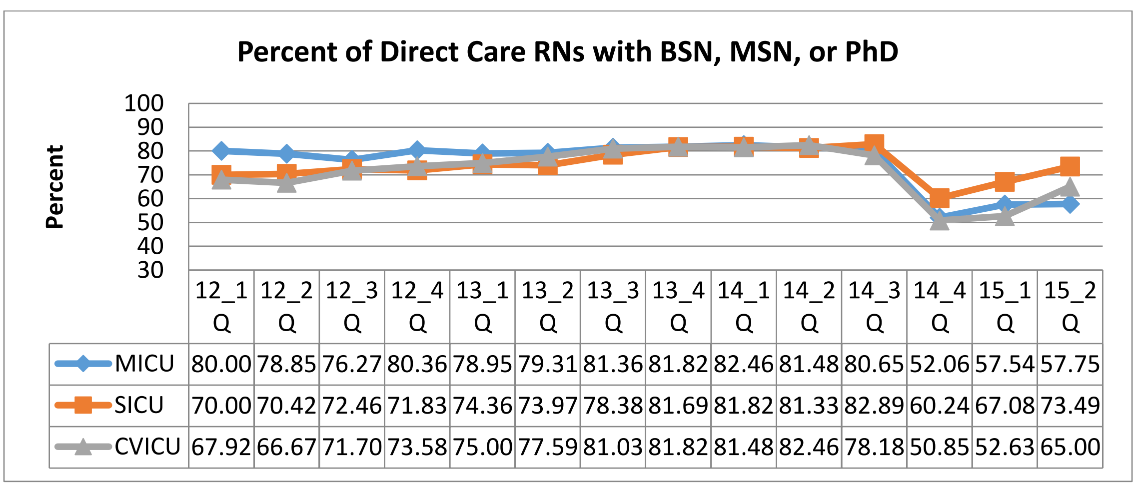
Percent of Direct Care RNs with Associate's Degree Nursing. The average percent of direct care RNs with associate's degrees in nursing during the 1st quarter of 2012 – 2nd quarter of 2015 in MICU was 18.17, SICU was 20.53, and CVICU was 21.68 (See Figure 4.5).



	12_1 Q	12_2 Q	12_3 Q	12_4 Q	13_1 Q	13_2 Q	13_3 Q	13_4 Q	14_1 Q	14_2 Q	14_3 Q	14_4 Q	15_1 Q	15_2 Q
MICU	20.00	21.15	23.73	19.64	21.05	20.69	18.64	18.18	17.54	18.52	19.35	12.33	15.07	15.49
SICU	30.00	29.58	27.54	28.17	25.64	26.03	21.62	18.31	18.18	18.67	17.11	6.02	6.10	14.46
CVICU	32.08	33.33	28.30	26.42	25.00	22.41	18.97	18.18	18.52	17.54	21.82	13.56	14.04	13.33

Figure 4.5. Percent of direct care RNs with associate's degrees in nursing.

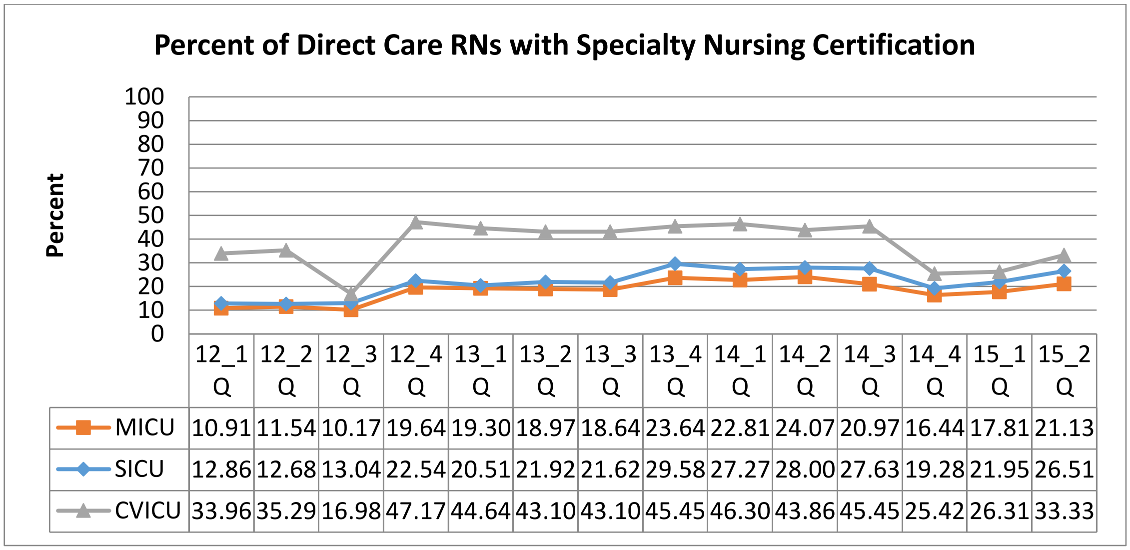
Percent of Direct Care RNs with BSN, MSN, or PhD. The average percent of direct care RNs with BSN, MSN, or PhD during the 1st quarter of 2012 – 2nd quarter of 2015 in MICU was 74.92, SICU was 74.28, and CVICU was 71.85 (See Figure 4.6).



	12_1 Q	12_2 Q	12_3 Q	12_4 Q	13_1 Q	13_2 Q	13_3 Q	13_4 Q	14_1 Q	14_2 Q	14_3 Q	14_4 Q	15_1 Q	15_2 Q
MICU	80.00	78.85	76.27	80.36	78.95	79.31	81.36	81.82	82.46	81.48	80.65	52.06	57.54	57.75
SICU	70.00	70.42	72.46	71.83	74.36	73.97	78.38	81.69	81.82	81.33	82.89	60.24	67.08	73.49
CVICU	67.92	66.67	71.70	73.58	75.00	77.59	81.03	81.82	81.48	82.46	78.18	50.85	52.63	65.00

Figure 4.6. Percent of direct care RNs with BSN, MSN, or PhD.

Percent of Direct Care RNs with Specialty Nursing Certification. The average percent of direct care RNs with specialty nursing certification during the 1st quarter of 2012 – 2nd quarter of 2015 in MICU was 18.29, SICU was 21.91, and CVICU was 37.88 (See Figure 4.7).



	12_1Q	12_2Q	12_3Q	12_4Q	13_1Q	13_2Q	13_3Q	13_4Q	14_1Q	14_2Q	14_3Q	14_4Q	15_1Q	15_2Q
MICU	10.91	11.54	10.17	19.64	19.30	18.97	18.64	23.64	22.81	24.07	20.97	16.44	17.81	21.13
SICU	12.86	12.68	13.04	22.54	20.51	21.92	21.62	29.58	27.27	28.00	27.63	19.28	21.95	26.51
CVICU	33.96	35.29	16.98	47.17	44.64	43.10	43.10	45.45	46.30	43.86	45.45	25.42	26.31	33.33

Figure 4.7. Percent of direct care RNs with specialty nursing certification.

Because of the skewed distribution of CAUTI events in the patient population, some attributes had no variance. For example, all patients had WBC counts of over 4,500 cells/mcL; all patients with a CAUTI had a history of a surgical procedure; more than 90% of patients with a CAUTI were white; and there was little variation in total RN hours per patient day. Therefore, WBC lab results, surgical procedure history, race, and total RN hours per patient day were removed for the final data analysis.

Final Attributes for Data Analysis

Table 4.4 shows the final attributes for the data analysis.

Table 4.4.

*Final Attributes for Analysis**

Attributes	Data Format	Attributes	Data Format
Young Adult (18 — 35)	Binary	Charlson Score = 0	Binary
Middle-aged Adult (36 — 55)	Binary	Charlson Score = 1–2	Binary
Young-old Adult (56 — 74)	Binary	Charlson Score ≥ 3	Binary
Old-old Adult (75 +)	Binary	Hospitalization within 6 Previous Months	Binary
Gender	Binary	Rationale for Continued Use of Catheter	Binary
Immunosuppression	Binary	Lab Result – Glucose	Binary
Pre-existing Catheter	Binary	Length of Stay	Numeric
Total Nursing Hours per Patient Day	Numeric	Percent of Direct Care RNs with Associate's Degree Nursing	Numeric
Percent of Direct Care RNs with Specialty Nursing Certification	Numeric	Percent of Direct Care RNs with BSN, MSN, or PhD	Numeric

***** Definition for variables are listed in Chapter 3.

Aim Two: Developing and Evaluating Predictive Models to Find the Best Predictive Model for Hospital-Acquired CAUTI

Using the final attributes, predictive models were developed; the evaluation and comparison criteria were (a) number of false negatives, (b) accuracy, (c) specificity, (d) sensitivity, (e) the area under the ROC curve, and (f) clinical interpretability. The number of false negatives refers to the number of patients who were classified as not having a CAUTI but did have CAUTI; false negatives are especially important in clinical settings because they prevent patients who need acute treatment from receiving proper care. Therefore, having less false negative value was preferred. The accuracy reflects a model's overall performance. Generally, accuracy is the higher the better. However, because of skewed number of true positives and true negatives, the portion of correctly classified true positives was not much reflected in the overall accuracy. Therefore, the accuracy was used as a minimum threshold for a model (75.00%). Both sensitivity and specificity are used to measure true positive rates from models. The area under the ROC curve indicates a model's performance; it is acceptable when it is over 0.7 (Redon et al., 2010). Clinical interpretability is also important because part of the purpose of this study was to discover new knowledge on hospital-acquired CAUTIs for clinical care.

The predictive models were developed using the following machine learning techniques: decision tree (DT), logistic regression (LR), and support vector machine (SVM). Although the result of SVM model was difficult to interpret, the model was included to compare the predictive performance with other two interpretable models.

Developing the Predictive Models

Three machine learning techniques were used to find the best predictive model for the study: decision tree (DT), logistic regression (LR), and support vector machine (SVM). Because there is a critical gap between the number of patients with CAUTIs and patients without CAUTIs in population, unbalanced classes had to be corrected. To counter class imbalance, a cost-sensitive classification method was used. Cost-sensitive classification uses misclassification costs as a penalty to counter class imbalance (Greiner, Grove, & Roth, 2002). The goal of this approach is to minimize the cost. Most classification algorithms tend to minimize the error rate regardless of misclassification errors – false negatives or false negatives – and treat the error equally. However, cost-sensitive classification in this study gives much higher cost (penalty) to misclassifying false negative than to false positives, to correct the imbalance of the data. Therefore, there was a need to find the optimal cost for the predictive models. Different cost values for the classifier were examined.

Cost-sensitive classifiers. To find the optimal cost for the model, different values (100, 200, and 300) were tested; Table 4.5 shows the results. The number of true positives (TP), false negatives (FN), false positives (FP), and true negatives (TN) are reported. The number of false negatives and accuracy were compared to select the most appropriate cost.

The number of patients who did not have CAUTI was 10,365 and the number of patients who had CAUTI was 55. The ratio was approximately 190:1. Therefore, the testing costs were determined around 200. The study used stratified tenfold cross-validation. In this procedure, the data are divided randomly into 10 approximately equal partitions; each portion is then used for testing, while the remainder are used for training (Witten et al., 2011). This procedure is repeated 10 times so that each instance has been used once for testing.

Table 4.5.

The result of comparing different costs in modeling

Cost 100	TP, FN, FP, TN	Accuracy (%)	Cost 200	TP, FN, FP, TN	Accuracy (%)	Cost 300	TP, FN, FP, TN	Accuracy (%)
DT	TP:33 FN:34 FP:1720 TN:9439	84.38		TP:54 FN:13 FP:2695 TN:8464	75.87		TP: 56 FN:11 FP:2869 TN:8290	74.35
LR	TP:43 FN:24 FP:1756 TN:9403	84.14		TP:50 FN:17 FP:2696 TN:8463	75.83		TP:55 FN:10 FP:4000 TN:7928	71.11
SVM	TP:43 FN:24 FP:1967 TN:9192	82.26		TP:54 FN:13 FP:3186 TN:7973	71.50		TP:57 FN:10 FP:4000 TN:7159	64.28

Results were compared for each of the costs. When the cost (penalty) was low, e.g. 100, the accuracy tended to be high. This was effective modeling in terms of overall accuracy, however, the model emphasized reduction of FP rather than FNs. Higher cost (penalty), e.g. 300, lowered the number of FNs. Instead of reducing the number of FPs, the models focused on reducing the number of FNs. Therefore, the accuracy became lower. The number of FNs was lowest when the cost was 300; however, the accuracy of all three models was below 75.0%. Therefore, cost was set to 200 for each classification modeling because 200 had the fewest possible false negatives.

Decision trees. Weka supports a number of DT models such as J48, RepTree, and BFTree. Among them, BFTree showed the best performance in terms of accuracy and the number of FNs. BFTree creates a decision tree using a best-first expansion of nodes (Witten et al., 2011). Best-first method explores all attributes and sorts them in order of the performance, and finds the most promising node. A heuristic search was used for binary split for nominal attributes. A heuristic function is used to search available information at each branching step to decide which node to follow. The Gini index (Gini impurity) was used for splitting criterion. The Gini index measures how often a randomly chosen element from the training set would be incorrectly labeled. The model used a post-pruning strategy on finding the best number of trees. The post-pruning operates after recursive splitting of DTs, and selects sub-trees that have the best validation accuracy with the smallest number of leaves (Osei-Bryson, 2007). The result from BFTree modeling is shown in Figure 4.8.

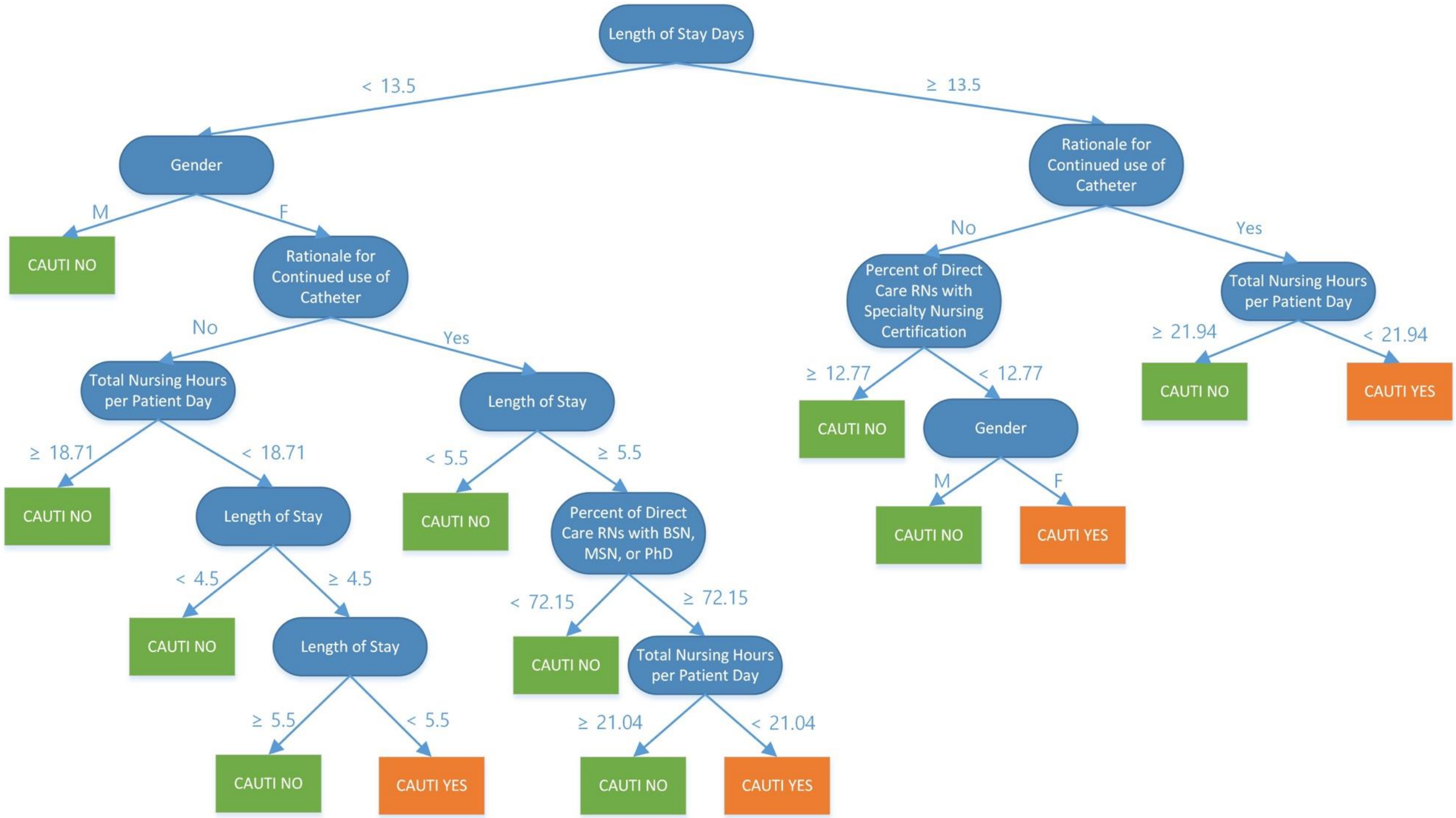


Figure 4.8. Model results from BFTree.

According to the BFTree results, the following were observed for the patients diagnosed with CAUTI:

1) If a patient was female, did not have a valid reason for continued use of catheter, stayed in a unit where total nursing hours per patient day was less than 18.71, and stayed in the hospital between 4.5 and 5.5 days, then the patient was likely to have a CAUTI.

2) If a patient was female, had a valid reason for continued use of the catheter, patient stayed in the hospital between 5.5 and 13.5 days, stayed in a unit where the percent of direct care RNs with BSN, MSN, or PhD was more than 72.15%, and stayed in a unit where total nursing hours per patient day was less than 21.04, then the patient was likely to have a CAUTI.

3) If a patient stayed in the hospital more than 13.5 days, did not have a valid reason for continued use of catheter, stayed in a unit where the percent of direct care RNs with specialty nursing certification was less than 12.77%, and was female, then the patient was likely to have a CAUTI.

4) If a patient stayed in the hospital more than 13.5 days, had a valid reason for continued use of catheter, and stayed in a unit where total nursing hours per patient day was less than 21.94, then the patient was likely to have a CAUTI.

Length of stay, gender, rationale for continued use of catheter, total nursing hours per patient day, percent of direct care RNs with specialty nursing certification, and percent of direct care RNs with BSN, MSN, or PhD, were used in modeling; thus, these factors were related to CAUTI occurrence.

Logistic regression. Table 4.6 shows each attribute's odds ratio (OR). The odd ratios show the magnitude of each attribute's contribution in modeling. Weka does not report confidence intervals (CIs) or p-value for each attribute; instead, the program provided an overall average CI, which was 95% to compare the predictive models.

Table 4.6.

Odds Ratios of Attributes from the Logistic Regression

Attributes	ORs	Attributes	ORs
Young Adult	0.29	Hospitalization within Previous 6 Months	0.65
Middle-aged Adult	0.69	Length of Stay	1.07
Young-old Adult	1.59	Lab Result – Glucose > 200 mg/dl	1.13
Old-old Adult	1.51	Pre-existing Urinary Catheter	0.57
Male Gender	0.21	Rationale for Continued Use of Catheter	14.96
Immunosuppression	0.27	Total Nursing Hours per Patient Day	0.83
Charlson Index Score = 0	0.94	Percent of Direct Care RNs with Associate's Degree in Nursing	1.04
Charlson Index Score = 1–2	0.96	Percent of Direct Care RNs with BSN, MSN, or PhD	1.04
Charlson Index Score ≥ 3	1.07	Percent of Direct Care RNs with Specialty Nursing Certification	0.99

Attributes with OR > 1 are associated with higher risk of CAUTI, whereas attributes with OR < 1 are associated with lower risk of CAUTI (Scotia, 2010). The factors associated with a higher risk of CAUTI are: Young-old adult, old-old adult, Charlson comorbidity index score ≥ 3, length of stay, glucose lab result higher than 200 mg/dl, present of rationale for continued use of catheter, percent of direct care RNs with associate's degree in nursing, and the percent of direct care RNs with BSN, MSN, or PhD. The factors associated with lower risk of CAUTI are: Young adult, middle-aged adult, male gender, immunosuppression, Charlson index score ≤ 2, hospitalization within previous 6 months, pre-existing urinary catheter, total nursing hours per patient day, and percent of direct care RNs with specialty nursing certification.

Support vector machines. SVM model uses a "black-box" approach, which means the decisions during the process are unknown and not easily explainable (Barbella et al., 2009). However, SVM model shows robust performance in classification and is widely used in machine learning tasks (Meyer & Wien, 2015; Westra et al., 2011). Therefore, SVM model was included in this study to compare the performance with other interpretable models. Support vector machines in this study used *linear kernel function* that leads to a linear decision hyperplane. The output of SVM model provides weights of each attribute. Although there is a limitation in terms of interpretation, the attribute weight indicates the attribute's relevance for modeling. That is, having larger absolute value for the weight means that the attribute is relatively important in discrimination of the two classes (Chang & Lin, 2008). The attributes with a positive weight contribute to the classification of CAUTI as Yes, whereas negative weighted attribute contribute to CAUTI as No. The table 4.7 shows the weights of each attribute.

Table 4.7.

The weights of each attribute from the SVM model

Attributes	Weights	Attributes	Weights
Young Adult	-0.65	Hospitalization within Previous 6 Months	-0.48
Middle-aged Adult	-0.03	Length of Stay	7.03
Young-old Adult	0.32	Lab Result – Glucose > 200 mg/dl	0.01
Old-old Adult	0.35	Pre-existing Urinary Catheter	-0.64
Male Gender	-1.02	Rationale for Continued Use of Catheter	1.96
Immunosuppression	-1.13	Total Nursing Hours per Patient Day	-0.30
Charlson Index Score = 0	0.13	Percent of Direct Care RNs with Associate’s Degree in Nursing	0.46
Charlson Index Score = 1–2	-0.18	Percent of Direct Care RNs with BSN, MSN, or PhD	0.70
Charlson Index Score ≥ 3	0.05	Percent of Direct Care RNs with Specialty Nursing Certification	-0.01

The SVM model had the fewest false negatives (n=13); however, its accuracy (71.47%) was the lowest among the three models. The SVM algorithm sacrificed accuracy and concentrated on classifying fewer false negatives.

Model Evaluation

Model evaluation and comparison used the number of false negatives, accuracy, sensitivity, specificity, precision, and the area under the ROC curve. The results of each model using the 200 cost are shown in Table 4.8. The cost was set to 200 for the modeling because 200 had the fewest possible false negatives.

Table 4.8.

Results from the Predictive Models

Cost 200	Number of FN	Accuracy (%)	Sensitivity	Specificity	Precision	ROC area	Clinical Interpretability
DT	TP: 54 **FN: 13** FP: 2695 TN: 8434	75.87	0.81	0.76	0.02	0.78	Yes
LR	TP: 50 **FN: 17** FP: 2696 TN: 8463	75.83	0.75	0.76	0.02	0.85	Yes
SVM	TP: 54 **FN: 13** FP: 3186 TN: 7973	71.50	0.80	0.71	0.02	0.76	No

The results showed that DT (BFTree) model had the highest reduction in FN (n=13), highest accuracy (75.87%) and sensitivity (0.81), and LR model had the largest area under the ROC curve (0.85). SVM model had the highest reduction in FN, tied with DT model. However, the accuracy of SVM model was below 75%. Also, DT model and LR model were clinically interpretable whereas SVM model was not. Since SVM model had accuracy below 75% and was not clinically interpretable, the results of DT and LR models were compared. The DT model showed better performance in reduction of FN, accuracy and sensitivity, and the LR model had larger area under the ROC curve. The specificity and precision were the same for both models. In terms of clinical interpretability, the result of DT model showed associated factors of CAUTI and relationship/interaction among those attributes. The results of LR model showed ORs of each attribute to represent risk factors of CAUTI.

Conclusion

The overall purpose of this study was to develop a predictive model for hospital-acquired CAUTIs using an EHR and nurse staffing data from multiple sources. The specific aims of this study were to 1) Create a quality, de-identified dataset combining multiple data sources for machine learning tasks, and 2) Develop and evaluate predictive models to find the best predictive model for hospital-acquired CAUTIs.

Three datasets—UMN AHC-IE, UMMC-NDNQI, and UMMC-ICU CAUTI—were mapped into an integrated dataset for analysis. Patient MRNs, dates for hospitalization, hospitalized unit, CAUTI diagnosis date, and NDNQI reported date were used as the key components to link UMN AHC-IE, UMMC-NDNQI, and UMMC-ICU

CAUTI datasets. Two software programs, SQL and Python, were used to extract and map the datasets. The Python program produced the final integrated dataset that was suitable for data analysis in Weka.

Predictive models were developed using three different data mining methods: DT (BFTree), LR, and SVM. Evaluation and comparison criteria were the number of false negatives, accuracy, sensitivity, specificity, precision, the area under the ROC curve, and clinical interpretability. The result showed that SVM model had the fewest false negatives (tied with DT model), but the accuracy was below 75%. Also, the SVM model result was difficult to interpret. Therefore, SVM model was ruled out for the best predictive model. The result from DT model showed the fewest FNs, with the highest accuracy and sensitivity. The result from LR model had larger area under the ROC curve. The values for specificity and precision were the same. In terms of clinical interpretability, the result of DT model showed associated attributes of CAUTI and relationships among them. The result of LR model presented individual OR for each attribute to show risk factors for CAUTI. Although both DT and LR models had strengths, the DT model showed better performance in three criteria (fewest FN, highest accuracy and sensitivity) and LR model showed better performance in one criterion (larger ROC curve). Therefore, the DT model was determined as the best model for this study.

The DT results showed that length of hospital stay, female gender, rationale for continued use of catheter, pre-existing catheter, total nursing hours per patient day, percent of direct care RNs with specialty nursing certification, percent of direct care RNs with BSN, MSN, or PhD, and Charlson comorbidity index score all contributed to CAUTI occurrence.

CHAPTER V

DISCUSSION

Nosocomial infections are closely related to patient outcomes such as morbidity and mortality, and higher costs of treatment. Urinary tract infections (UTI) are known as the most common type of nosocomial infections. The Centers for Disease Control and Prevention (CDC) estimated that most nosocomial UTIs occur in Intensive Care Units (ICUs) and are attributable to the use of indwelling urinary catheters. Studies and guidelines for catheter-associated UTI (CAUTI) prevention have existed for several years, but a recent report showed that CAUTI incidents have been increased and it is still prevalent nosocomial infections (Tambyah, 2015). Thus, there is a need to discover additional factors that may be related to CAUTI. The Centers for Medicare and Medicaid Services (CMS) no longer reimburses hospitals for hospital-acquired conditions such as CAUTI, which has resulted in hospitals increasing their attention to nosocomial infections and focusing on implementing evidence-based guidelines.

Catheter-associated urinary tract infections are one of the nursing-sensitive indicators because nurses are often responsible for the management of indwelling urinary catheters. In some studies, nurse-staffing related factors have influence on CAUTI (Hugonnet et al., 2007). Therefore, investigating more nurse-related factors that influence CAUTI occurrence was necessary. This study used data reported to the National Database of Nursing Quality Indicators (NDNQI), the collection of reliable nursing-sensitive indicators on a large scale.

The adoption and use of electronic health records (EHRs) driven by the CMS incentive program enabled collecting vast amounts of data enabling large-scale research. Several studies have reported a growing need for conducting clinical research using big data (Safran et al., 2007). Big data research using the EHRs is emerging as analytic methods for big data have been advanced and adopted in clinical fields.

The variables related to CAUTI occurrence are diverse, and this diversity makes the impact of such variables inconsistent among studies. Traditional statistical methods have limitations in investigating interactions among many variables and providing in-depth knowledge from large amount of data (Murdoch & Detsky, 2013; Zhao & Luan, 2006). Big data research using EHRs has advantages in discovering new insights and knowledge for CAUTIs from copious data.

To date, no studies have been found predicting hospital-acquired CAUTI using EHR and nurse staffing data through a big data research method. The overall purpose of this study was to develop a predictive model for hospital-acquired CAUTIs using data from multiple data sources. This study used existing EHRs from the University of Minnesota (UMN) Academic Health Center-Information Exchange (AHC-IE) Clinical Data Repository (CDR), NDNQI reported electronic data from the University of Minnesota Medical Center (UMMC), and the list of patients who were diagnosed as CAUTI in ICUs at the UMMC. The data from these multiple data sources were linked to an integrated dataset. The specific aims of this study were:

Aim 1: Create a quality, de-identified dataset combining multiple data sources for machine learning tasks.

Aim 2: Develop and evaluate predictive models to find the best predictive model for hospital-acquired CAUTI.

The study used the knowledge discovery in databases (KDD) approach for data analysis. This big data research analytic method includes a data mining process that uses machine learning techniques to build predictive models. Three data mining techniques were examined and compared in this study: decision trees (DT), logistic regression (LR), and support vector machines (SVM). The evaluation criteria were the number of false negatives (FN), accuracy, sensitivity, specificity, precision, area under the ROC curve, and clinical interpretability. The results demonstrated that the prediction model using DT was the best model for the aim of this study.

In the following sections, the study results for each aims are discussed and conclusions are drawn from the study findings. Strengths and limitations of this study are described, and findings are linked to the conceptual framework of the study. Then implications for future study are discussed.

Significant Findings and Relationship to Existing Literature

Creating an Integrated Dataset

Aim 1 of the study was to create a quality, de-identified dataset combining multiple data sources for machine learning tasks. Three datasets were combined for this study as previously described.

The researcher acquired a limited EHR dataset from the UMN clinical data repository (the UMN AHC-IE dataset), which contained data for patient demographics, problem list, surgical history, diagnosis, admission/discharge date and time, encounter information, lab information, medications, and flowsheets. The data were de-identified and the dataset was placed in the UMN data shelter. The NDNQI reported nurse staffing data (the UMMC-NDNQI dataset) and the list of patients who acquired CAUTIs in the three ICUs (UMMC-ICU CAUTI dataset) were also transferred to this UMN data shelter for mapping.

Data preparation included preprocessing, transformation, and consolidating of data for analysis (Fayyad et al., 1996). As mentioned in previous studies (Pyle, 1999; Zhang, Zhang, & Yang, 2003), data preparation step required a considerable amount of time. Missing, misspelled, and inconsistent data were searched, adjusted, and corrected. To find the optimal format and interval for data transformation, multiple options were tested. Integration of three datasets comprised two steps. First, the data within the UMN AHC-IE and the UMMC-ICU CAUTI datasets were linked using the key variables including patient medical record number (MRN), department ID, and CAUTI diagnosis date. This step used Structured Query Language (SQL) for data extraction and combination. The second step was to map this merged dataset with the UMMC-NDNQI dataset using department ID and NDNQI reporting period. The Python program was used to preprocess the data and create the final integrated dataset for analytic software, Weka.

Because creating an integrated dataset was carried out within the UMN secure data shelter, it was difficult to use outside resources such as additional machine learning packages supporting Python program. Using limited resource for the integration made the experience of data preparation challenging. Also, due to low computing power/capability of the data shelter, the process of data preparation and integration of datasets was slow.

When developing predictive models, cost-sensitive classification method was employed using 200 cost. This method was used to correct imbalanced data (Turney, 1995). The proportion of patients with CAUTI to patients without CAUTI was approximately 1:190, therefore, 200 cost was considered as ideal value. Different costs were also tested, however, 200 cost showed the fewest FNs with highest accuracy possible.

The outcome variable in this study referred to patients who acquired a CAUTI during hospitalization between January 1, 2012, and June 30, 2015, at one of three ICUs – Medical ICU (MICU), Surgical ICU (SICU), and Cardiovascular ICU (CVICU) – at UMMC; whose MRN appeared on the list of patients having CAUTIs in the ICUs; and who did not have a CAUTI/ UTI when admitted.

The factors associated with CAUTI used for analysis from the UMN AHC-IE dataset were the following: age, gender, immunosuppression, glucose lab results, length of stay, Charlson comorbidity index score, hospitalization within 6 previous months, pre-existing catheter, and rationale for continued use of catheter. The factors assoaciated with CAUTI for analysis from the UMMC-NDNQI dataset were total nursing hours per patient day, percent of direct care RNs with associate’s degree nursing, percent of direct care RNs with BSN, MSN, or PhD, and percent of direct care RNs with specialty nursing certification.

Predictive Models

The second aim of this study was to develop and evaluate predictive models to find the best predictive model for hospital-acquired CAUTI. To address this goal, predictive models derived from three machine learning techniques were examined and the results were compared to determine the best predictive model for the hospital-acquired CAUTIs. The evaluation criteria were predictive performance as well as clinical interpretability. Although the SVM model result was difficult to interpret, the model showed high predictive performance. Therefore, SVM model was included in the study to compare the performance with the other two interpretable models.

Both DT and LR models were clinically interpretable, showing associated factors of hospital-acquired CAUTI. The DT model created rules that combine multiple variables, making it clear to interpret relationships among associated factors of CAUTI. The LR model result showed the effect of a single variable, indicating what factors were associated with a higher risk of CAUTI.

Although the results from DT and LR models both had strengths, the DT model showed better performance in three criteria (reduction in FN, higher accuracy and sensitivity) and LR model showed better performance in one criterion (larger ROC curve). Number of FNs and accuracy were important criteria when evaluating the models in this study. Therefore, the DT model was determined as the best model for this study.

Factors Associated with CAUTI

The results from the DT and LR models explain what factors are associated with hospital-acquire CAUTI. The description of factors associated with CAUTI from the DT and LR model are followed.

According to the DT model result, length of stay, gender, rationale for continued use of catheter, total nursing hours per patient day, percent of direct care RNs with specialty nursing certification, and percent of direct care RNs with BSN, MSN, or PhD contributed in modeling. Thus, these factors were related to CAUTI occurrence.

The DT model results demonstrated that female gender, less total nursing hours per patient day (less than 18.71), and a low percent of direct care RNs with specialty nursing certification (less than 12.77%) was related to CAUTI occurrence. If a patient stayed more than 13.5 days, then the patient was likely to have a CAUTI even if the patient had a valid reason for continued use of catheter.

According to LR model result, the factors associated with a higher risk of CAUTI are: young-old and old-old adult, Charlson comorbidity index score ≥ 3, longer length of stay, glucose lab result higher than 200 mg/dl, presence of rationale for continued use of catheter, higher percent of direct care RNs with associate's degree in nursing, and higher percent of direct care RNs with BSN, MSN, or PhD. The factors associated with lower risk of CAUTI are: Young adult, middle-aged adult, male gender, immunosuppression, Charlson index score ≤ 2, hospitalization within previous 6 months, pre-existing urinary catheter, total nursing hours per patient day, and percent of direct care RNs with specialty nursing certification.

An unexpected finding was that higher percent of direct care RNs with BSN, MSN, or PhD was related to CAUTI occurrence. According to DT model result, if a patient was female, had a valid reason for continued use of the catheter, the patient stayed in the hospital between 5.5 and 13.5 days, stayed in a unit where the percent of direct care RNs with BSN, MSN, or PhD was more than 72.15%, and stayed in a unit where total nursing hours per patient day was less than 21.04, then the patient was likely to have a

CAUTI. Perhaps, even though a unit had more than 72.15 % of direct care RNs with BSN or higher education, because total nursing hours per patient day was less than 21.04, the patient in that unit was likely to have a CAUTI. It is also possible that there was limited variation in the higher percent of direct care RNs with BSN, MSN, or PhD, resulting spurious finding. Another possible explanation is that because there were some patients with CAUTI in ICUs, nurses with BSN or higher education were needed to provide better care for patients with CAUTI. If that is the case, then the percent of direct care RNs with BSN, MSN, or PhD is not a risk factor for CAUTI but it is a result from the CAUTI occurrence in ICUs. Therefore, further investigation is needed with a larger sample size of patients with CAUTI.

The results from DT and LR models both found that longer length of stay and presence of rationale for continued use of catheter were associated factors of hospital-acquired CAUTI. Perhaps, although there was a valid reason for continued use of catheter, continued use of catheter contributed to CAUTI occurrence.

The associated factors found from DT and LR models support the findings from previous studies. Previous studies presented that female gender (Tambyah et al., 1999), older age (>50) (Saint & Lipsky, 1999), longer length of stay (Leone et al., 2001), severe underlying disease (Tissot et al., 2001), and RN staffing (Kovner & Gergen, 1998) were the factors associated with CAUTI. This study adds new information that less total nursing hours per patient day, lower percent of direct care RNs with specialty nursing certification, and higher percent of direct care RNs with associate’s degree in nursing contributed to CAUTI occurrence.

Differences Between Patients with and without CAUTI

Using the integrated dataset that was created, the characteristics of those who have CAUTI and those who do not have CAUTI were compared. The results were also compared with the findings from previous studies.

Age. It was found from the literature that older age (>50) was a predictor of CAUTI (Tambyah et al., 1999). CAUTIs in young-old adult group (56 ≤ and ≤ 74) showed the highest frequency, whereas CAUTI in young adult group (18 ≤ and ≤ 35) was least frequent. The LR model showed that young-old and old-old adults had higher risk of CAUTI.

Gender. Many previous studies have found that female gender is a predictor of CAUTI (Nicolle, 2012). The females acquired CAUTIs more than the males, even if there were more male patients. The DT model showed that female fender was related to CAUTI occurrence.

Immunosuppression. Most of the CAUTIs occurred in the population who was not immunosuppressed. This is the opposite of the findings from previous studies (Tissot et al., 2001). Perhaps, this was because the majority of population (90.7%) was not in immunosuppression.

Lab Result – Glucose. More CAUTIs occurred in the population where patients had glucose level higher than 200mg/dl. The LR model indicated that glucose lab result higher than 200 mg/dl was related to higher risk of CAUTI. This result is matched with the finding from previous studies that higher glucose level is associated with CAUTI (Hagerty et al., 2015).

Length of Stay. Previous studies have found that longer length of stay is associated with CAUTI (Leone et al., 2003). Despite the fact that majority of hospitalization was fewer than 20 days and almost half of CAUTIs occurred for hospital stays fewer than 20 days, both DT and LR models showed that longer length of stay was associated with hospital-acquired CAUTI.

Charlson Comorbidity Index Score. More than half of the CAUTI incidents occurred in the population with Charlson Comorbidity Index score ≥ 3. The LR model showed that Charlson comorbidity index score ≥ 3 was related to higher risk for CAUTI. This result is congruent with previous findings that severe underlying illness is associated with CAUTI (Gould et al., 2010).

Hospitalization within 6 Previous Months. More than two thirds of CAUTIs occurred in the population where patients had previous hospitalization within 6 previous months. Previous hospitalization is associated with CAUTI found from previous studies (Johansen et al., 2006).

Pre-existing Catheter. Majority of CAUTI incidents occurred in the population where patients did not have pre-existing indwelling catheter. This may be because most of the population did not have pre-existing catheter.

Rationale for Continued Use of Catheter. The majority of CAUTIs (92.7%) occurred in the population who had a rationale for using a catheter. Although there were more patients who did not have rationale for continued use of catheter, CAUTIs occurred in the population who had rationale. Perhaps, even if a valid reason exists for continued use of catheter, continued/longer use of catheter contributes to CAUTI occurrence. Both

DT and LR models found that presence of rationale for continued use of catheter was related to higher risk of CAUTI.

Linking This Study to Theory

Systems theory was used as the conceptual framework in this study for discovering new knowledge. The theory provided a way of organizing variables, and was not intended for hypotheses testing. Using the conceptual framework for literature review and modeling process allowed firm theoretical background for this study.

Known associated factors of hospital-acquired CAUTIs were categorized into four parts: patient factors, environmental factors, staff factors, interventions. In this study, patient and nurse staff factors were the inputs, hospital environment as the environment, and hospital-acquired CAUTIs as the output. Throughput, or specific interventions, were not measured in this study.

According to the findings of this study, the input, patient factors including age, female gender, glucose lab results > 200 mg/dl, and Charlson comorbidity index score ≥ 3, and nurse staffing factors including total nursing hours per patient day, percent of direct care RNs with specialty nursing certification, percent of direct care RNs with associate’s degree in nursing, and percent of direct care RNs with BSN, MSN, or PhD contributed to the output, CAUTI occurrence. Environmental factors such as longer length of stay and present of rationale for continued use of catheter also contributed to the output, CAUTI occurrence.

Strengths and Limitations of this Study

One of the strengths of this study is the use of combination data from multiple sources. There is yet no research for a predictive model of hospital-acquired CAUTI using a combination of data from EHR and nurse staffing factors collected as part of the NDNQI submission. There were only a few studies in nursing used combination of data, especially EHRs and nurse staffing data. Combination of data from multiple sources enabled interoperable data use in a large scale and is useful to uncover new knowledge. In addition to EHRs and NDNQI submitted data, there are plenty of health data electronically available at national level, such as cancer statistics, national sample survey on RNs, or nursing home compare. This study demonstrated each step for integrating different datasets; therefore, the study can be a good example for future studies that use combined datasets.

Another strength of this study is the use of big data. Big data often encompasses the meaning of three V's including *Volume, Variety,* and *Velocity* (Gandomi & Haider, 2015; Raghupathi & Raghupathi, 2014). *Volume* refers to the size and scale of data, *variety* refers to different formats/types of data, and *velocity* refers to generating rate and analyzing speed of data (Sagiroglu & Sinanc, 2013). This study used large amounts of data stored in the UMN data warehouse, with more than 2.4 million patient data. The forms of data content were diverse; admission date/time, nurse's note, a choice list in flowsheets, NDNQI submitted nurse staffing data, and the list of patients who acquired CAUTIs in ICUs were the examples of various types of data. The use of electronic sensors and health monitors creates data in an unprecedented speed. Real-time health information exchange enabled rapid interactions among care providers and patients. The predictive model developed in this study, when implemented in real setting, can predict or detect a patient at high-risk of CAUTI as early as possible. Immediate response and proper treatment can reduce CAUTI incidents.

Another strength of this study is using cutting-edge analytic methods for knowledge discovery. Big data analytic method using machine learning techniques enabled investigating unknown patterns and relationships among the factors. Since there are many factors associated with hospital-acquired CAUTIs, it is important to look at each factor's association in addition to the interactions of those factors. To date, there are not many studies that conducted nursing research using big data analytics. There is a growing need for nurses to discover new knowledge from the data generated from EHRs, to provide evidence-based care tailored to nursing practice. Also, being the nation's largest health care profession (Health Resources and Services Administration, 2010), the use of EHR and big data analytics for nursing research can enhance visibility of the care that nurses provide.

Also, the data preparation process strengthened this study. Data were thoroughly inspected, cleaned, transformed, and consolidated into an integrated data file for analysis. Although the work required significant amount of time and efforts, the data quality of final integrated dataset was satisfying.

One of the limitations of this study was that the data collection was limited to a single health system. There may be an issue with generalizing the findings of this study in other regions.

Another limitation is skewed distribution of CAUTI patients. Although unbalanced data were corrected using cost-sensitive approach, using more patients with CAUTI would have improved the validity of outcome.

Another limitation was about the data available for research in the UMN data shelter. The EHR in the UMN data shelter contains very large volumes of variable data. While some data were standardized and structured, a substantial amount of data was semi-structured or unstructured. This study did not use unstructured data such as text or provider's notes. Using all the data documented in the EHRs would have provided more in-depth knowledge about hospital-acquired CAUTIs.

Using secure data shelter provided high level of protection, but it also was a limitation. It was difficult to use outside resources such as additional machine learning packages supporting Python program. Also, due to low computing capability of the data shelter, the data preparation and analytic process was slow.

Another limitation is related to diversity of factors related to hospital-acquired CAUTI. Although this study used several variables that were considered as the factors associated with CAUTI, there might have been other factors related to CAUTI. For example, nurses are mostly responsible for urinary catheter care, but physicians can also look at the catheter. Since infection is very sensitive to contact, physician factor (such as rounding frequency) may have influence on CAUTI occurrence. Also, urinary catheter is usually placed in an operating room. But there was no way in this study to investigate whether the placement was aseptic. In hindsight, there are additional variables that could strengthen future studies such as length of time that urinary catheter was in regardless of reason, and length of ICU stay would have provided more information about CAUTI occurrence.

The level of data was another limitation of this study. While patient data were individual level, nurse staffing data from NDNQI dataset were unit level. Individual nurse staffing data in a unit was collected and then unit average was reported to NDNQI. Therefore, a patient's individual data was linked to the unit's average nurse staffing data where the patient stayed. CAUTI is a nursing-sensitive outcome indicator, and is strongly influenced by care provided. It would have been accurate to look at individual nurse's data mapping with patient data. However, nurses shift at least three times a day and a patient receives different care by several nurses. Also, it may take a couple of days to develop CAUTI, which means there may be number of nurses who contributed to CAUTI incident. Therefore, it might have been difficult to track which nurses' characteristics have influenced patient outcome even if nurse's data were individual level.

Implications for Future Study and Translation to the Clinical Setting

Implications for Future Research

The goal of this study was to develop a predictive model for hospital-acquired CAUTIs using EHRs and nurse staffing data from multiple data sources. An integrated dataset was created, predictive models were developed and compared, and factors related to CAUTI occurrence were found. Several suggestions for future studies can improve the efficacy of research process and validity of predictive models.

Future research is required with a larger sample with more patients with CAUTI to verify the results. This study was carried out in a single health system, using limited number of patients with CAUTI. It would have enhanced reliability of the findings if there were more patients with CAUTI.

Time would have been saved if terminologies were standardized in the datasets. For example, department IDs were recorded differently in UMN AHC-IE and UMMC-ICU CAUTI datasets. These IDs were manually converted into standardized format. In case of the variable for *Race*, there were four ways to document “American Indian or Alaska Native”, which meant those four choices were considered as different answers for the *Race*. Searching and correcting unstandardized data slows the process and lowers research efficacy. Collecting and releasing data in a standardized format would greatly improve the quality of data extracted from EHRs for future research.

The quality of NDNQI data was good and ready for research, because the data were cleaned before reporting to NDNQI. Significant amount of time and effort could have been saved if the EHR was preprocessed and integrated with other datasets such as nursing administrative data in a data warehouse, in which the linkages already exist.

Future work is required in predictive modeling using LR. The analytic software, Weka, did not provide confidence intervals (CIs) and p-values for each attribute in LR model. There is a need for future study using different analytic software to look at detailed results of attributes for LR model.

The findings from this study can be extended by future studies. Adding more factors associated with CAUTI, such as aseptic placement techniques or nurse-driven interventions, could provide new information about factors related to CAUTI. Comparing patient outcomes among departments is also possible. Department tailored clinical guidelines would contribute to reduction of CAUTI occurrence.

Research using other nursing-sensitive outcome indicators, such as pain or patient falls, can be another implication for future research. Nursing-sensitive outcome indicators depend on the quantity and quality of care nurses provide. Therefore, patient outcomes studies using these nursing-sensitive indicators would contribute to new knowledge building for nursing research. The NDNQI has standardized, quality nursing data at national level; this dataset can provide new insights about nursing care and nurses' role in health care.

Big data research using clinical data requires multiple expertise including health care, data bases, informatics, and computer science. Multidisciplinary efforts are needed as a team when conducting research using EHR to find the best data-driven model.

I think you need one more paragraph that ties together NMDS (standardized nursing clinical data) NDNQI, NMMDS, and common data elements – this is a big informatics issue for inclusion of nursing data – evident in NDNQI and NMMDS but not in the CDE's.

Implications for Nursing Practice

Findings from this study have significant implications for nursing practice. The study result demonstrated that more direct care RN's possession of specialty nursing certification, lower percent of direct care RNs with associate's degrees in nursing, and more total nursing hours per patient day can contribute to reduction of CAUTI occurrence. The study results support findings from previous studies that nurse specialty certification and nurse’s education at the baccalaureate level or higher are associated with better patient outcomes (Aiken et al., 2003; Bliss et al., 2013; Kendall-Gallagher, Aiken, Sloane, & Cimiotti, 2011; Westra, Bliss, Savik, Hou, & Borchert, 2013; Wilkerson, 2011). Nurses prepared with specialty certification can advocate patient health outcomes. The policies on future nurse workforce planning should aim at promoting baccalaureate or higher nursing education. The findings also suggest that more nursing hours are related to better patient outcome. Therefore, it is important to ensure that the supply of nurse workforce is sufficient for a quality care.

The study has found that even if there were valid reasons for leaving in a urinary catheter, having a catheter more than 48 hours could contribute to CAUTI occurrence. Findings support removing urinary catheter as soon as possible; the American Nurses Association (ANA) suggests that the overuse of urinary catheters should be reduced, and the catheter should be removed per nurse-driven evidence-based protocols (ANA, 2016).

Conclusion

Despite the fact that there are a number of clinical guidelines and studies about hospital-acquired CAUTIs, the rate of CAUTI occurrence is still rising. Review of previous literature addressed a gap in knowledge about the factors associated with hospital-acquired CAUTI. Hospitals are focusing on preventing hospital-acquired CAUTI, as the CMS does not provide payment for hospital-acquired infections anymore. Therefore, there was a need to explore additional factors associated with CAUTI and develop predictive model to detect patients at high risk.

This study developed a predictive model for hospital-acquired CAUTIs using EHR and nurse staffing data from multiple data sources. Research using large amounts of data could provide additional knowledge about hospital-acquired CAUTI. To address the first aim of the study, three datasets were combined into an integrated dataset. All data were electronically available and securely contained in UMN data warehouse. After integrating the datasets, data were cleaned and prepared for analysis. For the second aim of the study, three predictive models were created using the following data mining method: DT, LR, and SVM. The models were evaluated and DT model was determined as the best predictive model for hospital-acquired CAUTI.

The findings from this study have presented factors associated with hospital-acquired CAUTI. The study results demonstrated that female gender, old adult (≥56), Charlson comorbidity index score ≥ 3, longer length of stay, glucose lab result > 200 mg/dl, present of rationale for continued use of catheter, higher percent of direct care RNs with associate’s degree in nursing, less total nursing hours per patient day, and lower percent of direct care RNs with specialty nursing certification was related to CAUTI occurrence.

Implications for future research include the use of different analytic software to investigate detailed results for LR model, adding more factors associated with CAUTI in modeling, using a larger sample with more number of CAUTI patients, and patient outcomes research using nursing-sensitive indicators. This study has important implications for nursing practice. According to the study results, nurse specialty certification, nurse’s education at the baccalaureate level or higher, and more nursing hours per patient day were associated with better patient outcomes. Therefore, considerable efforts are needed to promote possession of nurse specialty certification and higher level of nursing education, as well as enough supply of nursing workforce.

REFERENCES

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. *The Journal of Nursing Administration, 38*(5), 223-229. doi:10.1097/01.NNA.0000312773.42352.d7

Aiken, L. H., Sloane, D. M., Bruyneel, L., Heede, K. V., Griffiths, P., Busse, R., . . . Lesaffre, E. (2014). Nurse staffing and education and hospital mortality in nine european countries: A retrospective observational study. *The Lancet, 383*(9931), 1824-1830.

Aiken, L. H., Smith, H. L., & Lake, E. T. (1994). Lower medicare mortality among a set of hospitals known for good nursing care. *Medical Care,* , 771-787.

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. *Jama, 290*(12), 1617-1623.

Alexaitis, I., & Broome, B. (2014). Implementation of a nurse-driven protocol to prevent catheter-associated urinary tract infections. *Journal of Nursing Care Quality, 29*(3), 245-252. doi:10.1097/NCQ.0000000000000041 [doi]

Al-Habdan, I., Sadat-Ali, M., Corea, J. R., Al-Othman, A., Kamal, B. A., & Shriyan, D. S. (2003). Assessment of nosocomial urinary tract infections in orthopaedic patients: A prospective and comparative study using two different catheters. *International Surgery, 88*(3), 152-154.

ANA. (1999). *Nursing-sensitive quality indicators for acute care settings and ANA's safety & quality initiative* American Nurses Association.

ANA. (2016). ANA CAUTI Prevention Tool. American Nurses Association. Retrieved from http://nursingworld.org/ANA-CAUTI-Prevention-Tool

Andreessen, L., Wilde, M. H., & Herendeen, P. (2012). Preventing catheter-associated urinary tract infections in acute care: The bundle approach. *Journal of Nursing Care Quality*, *27*(3), 209-217.

Appelgren, P., Hellstrom, I., Weitzberg, E., Soderlund, V., Bindslev, L., & Ransjo, U. (2001). Risk factors for nosocomial intensive care infection: A long-term prospective analysis. *Acta Anaesthesiologica Scandinavica*, *45*(6), 710-719.

Askarian, M., Hosseini, R. S., Kheirandish, P., & Memish, Z. A. (2003). Incidence of urinary tract and bloodstream infections in ghotbeddin burn center, shiraz 2000-2001. *Burns*, *29*(5), 455-459.

Bagshaw, S., & Laupland, K. (2006). Epidemiology of intensive care unit-acquired urinary tract infections. *Current Opinion in Infectious Diseases, 19*(1), 67-71. doi:10.1097/01.qco.0000200292.37909.e0

Bakken, S., Hyun, S., Friedman, C., & Johnson, S. (2005). ISO reference terminology models for nursing: Applicability for natural language processing of nursing narratives. *International Journal of Medical Informatics, 74*(7), 615-622.

Balkhy, H. H., Cunningham, G., Chew, F. K., Francis, C., Al Nakhli, D. J., Almuneef, M. A., & Memish, Z. A. (2006). Hospital- and community-acquired infections: A point prevalence and risk factors survey in a tertiary care center in saudi arabia. *International Journal of Infectious Diseases, 10*(4), 326-333.

Barbella, D., Benzaid, S., Christensen, J., Jackson, B., Qin, X., & Musicant, D. (2009). Understanding support vector machine classifications via a recommender system-like approach. Paper presented at the *DMIN,* 305-311.

Beck, J. R., & Shultz, E. K. (1986). The use of relative operating characteristic (ROC) curves in test performance evaluation. *Archives of Pathology & Laboratory Medicine, 110*(1), 13-20.

Bellazzi, R., & Zupan, B. (2008). Predictive data mining in clinical medicine: Current issues and guidelines. *International Journal of Medical Informatics, 77*(2), 81-97. doi:10.1016/j.ijmedinf.2006.11.006

Bernard, M. S., Hunter, K. F., & Moore, K. N. (2012). A review of strategies to decrease the duration of indwelling urethral catheters and potentially reduce the incidence of catheter- associated urinary tract infections. *Urologic Nursing, 32*(1), 29-37.

Beveridge, L. A., Davey, P. G., Phillips, G., & McMurdo, M. E. (2011). Optimal management of urinary tract infections in older people. *Clinical Interventions in Aging, 6*, 173-180.

Blegen, M. A., Goode, C. J., & Reed, L. (1998). Nurse staffing and patient outcomes. *Nursing Research, 47*(1), 43-50.

Bliss, D., Westra, B., Savik, K., & Hou, Y. (2013). Effectiveness of wound, ostomy and Continence–Certified nurses on individual patient outcomes in home health care. *Journal of Wound, Ostomy, and Continence Nursing, 40*(2), 135-142. doi:10.1097/WON.0b013e3182850831

Blodgett, T. J. (2009). Reminder systems to reduce the duration of indwelling urinary catheters: A narrative review. *Urologic Nursing, 29*(5), 369-78; quiz 379.

Blumenthal, D., & Tavenner, M. (2010). The "meaningful use" regulation for electronic health records. *New England Journal of Medicine, 363*(6), 501-504.

Bochicchio, G. V., Joshi, M., Shih, D., Bochicchio, K., Tracy, K., & Scalea, T. M. (2003). Reclassification of urinary tract infections in critically ill trauma patients: A time-dependent analysis. *Surgical Infections, 4*(4), 379-385. doi:10.1089/109629603322761436

Bonadio, M., Pichierri, G., Costarelli, S., Morelli, G., Tanzilli, P., Tartaglia, T., & Mori, S. (2005). Catheter-associated urinary tract infections (CAUTI): Distribution of uropathogens and patterns of antimicrobial resistance in an italian hospital (1996-2003). *Journal of Chemotherapy, 17*(5), 560-562.

Bouza, E., San Juan, R., Munoz, P., Voss, A., Kluytmans, J., & Co-operative Group of the European Study Group on Nosocomial,Infections. (2001). A european perspective on nosocomial urinary tract infections II. report on incidence, clinical characteristics and outcome (ESGNI-004 study). european study group on nosocomial infection. *Clinical Microbiology & Infection, 7*(10), 532-542.

Boyle, D. K., Cramer, E., Potter, C., Gatua, M. W., & Stobinski, J. X. (2014). The relationship between direct-care RN specialty certification and surgical patient outcomes. *AORN Journal, 100*(5), 511-528.

Bradford, J. P., Kunz, C., Kohavi, R., Brunk, C., & Brodley, C. E. (1998). Pruning decision trees with misclassification costs. Paper presented at the *European Conference on Machine Learning*, 131-136.

Breault, J. L., Goodall, C. R., & Fos, P. J. (2005). Data mining a diabetic data warehouse. *Artificial Intelligence in Medicine*, *26*(1), 37-54. doi:10.1016/S0933-3657(02)00051-9

Brosnahan, J., Jull, A., & Tracy, C. (2004). Types of urethral catheters for management of short-term voiding problems in hospitalised adults. *Cochrane Database of Systematic Reviews,* (1), 004013.

Brown, M. L., & Kros, J. F. (2003). Data mining and the impact of missing data. *Industrial Management & Data Systems, 103*(8), 611-621.

Bruminhent, J., Keegan, M., Lakhani, A., Roberts, I. M., & Passalacqua, J. (2010). Effectiveness of a simple intervention for prevention of catheter-associated urinary tract infections in a community teaching hospital. *American Journal of Infection Control, 38*(9), 689-693.

Burton, D. C., Edwards, J. R., Srinivasan, A., Fridkin, S. K., & Gould, C. V. (2011). Trends in catheter-associated urinary tract infections in adult intensive care Units—United states, 1990–2007. *Infection Control & Hospital Epidemiology, 32*(08), 748-756. doi:10.1086/660872

Cardo, D., Dennehy, P. H., Halverson, P., Fishman, N., Kohn, M., Murphy, C. L., & Whitley, R. J. (2010). Moving toward elimination of healthcare-associated infections: A call to action. *Infection Control & Hospital Epidemiology, 31*(11), 1101-1105.

CDC. (2007). *Laboratory procedure manual: Complete blood count.* Centers for Disease Control and Prevention. Retrieved from http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/l25_c_met_complete_blood_count.pdf

CDC. (2015a). Frequently asked questions about catheter-associated urinary tract infections. Retrieved from http://www.cdc.gov/HAI/ca_uti/cauti_faqs.html

CDC. (2015b). *National and state healthcare-associated infections progress report.* Centers for Disease Control and Prevention.

CDC/NHSN. (2016). *Urinary tract infection (catheter-associated urinary tract infection [CAUTI] and non-catheter-associated urinary tract infection [UTI]) and other urinary system infection [USI]) events* . Retrieved from http://www.cdc.gov/nhsn/pdfs/pscmanual/7psccauticurrent.pdf

CDPH. (2016). Healthcare-associated infections (HAIs) definition by california department of public health (CDPH). Retrieved from https://www.cdph.ca.gov/programs/hai/Pages/HealthcareAssociatedInfections.aspx

Chan, J. Y., Semenov, Y. R., & Gourin, C. G. (2013). Postoperative urinary tract infection and short-term outcomes and costs in head and neck cancer surgery. *Otolaryngology - Head & Neck Surgery*, *148*(4), 602-610.

Chang, Y., & Lin, C. (2008). Feature ranking using linear SVM. Paper presented at the *WCCI Causation and Prediction Challenge*, 53-64.

Chant, C., Smith, O. M., Marshall, J. C., & Friedrich, J. O. (2011). Relationship of catheter-associated urinary tract infection to mortality and length of stay in critically ill patients: A systematic review and meta-analysis of observational studies. *Critical Care Medicine*, *39*(5), 1167-1173.

Charlson, M. E., Pompei, P., Ales, K. L., & MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases, 40*(5), 373-383.

Chaudhry, B., Wang, J., Wu, S., Maglione, M., Mojica, W., Roth, E., Morton, S.C., & Shekelle, P.G. (2006). Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. *Annals of Internal Medicine, 144*(10), 742-752.

Chen, Y. Y., Chi, M. M., Chen, Y. C., Chan, Y. J., Chou, S. S., & Wang, F. D. (2013). Using a criteria-based reminder to reduce use of indwelling urinary catheters and decrease urinary tract infections. *American Journal of Critical Care, 22*(2), 105-114.

Chenoweth, C., & Saint, S. (2013). Preventing catheter-associated urinary tract infections in the intensive care unit. *Critical Care Clinics, 29*(1), 19-32.

Chenoweth, C. E., & Saint, S. (2011). Urinary tract infections. *Infectious Disease Clinics of North America, 25*(1), 103-115. doi:10.1016/j.idc.2010.11.005; 10.1016/j.idc.2010.11.005

Cimiotti, J. P., Aiken, L. H., Sloane, D. M., & Wu, E. S. (2012). Nurse staffing, burnout, and health care–associated infection. *American Journal of Infection Control, 40*(6), 486-490.

Cios, K. J., & Kurgan, L. A. (2005). Trends in data mining and knowledge discovery. *Advanced techniques in knowledge discovery and data mining* (pp. 1-26) Springer.

Clarke, K., Tong, D., Pan, Y., Easley, K.A., Norrick, B., Ko, C., Wang, A., Razavi, B., & Stein, J. (2013). Reduction in catheter-associated urinary tract infections by bundling interventions. *International Journal for Quality in Health Care, 25*(1), 43-49.

Clec'h, C., Schwebel, C., Francais, A., Toledano, D., Fosse, J. P., Garrouste-Orgeas, M., . . . OutcomeRea Study, G. (2007). Does catheter-associated urinary tract infection increase mortality in critically ill patients?. *Infection Control & Hospital Epidemiology, 28*(12), 1367-1373.

CMS. (2016). Electronic health records (EHR) incentive programs : Centers for medicare & medicaid services. Retrieved from https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/index.html?redirect=/ehrincentiveprograms/

Conway, L. J., & Larson, E. L. (2012). Guidelines to prevent catheter-associated urinary tract infection: 1980 to 2010. *Heart & Lung, 41*(3), 271-283. doi:10.1016/j.hrtlng.2011.08.001

Conway, L. J., Pogorzelska, M., Larson, E., & Stone, P. W. (2012). Adoption of policies to prevent catheter-associated urinary tract infections in united states intensive care units. *American Journal of Infection Control, 40*(8), 705-710. doi:10.1016/j.ajic.2011.09.020

Cope, M., Cevallos, M. E., Cadle, R. M., Darouiche, R. O., Musher, D. M., & Trautner, B. W. (2009). Inappropriate treatment of catheter-associated asymptomatic bacteriuria in a tertiary care hospital. *Clinical Infectious Diseases*, *48*(9), 1182-1188.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods* Cambridge university press.

Crouzet, J., Bertrand, X., Venier, A. G., Badoz, M., Husson, C., & Talon, D. (2007). Control of the duration of urinary catheterization: Impact on catheter-associated urinary tract infection. *Journal of Hospital Infection*, *67*(3), 253-257.

Curtin, L. (2003). An Integrated Analysis of Nurse Staffing and Related Variables: Effects on Patient Outcomes. *Online Journal of Issues in Nursing*, *8*(3), 118-129.

Dinc, L., & Erdil, F. (2000). The effectiveness of an educational intervention in changing nursing practice and preventing catheter-related infection for patients receiving total parenteral nutrition. *International Journal of Nursing Studies, 37*(5), 371-379.

Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: A methodology review. *Journal of Biomedical Informatics, 35*(5), 352-359. doi:10.1016/S1532-0464(03)00034-0

Duncan, J., Montalvo, I., & Dunton, N. (2011). *NDNQI case studies in nursing quality improvement*. Silver Spring, Maryland: American Nurses Association.

Elpern, E. H., Killeen, K., Ketchem, A., Wiley, A., Patel, G., & Lateef, O. (2009). Reducing use of indwelling urinary catheters and associated urinary tract infections. *American Journal of Critical Care, 18*(6), 535-541.

Esposito, S., Noviello, S., Leone, S., Marvaso, A., Drago, L., Marchetti, F., & LEVT06 Study, G. (2006). A pilot study on prevention of catheter-related urinary tract infections with fluoroquinolones. *Journal of Chemotherapy, 18*(5), 494-501.

Fakih, M. G., Dueweke, C., Meisner, S., Berriel-Cass, D., Savoy-Moore, R., Brach, N., . . . Saravolatz, L. D. (2008). Effect of nurse-led multidisciplinary rounds on reducing the unnecessary use of urinary catheterization in hospitalized patients. *Infection Control and Hospital Epidemiology : The Official Journal of the Society of Hospital Epidemiologists of America, 29*(9), 815-819. doi:10.1086/589584; 10.1086/589584

Falkiner, F. R. (1993). The insertion and management of indwelling urethral catheters--minimizing the risk of infection. *Journal of Hospital Infection, 25*(2), 79-90.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine, 17*(3), 37.

FernáNdez, A., LóPez, V., Galar, M., Del Jesus, M. J., & Herrera, F. (2013). Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches. *Knowledge-Based Systems, 42*, 97-110.

Finkelstein, R., Rabino, G., Kassis, I., & Mahamid, I. (2000). Device-associated, device-day infection rates in an israeli adult general intensive care unit. *Journal of Hospital Infection, 44*(3), 200-205.

Foxman, B. (2002). Epidemiology of urinary tract infections: Incidence, morbidity, and economic costs. *The American Journal of Medicine, 113*(1), 5-13. doi:10.1016/S0002-9343(02)01054-9

Freitas, A. A. (2002). *Data mining and knowledge discovery with evolutionary algorithms.* Springer Science & Business Media.

Fuchs, M. A., Sexton, D. J., Thornlow, D. K., & Champagne, M. T. (2011). Evaluation of an evidence-based, nurse-driven checklist to prevent hospital-acquired catheter-associated urinary tract infections in intensive care units. *Journal of Nursing Care Quality, 26*(2), 101-109. doi:10.1097/NCQ.0b013e3181fb7847 [doi]

Gallagher, R. M., & Rowell, P. A. (2003). Claiming the future of nursing through Nursing-sensitive quality indicators. *Nursing Administration Quarterly, 27*(4), 273-284.

Ganas, S. (2009). Data mining and predictive modeling with excel 2007. *Casualty Actuarial Society Forum*, (Winter)

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*(2), 137-144.

Garner, J., William, J., Emori, T. G., Horan, T., & Hughes, J. (1988). CDC definitions for nosocomial infections, 1988. *American Journal of Infection Control, 16*(3), 128-140. doi:10.1016/0196-6553(88)90053-3

Gartner. Retrieved July 20, 2016 from the Gartner IT Glossary: http://www.gartner.com/it-glossary/big-data/

Gentry, H., & Cope, S. (2005). Using silver to reduce catheter-associated urinary tract infections. *Nursing Standard, 19*(50), 51-54.

Ghezzi, F., Serati, M., Cromi, A., Uccella, S., Salvatore, S., & Bolis, P. (2007). Prophylactic single-dose prulifloxacin for catheter-associated urinary tract infection after tension-free vaginal tape procedure. *International Urogynecology Journal, 18*(7), 753-757.

Giudici, P. (2003). Applied data mining: Statistical Methods for Business and Industry. England: Wiley & Sons Ltd.

Givens, C. D., & Wenzel, R. P. (1980). Catheter-associated urinary tract infections in surgical patients: A controlled study on the excess morbidity and costs. *The Journal of Urology, 124*(5), 646-648.

Gokula, R. M., Smith, M. A., & Hickner, J. (2007). Emergency room staff education and use of a urinary catheter indication sheet improves appropriate use of foley catheters. *American Journal of Infection Control, 35*(9), 589-593.

Gould, C., Umscheid, C., Agarwal, r., Kuntz, G., Pegues, D., & HICPAC. (2010). Guideline for prevention of Catheter-Associated urinary tract infections 2009. *Infection Control and Hospital Epidemiology, 31*(4), 319-326. doi:10.1086/651091

Graves, N., Tong, E., Morton, A. P., Halton, K., Curtis, M., Lairson, D., & Whitby, M. (2007). Factors associated with health care-acquired urinary tract infection. *American Journal of Infection Control, 35*(6), 387-392. doi:http://dx.doi.org/10.1016/j.ajic.2006.09.006

Greene, M. T., Chang, R., Kuhn, L., Rogers, M. A., Chenoweth, C. E., Shuman, E., & Saint, S. (2012). Predictors of hospital-acquired urinary tract-related bloodstream infection. *Infection Control and Hospital Epidemiology, 33*(10), 1001-1007. doi:10.1086/667731

Greiner, R., Grove, A. J., & Roth, D. (2002). Learning cost-sensitive active classifiers. *Artificial Intelligence, 139*(2), 137-174.

Guggenbichler, J. P., Assadian, O., Boeswald, M., & Kramer, A. (2011). Incidence and clinical implication of nosocomial infections associated with implantable biomaterials - catheters, ventilator-associated pneumonia, urinary tract infections. *GMS Krankenhaushygiene Interdisziplinar, 6*(1), Doc18. doi:10.3205/dgkh000175; 10.3205/dgkh000175

Hagerty, T., Kertesz, L., Schmidt, J. M., Agarwal, S., Claassen, J., Mayer, S. A., Larson, E.L., & Shang, J. (2015). Risk factors for catheter-associated urinary tract infections in critically ill patients with subarachnoid hemorrhage. *The Journal of Neuroscience Nursing : Journal of the American Association of Neuroscience Nurses, 47*(1), 51-54. doi:10.1097/JNN.0000000000000111 [doi]

Hakvoort, R. A., Elberink, R., Vollebregt, A., Ploeg, T., & Emanuel, M. H. (2004). How long should urinary bladder catheterisation be continued after vaginal prolapse surgery? A randomised controlled trial comparing short term versus long term catheterisation after vaginal prolapse surgery. *BJOG : An International Journal of Obstetrics and Gynaecology, 111*(8), 828-830. doi:10.1111/j.1471-0528.2004.00181.x

Hampton, S. (2004). Nursing management of urinary tract infections for catheterized patients. *British Journal of Nursing*, *13*(20), 1180.

Hart, S. (2008). Urinary catheterisation. *Nursing Standard, 22*(27), 44-48. doi:10.7748/ns2008.03.22.27.44.c6430

Hashmi, S. H., & Edward, K., Selwyn, R.,Jonathan, G. (2003). Urinary tract infection in surgical patients. *The American Journal of Surgery, 186*(1), 53-56. doi:10.1016/S0002-9610(03)00120-X

Hazelett, S. E., Tsai, M., Gareri, M., & Allen, K. (2006). The association between indwelling urinary catheter use in the elderly and urinary tract infection in acute care. *BMC Geriatrics, 6*, 15.

Health Resources and Services Administration. (2010). *The registered nurse population: Findings from the 2008 national sample survey of registered nurses.* U.S. Department of Health and Human Services.

Hersh, W. R. (2007). Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance. *Clin Pharmacol Ther, 81*, 126-128.

HIMSS. (2016). Electronic health records definition. Retrieved from http://www.himss.org/library/ehr/?navItemNumber=13261

Hooton, T. M., Bradley, S. F., Cardenas, D. D., Colgan, R., Geerlings, S. E., Rice, J. C., Saint, S., Schaeffer, A. J., Tambayh, P. A., Tenke, P., & Nicolle, L. E. (2010). Diagnosis, prevention, and treatment of catheter-associated urinary tract infection in adults: 2009 international clinical practice guidelines from the infectious diseases society of america. *Clinical Infectious Diseases, 50*(5), 625-663.

Horan, T.C., & Gaynes, R.P. (2004). Surveillance of nosocomial infections. In C. G. Mayhall (Ed.), *Hospital epidemiology and infection control* (3rd ed., pp. 1659-1702). Philadelphia: Lippincott Williams & Wilkins.

Horan, T. C., Andrus, M., & Dudeck, M. A. (2008). CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. *American Journal of Infection Control, 36*(5), 309-332. doi:10.1016/j.ajic.2008.03.002; 10.1016/j.ajic.2008.03.002

Hsiao, C., Hing, E., Socey, T. C., & Cai, B. (2011). Electronic health record systems and intent to apply for meaningful use incentives among office-based physician practices: United states, 2001–2011. *System, 18*(17.3), 17.3.

Huang, W. C., Wann, S. R., Lin, S.L., Kunin, C. M., Kung, M. H., Lin, C. H., Hsu, C. W., Liu, C. P., Lee, S.J., Liu, Y. C., & Lai, K. H. (2004). Catheter-associated urinary tract infections in intensive care units can be reduced by prompting physicians to remove unnecessary catheters. *Infection Control & Hospital Epidemiology, 25*(11), 974-978.

Hugonnet, S., Villaveces, A., & Pittet, D. (2007). Nurse staffing level and nosocomial infections: Empirical evaluation of the case-crossover and case-time-control designs. *American Journal of Epidemiology, 165*(11), 1321-1327. doi:kwm041 [pii]

Huth, T. S., Burke, J. P., Larsen, R. A., Classen, D. C., & Stevens, L. E. (1992). Randomized trial of meatal care with silver sulfadiazine cream for the prevention of catheter-associated bacteriuria. *The Journal of Infectious Diseases, 165*(1), 14-18.

International Council of Nurses. (2009). Nursing sensitive outcome indicators. Retrieved from http://www.icn.ch/images/stories/documents/publications/fact_sheets/15c_FS-Nursing_Sensitive_Outcome_Indicators.pdf

IOM. (2001). *Crossing the quality chasm: A new health system for the 21st century* National Academies Press.

IOM. (2010). The future of nursing: Focus on education. Retrieved from http://www.iom.edu/~/media/Files/Report%20Files/2010/The-Future-of-Nursing/Nursing%20Education%202010%20Brief.pdf

Jackson, M., Chiarello, L. A., Gaynes, R. P., & Gerberding, J. L. (2002). Nurse staffing and health care-associated infections: Proceedings from a working group meeting. *American Journal of Infection Control, 30*(4), 199-206.

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis, 6*(5), 429-449.

Jin, R., Breitbart, Y., & Muoh, C. (2009). Data discretization unification. *Knowledge and Information Systems, 19*(1), 1-29.

Johansen, T. E. B., Cek, M., Naber, K. G., Stratchounski, L., Svendsen, M. V., & Tenke, P. (2006). Hospital acquired urinary tract infections in urology departments: Pathogens, susceptibility and use of antibiotics: Data from the PEP and PEAP-studies. *International Journal of Antimicrobial Agents, 28, Supplement 1*(0), 91-107. doi:http://dx.doi.org/10.1016/j.ijantimicag.2006.05.005

Joint Commission. (2009). *Implementation guide for the NQF endorsed nursing-sensitive care measure set.* Oakbrook Terrace, Illinois: The Joint Commission.

Jones, S. S., Adams, J. L., Schneider, E. C., Ringel, J. S., & McGlynn, E. A. (2010). Electronic health record adoption and quality improvement in US hospitals. *The American Journal of Managed Care, 16*(12 Suppl HIT), SP64-71. doi:12788 [pii]

Kaiser, J. (2014). Dealing with missing values in data. *Journal of Systems Integration, 5*(1), 42-51.

Kane, R. L., Shamliyan, T. A., Mueller, C., Duval, S., & Wilt, T. J. (2007). The association of registered nurse staffing levels and patient outcomes: Systematic review and meta-analysis. *Medical Care, 45*(12), 1195-1204. doi:10.1097/MLR.0b013e3181468ca3 [doi]

Karchmer, T. B., Giannetta, E. T., Muto, C. A., Strain, B. A., & Farr, B. M. (2000). A randomized crossover study of silver-coated urinary catheters in hospitalized patients. *Archives of Internal Medicine, 160*(21), 3294-3298.

Katz, D., & Kahn, R. L. (1978). The social psychology of organizations.

Keerasuntonpong, A., Thearawiboon, W., Panthawanan, A., Judaeng, T., Kachintorn, K., Jintanotaitavorn, D., Suddhisanont, L., Waitayapiches, S., Tiangrim, S., & Thamlikitkul, V. (2003). Incidence of urinary tract infections in patients with short-term indwelling urethral catheters: A comparison between a 3-day urinary drainage bag change and no change regimens. *American journal of infection control*, *31*(1), 9-12.

Keim, D., Qu, H., & Ma, K. (2013). Big-data visualization. *IEEE Computer Graphics and Applications*, *33*(4), 20-21.

Kelley, T. F., Brandon, D. H., & Docherty, S. L. (2011). Electronic nursing documentation as a strategy to improve quality of patient care. *Journal of Nursing Scholarship, 43*(2), 154-162.

Kendall-Gallagher, D., & Blegen, M. A. (2009). Competence and certification of registered nurses and safety of patients in intensive care units. *American Journal of Critical Care : An Official Publication, American Association of Critical-Care Nurses*, *18*(2), 106-13; quiz 114. doi:10.4037/ajcc2009487 [doi]

Kendall-Gallagher, D., Aiken, L. H., Sloane, D. M., & Cimiotti, J. P. (2011). Nurse specialty certification, inpatient mortality, and failure to rescue. *Journal of Nursing Scholarship, 43*(2), 188-194.

Klevens, R. M., Edwards, J. R., Richards Jr, C. L., Horan, T. C., Gaynes, R. P., Pollock, D. A., & Cardo, D. M. (2007). Estimating health care-associated infections and deaths in US hospitals, 2002. *Public Health Reports,* , 160-166.

Kohavi, R., & Quinlan, J. R. (2002). Data mining tasks and methods: Classification: Decision-tree discovery. Paper presented at the *Handbook of Data Mining and Knowledge Discovery,* 267-276.

Kohavi, R., & Sommerfield, D. (1995). Feature subset selection using the wrapper method: Overfitting and dynamic search space topology. Paper presented at the *KDD,* 192-197.

Kohn, L. T., Corrigan, J. M., & Donaldson, M. S. (2010). Institute of medicine (2000). to err is human: Building a safer health system.

Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). Data preprocessing for supervised leaning. *International Journal of Computer Science, 1*(2), 111-117.

Kovner, C., & Gergen, P. J. (1998). Nurse staffing levels and adverse events following surgery in US hospitals. *Image: The Journal of Nursing Scholarship, 30*(4), 315-321.

Krein, S. L., Kowalski, C. P., Hofer, T. P., & Saint, S. (2012). Preventing hospital-acquired infections: A national survey of practices reported by U.S. hospitals in 2005 and 2009. *Journal of General Internal Medicine, 27*(7), 773-779. doi:10.1007/s11606-011-1935-y; 10.1007/s11606-011-1935-y

Ksycki, M. F., & Namias, N. (2009). Nosocomial urinary tract infection. *The Surgical Clinics of North America, 89*(2), 475-481. doi:10.1016/j.suc.2008.09.012

Kubler, A., Duszynska, W., Rosenthal, V. D., Fleischer, M., Kaiser, T., Szewczyk, E., & Barteczko-Grajek, B. (2012). Device-associated infection rates and extra length of stay in an intensive care unit of a university hospital in wroclaw, poland: International nosocomial infection control consortium's (INICC) findings. *Journal of Critical Care, 27*(1), 105.e5-105.10.

Kunin, C. M. (2006). Urinary-catheter-associated infections in the elderly. *International Journal of Antimicrobial Agents, 28*(Suppl 1), S78-81.

Kurtzman, E. T., & Corrigan, J. M. (2007). Measuring the contribution of nursing to quality, patient safety, and health care outcomes. *Policy, Politics & Nursing Practice, 8*(1), 20-36. doi:8/1/20 [pii]

Lankshear, A. J., Sheldon, T. A., & Maynard, A. (2005). Nurse staffing and healthcare outcomes: A systematic review of the international research evidence. *ANS.Advances in Nursing Science, 28*(2), 163-174. doi:00012272-200504000-00008 [pii]

Laupland, K. B., Bagshaw, S. M., Gregson, D. B., Kirkpatrick, A. W., Ross, T., & Church, D. L. (2005). *Intensive care unit-acquired urinary tract infections in a regional critical care system*. London: Current Science Ltd. doi:10.1186/cc3023

Laupland, K., Zygun, D., H D Davies, Church, D., Louie, T., & Doig, D. (2002). Incidence and risk factors for acquiring nosocomial urinary tract infection in the critically ill. *Journal of Critical Care, 17*(1), 50-57. doi:10.1053/jcrc.2002.33029

Lavrač, N. (1999). Selected techniques for data mining in medicine. *Artificial Intelligence in Medicine, 16*(1), 3-23.

Lee, J. B. L., & Neild, G. H. (2007). Urinary tract infection. *Medicine, 35*(8), 423-428.

Lee, S.J., Kim, S.W., Cho, Y.H., Shin, W.S., Lee, S.E., Kim, C.S., Hong, S.J., Chung, B.H., Kim, J.J., & Yoon, M.S. (2004). A comparative multicentre study on the incidence of catheter-associated urinary tract infection between nitrofurazone-coated and silicone catheters. *International Journal of Antimicrobial Agents, 24*(Suppl 1), S65-9.

Leone, M., Garnier, F., Avidan, M., & Martin, C. (2004). Catheter-associated urinary tract infections in intensive care units. *Microbes and Infection, 6*(11), 1026-1032. doi:10.1016/j.micinf.2004.05.016

Leone, M., Albanese, J., Garnier, F., Sapin, C., Barrau, K., Bimar, M. C., & Martin, C. (2003). Risk factors of nosocomial catheter-associated urinary tract infection in a polyvalent intensive care unit. *Intensive Care Medicine, 29*(7), 1077-1080.

Leone, M., Garnier, F., Antonini, F., Bimar, M. C., Albanese, J., & Martin, C. (2003). Comparison of effectiveness of two urinary drainage systems in intensive care unit: A prospective, randomized clinical trial. *Intensive Care Medicine, 29*(4), 551-554.

Leone, M., Garnier, F., Dubuc, M., Bimar, M. C., & Martin, C. (2001). Prevention of nosocomial urinary tract infection in ICU patients: Comparison of effectiveness of two urinary drainage systems. *Chest, 120*(1), 220-224.

Leslie, C. S., Eskin, E., & Noble, W. S. (2002). The spectrum kernel: A string kernel for SVM protein classification. Paper presented at the *Pacific Symposium on Biocomputing, , 7*(7) 566-575.

Lichtig, L. K., Knauf, R. A., & Milholland, D. K. (1999). Some impacts of nursing on acute care hospital outcomes. *Journal of Nursing Administration, 29*(2), 25-33.

Liedberg, H., & Lundeberg, T. (1990). Silver alloy coated catheters reduce catheter-associated bacteriuria. *British Journal of Urology, 65*(4), 379-381.

Liedberg, H., Lundeberg, T., & Ekman, P. (1990). Refinements in the coating of urethral catheters reduces the incidence of catheter-associated bacteriuria. an experimental and clinical study. *European Urology, 17*(3), 236-240.

Linder, J. A., Ma, J., Bates, D. W., Middleton, B., & Stafford, R. S. (2007). Electronic health record use and the quality of ambulatory care in the united states. *Archives of Internal Medicine, 167*(13), 1400-1405.

Ling, C. X., & Sheng, V. S. (2011). Cost-sensitive learning. *Encyclopedia of machine learning* (pp. 231-235) Springer.

Lo, E., Nicolle, L., Classen, D., Arias, K.M., Podgorny, K., Anderson, D.J., Burstin, H., Calfee, D.P., Coffin, S.E., Dubberke, E.R., & Fraser, V. (2008). Strategies to prevent Catheter-Associated urinary tract infections in acute care hospitals. *Infection Control and Hospital Epidemiology, 29*(S1), S41-S50. doi:10.1086/591066

Lo, E., Nicolle, L.E., Coffin, S.E., Gould, C., Maragakis, L.L., Meddings, J., Pegues, D.A., Pettis, A.M., Saint, S., & Yokoe, D.S (2014). Strategies to prevent catheter-associated urinary tract infections in acute care hospitals: 2014 update. *Infection Control & Hospital Epidemiology, 35*(S2), S32-S47.

Long, W. J., Griffith, J. L., Selker, H. P., & D'agostino, R. B. (1993). A comparison of logistic regression to decision-tree induction in a medical domain. *Computers and Biomedical Research, 26*(1), 74-97.

Lundeberg, T. (1986). Prevention of catheter-associated urinary-tract infections by use of silver-impregnated catheters. *Lancet, 2*(8514), 1031.

Macfarlane, D. E. (1985). Prevention and treatment of catheter-associated urinary tract infections. *Journal of Infection, 10*(2), 96-106.

Magill, S.S., Edwards, J.R., Bamberg, W., Beldavs, Z.G., Dumyati, G., Kainer, M.A., Lynfield, R., Maloney, M., McAllister-Hollod, L., Nadle, J., & Ray, S.M (2014).

Multistate point-prevalence survey of health care–associated infections. *New England Journal of Medicine, 370*(13), 1198-1208.

Maki, D. G., & Tambyah, P. A. (2001). Engineering out the risk for infection with urinary catheters. *Emerging Infectious Diseases, 7*(2), 342-347. doi:10.3201/eid0702.700342

Maki, D. G. (1997). A prospective, randomized, investigator-blinded trial of a novel nitrofurazone-impregnated urinary catheter. *Infection Control and Hospital Epidemiology*, *18*, 50.

Manojlovich, M., Sidani, S., Covell, C. L., & Antonakos, C. L. (2011). Nurse dose: Linking staffing variables to adverse patient outcomes. *Nursing Research, 60*(4), 214-220. doi:10.1097/NNR.0b013e31822228dc [doi]

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity.

Marra, A. R., Sampaio Camargo, T. Z., Goncalves, P., Sogayar, A. M., Moura, D. F.,Jr, Guastelli, L. R., . . . Edmond, M. B. (2011). Preventing catheter-associated urinary tract infection in the zero-tolerance era. *American Journal of Infection Control, 39*(10), 817-822.

Meddings, J., Rogers, M. A., Macy, M., & Saint, S. (2010). Systematic review and meta-analysis: Reminder systems to reduce catheter-associated urinary tract infections and urinary catheter use in hospitalized patients. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America, 51*(5), 550-560. doi:10.1086/655133; 10.1086/655133

Meddings, J., Saint, S., & McMahon, L. F.,Jr. (2010). Hospital-acquired catheter-associated urinary tract infection: Documentation and coding issues may reduce financial impact of medicare's new payment policy. *Infection Control & Hospital Epidemiology, 31*(6), 627-633.

Meddings, J. A., Reichert, H., Rogers, M. A., Saint, S., Stephansky, J., & McMahon, L. F. (2012). Effect of nonpayment for hospital-acquired, catheter-associated urinary tract infection: A statewide analysis. *Annals of Internal Medicine, 157*(5), 305-312.

Merle, V., Germain, J. M., Bugel, H., Nouvellon, M., Lemeland, J. F., Czernichow, P., & Grise, P. (2002). Nosocomial urinary tract infections in urologic patients: Assessment of a prospective surveillance program including 10,000 patients. *European Urology, 41*(5), 483-489.

Meyer, D., & Wien, F. T. (2015). Support vector machines. *The Interface to Libsvm in Package e1071,*

Meyer, R. M., & O'Brien-Pallas, L. L. (2010). Nursing services delivery theory: An open system approach. *Journal of Advanced Nursing, 66*(12), 2828-2838.

Mnatzaganian, G., Galai, N., Sprung, C. L., Zitser-Gurevich, Y., Mandel, M., Ben-Hur, D., Gurman, G., Klein, M., Lev, A., Levi, L., & Bar-Lavi, Y. (2005). Increased risk of bloodstream and urinary infections in intensive care unit (ICU) patients compared with patients fitting ICU admission criteria treated in regular wards. *The Journal of Hospital Infection*, *59*(4), 331-342. doi:10.1016/j.jhin.2004.07.028

Montalvo, I. (2007). The National Database of Nursing Quality IndicatorsTM (NDNQI®). *OJIN: The Online Journal of Issues in Nursing*, *12*(3).

Montalvo, I. (2008). *The centers for medicare and medicaid services (CMS) hospital value-based purchasing (VBP) program implementation plan*. Silver Spring, MD: American Nurses Association.

Morgan, D. J., Meddings, J., Saint, S., Lautenbach, E., Shardell, M., Anderson, D., . . . SHEA Research, N. (2012). Does nonpayment for hospital-acquired catheter-associated urinary tract infections lead to overtesting and increased antimicrobial prescribing? *Clinical Infectious Diseases, 55*(7), 923-929.

Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *Jama*, *309*(13), 1351-1352.

Muzzi-Bjornson, L., & Macera, L. (2011). Preventing infection in elders with long-term indwelling urinary catheters. *Journal of the American Academy of Nurse Practitioners, 23*(3), 127-134.

NDNQI. (2012). *Guidelines for data collection on the american nurses Association's national quality forum endorsed measures: Nursing care hours per patient day, skill mix, falls, falls with injury.* Overland Park, KS: National Database of Nursing Quality Indicators.

Needleman, J., Buerhaus, P., Mattke, S., Stewart, M., & Zelevinsky, K. (2002). Nurse-staffing levels and the quality of care in hospitals. *New England Journal of Medicine, 346*(22), 1715-1722.

Nicolle, L. E., Bradley, S., Colgan, R., Rice, J. C., Schaeffer, A., & Hooton, T. M. (2005). Infectious diseases society of america guidelines for the diagnosis and treatment of asymptomatic bacteriuria in adults. *Clinical Infectious Diseases, 40*(5), 643-654. Retrieved from http://www.jstor.org/stable/4463100

Nicolle, L. (2012). Urinary catheter-associated infections. *Infectious Disease Clinics of North America, 26*(1), 13-27. doi:10.1016/j.idc.2011.09.009

Niel-Weise, B. S., Arend, S. M., & van den Broek, P. J. (2002). Is there evidence for recommending silver-coated urinary catheters in guidelines? *Journal of Hospital Infection, 52*(2), 81-87.

NIH. (2012). Urinary tract infection in adults- national institute of diabetes and digestive and kidney diseases. Retrieved from http://www.niddk.nih.gov/health-information/health-topics/urologic-disease/urinary-tract-infections-in-adults/Pages/facts.aspx

NNIS. (2000). National nosocomial infections surveillance (NNIS) system report, data summary from january 1992-april 2000, issued june 2000. *American Journal of Infection Control*, *28*(6), 429-448.

NNIS. (2001). National nosocomial infections surveillance (NNIS) system report, data summary from january 1992-june 2001, issued august 2001. *American Journal of Infection Control*, *29*(6), 404-421. doi:10.1067/mic.2001.119952

NNIS. (2003). National nosocomial infections surveillance (NNIS) system report, data summary from january 1992 through june 2003, issued august 2003. *American Journal of Infection Control, 31*(8), 481-498. doi:http://dx.doi.org/10.1016/j.ajic.2003.09.002

NQF. (2004). *National voluntary consensus standards for nursing-sensitive care: An initial performance measure set.* Washington, DC: National Quality Forum.

NQF. (2011). Safe practices for better healthcare: 2010 update. Retrieved from http://www.qualityforum.org/Publications/2010/04/Safe_Practices_for_Better_Healthcare_%E2%80%93_2010_Update.aspx

ODPHP. (2016). National action plan to prevent health care-associated infections: Road map to elimination. Retrieved from http://health.gov/hcq/prevent-hai.asp

Oman, K. S., Makic, M., Fink, R., Schraeder, N., Hulett, T., Keech, T., & Wald, H. (2012). Nurse-directed interventions to reduce catheter-associated urinary tract infections. *American Journal of Infection Control*, *40*(6), 548-553.

Osei-Bryson, K. (2007). Post-pruning in decision tree induction using multiple performance measures. *Computers & Operations Research, 34*(11), 3331-3345.

Owen, R. M., Perez, S. D., Bornstein, W. A., & Sweeney, J. F. (2012). Impact of surgical care improvement project inf-9 on postoperative urinary tract infections: Do exemptions interfere with quality patient care?. *Archives of Surgery*, *147*(10), 946-953.

Pal, N. R., & Jain, L. (2007). *Advanced techniques in knowledge discovery and data mining.* Springer Science & Business Media.

Palmer, J. A., Lee, G. M., Dutta-Linn, M. M., Wroe, P., & Hartmann, C. W. (2013). Including catheter-associated urinary tract infections in the 2008 CMS payment policy: A qualitative analysis. *Urologic Nursing*, *33*(1), 15-23.

Parker, D., Callan, L., Harwood, J., Thompson, D., Webb, M., Wilde, M., & Willson, M. (2009). Catheter-associated urinary tract infections: Fact sheet. *Journal of Wound, Ostomy, and Continence Nursing*, *36*(2), 156-159. doi:10.1097/01.WON.0000347656.94969.99

Parker, D., Callan, L., Harwood, J., Thompson, D., Wilde, M., & Gray, M. (2009). Nursing interventions to reduce the risk of catheter-associated urinary tract infection. part 1: Catheter selection. *Journal of Wound, Ostomy, and Continence Nursing, 36*(1), 23-34. doi:10.1097/01.WON.0000345173.05376.3e

Pellowe, C., & Pratt, R. (2004). Catheter-associated urinary tract infections: Primary care guidelines. *Nursing Times, 100*(2), 53-55.

Penny, K. I., & Chesney, T. (2006). Imputation methods to deal with missing values when data mining trauma injury data. Paper presented at the *Information Technology Interfaces, 2006. 28th International Conference on,* 213-218.

Peschers, U. M., Kempf, V., Jundt, K., Autenrieth, I., & Dimpfl, T. (2001). Antibiotic treatment to prevent urinary tract infections after urodynamic evaluation. *International Urogynecology Journal, 12*(4), 254-257.

Pickard, R., Lam, T., MacLennan, G., Starr, K., Kilonzo, M., McPherson, G., . . . N'Dow, J. (2012). Antimicrobial catheters for reduction of symptomatic urinary tract infection in adults requiring short-term catheterisation in hospital: A multicentre randomised controlled trial. *Lancet, 380*(9857), 1927-1935.

Pickard, R., Lam, T., Maclennan, G., Starr, K., Kilonzo, M., McPherson, G., . . . N'dow, J. (2012). Types of urethral catheter for reducing symptomatic urinary tract infections in hospitalised adults requiring short-term catheterisation: Multicentre randomised controlled trial and economic evaluation of antimicrobial- and antiseptic-impregnated urethral catheters (the CATHETER trial). *Health Technology Assessment (Winchester, England), 16*(47), 1-197.

Platt, R., Polk, B. F., Murdock, B., & Rosner, B. (1982). Mortality associated with nosocomial urinary-tract infection. *New England Journal of Medicine, 307*(11), 637-642.

Platt, R., Polk, B. F., Murdock, B., & Rosner, B. (1986). Risk factors for nosocomial urinary tract infection. *American Journal of Epidemiology, 124*(6), 977-985.

Plowman, R., Graves, N., Esquivel, J., & Roberts, J. A. (2001). An economic model to assess the cost and benefits of the routine use of silver alloy coated urinary catheters to reduce the risk of urinary tract infections in catheterized patients. *Journal of Hospital Infection, 48*(1), 33-42.

Polit, D. F., & Beck, C. T. (2008). *Nursing research: Generating and assessing evidence for nursing practice* (9th edition ed.) Lippincott Williams & Wilkins.

Pratt, R., Pellowe, C., Loveday, H. P., Robinson, N., Smith, G., Barrett, S., . . . Wilcox, M. (2001). The epic project: Developing national evidence-based guidelines for preventing healthcare associated infections. phase I: Guidelines for preventing hospital-acquired infections. department of health (england). *Journal of Hospital Infection, 47 Suppl,* S3-82. doi:10.1053/jhin.2000.0888

Pratt, R. J., Pellowe, C. M., Wilson, J. A., Loveday, H. P., Harper, P. J., Jones, S. R. L. J., McDougall, C., & Wilcox, M. H. (2007). epic2: National evidence-based guidelines for preventing healthcare-associated infections in NHS hospitals in england. *Journal of Hospital Infection, 65*, S1-S59. doi:10.1016/S0195-6701(07)60002-4

Press Ganey Associates, I. (2015). National database of nursing quality indicators: NDNQI. Retrieved from http://www.pressganey.com/solutions/clinical-quality/nursing-quality

Puri, J., Mishra, B., Mal, A., Murthy, N. S., Thakur, A., Dogra, V., & Singh, D. (2002). Catheter associated urinary tract infections in neurology and neurosurgical units. *Journal of Infection, 44*(3), 171-175.

Pyle, D. (1999). Data Preparation for Data Mining (The Morgan Kaufmann Series in Data Management Systems).

Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J. C., . . . Ghali, W. A. (2005). Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical Care,* , 1130-1139.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems, 2*, 3-2501-2-3. eCollection 2014. doi:10.1186/2047-2501-2-3 [doi]

Rasslan, O., Seliem, Z. S., Ghazi, I. A., El Sabour, M. A., El Kholy, A. A., Sadeq, F. M., . . . Rosenthal, V. D. (2012). Device-associated infection rates in adult and pediatric intensive care units of hospitals in egypt. international nosocomial infection control consortium (INICC) findings. *Journal of Infection and Public Health, 5*(6), 394-402.

Rebmann, T., & Greene, L. (2010). Preventing catheter-associated urinary tract infections: An executive summary of the association for professionals in infection control and epidemiology, inc, elimination guide. *American Journal of Infection Control, 38*(8), 644-646. doi:10.1016/j.ajic.2010.08.003

Redon, J., Coca, A., Lazaro, P., Aguilar, M. D., Cabanas, M., Gil, N., . . . Aranda, P. (2010). Factors associated with therapeutic inertia in hypertension: Validation of a predictive model. *Journal of Hypertension*, *28*(8), 1770-1777. doi:10.1097/HJH.0b013e32833b4953 [doi]

Riley, D. K., Classen, D. C., Stevens, L. E., & Burke, J. P. (1995). A large randomized clinical trial of a silver-impregnated urinary catheter: Lack of efficacy and staphylococcal superinfection. *The American Journal of Medicine, 98*(4), 349-356. doi:10.1016/S0002-9343(99)80313-1

Robert, J., Fridkin, S. K., Blumberg, H. M., Anderson, B., White, N., Ray, S. M., . . . Jarvis, W. R. (2000). The influence of the composition of the nursing staff on primary bloodstream infection rates in a surgical intensive care unit. *Infection Control & Hospital Epidemiology, 21*(1), 12-17.

Rosenthal, V. D., Guzman, S., & Orellano, P. W. (2003). Nosocomial infections in medical-surgical intensive care units in argentina: Attributable mortality and length of stay. *American Journal of Infection Control, 31*(5), 291-295.

Rosenthal, V. D., Guzman, S., & Safdar, N. (2004). Effect of education and performance feedback on rates of catheter-associated urinary tract infection in intensive care units in argentina. *Infection Control & Hospital Epidemiology, 25*(1), 47-50.

Safran, C., Bloomrosen, M., Hammond, W. E., Labkoff, S., Markel-Fox, S., Tang, P. C., & Detmer, D. E. (2007). Toward a national framework for the secondary use of health data: An american medical informatics association white paper. *Journal of the American Medical Informatics Association, 14*(1), 1-9.

Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. Paper presented at the *Collaboration Technologies and Systems (CTS), 2013 International Conference on,* 42-47.

Saint, S. (2000). Clinical and economic consequences of nosocomial catheter-related bacteriuria. *American Journal of Infection Control, 28*(1), 68-75. doi:10.1016/S0196-6553(00)90015-4

Saint, S., & Chenoweth, C. E. (2003). Biofilms and catheter-associated urinary tract infections. *Infectious Disease Clinics of North America, 17*(2), 411-432.

Saint, S., Kaufman, S. R., Rogers, M. A., Baker, P. D., Boyko, E. J., & Lipsky, B. A. (2006). Risk factors for nosocomial urinary tract–related bacteremia: A case-control study. *American Journal of Infection Control, 34*(7), 401-407. doi:10.1016/j.ajic.2006.03.001

Saint, S., Kowalski, C.P., Kaufman, S.R., Hofer, T.P., Kauffman, C.A., Olmsted, R.N., Forman, J., Banaszak-Holl, J., Damschroder, L., & Krein, S.L. (2008). Preventing hospital-acquired urinary tract infection in the united states: A national study. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America, 46*(2), 243-250. doi:10.1086/524662; 10.1086/524662

Saint, S., Meddings, J. A., Calfee, D., Kowalski, C. P., & Krein, S. L. (2009). Catheter-associated urinary tract infection and the medicare rule changes. *Annals of Internal Medicine, 150*(12), 877-884.

Saint, S., Kaufman, S. R., Thompson, M., Rogers, M. A., & Chenoweth, C. E. (2005). A reminder reduces urinary catheterization in hospitalized patients. *Joint Commission Journal on Quality and Patient Safety / Joint Commission Resources, 31*(8), 455-462.

Saint, S., & Lipsky, B. A. (1999). Preventing catheter-related bacteriuria: Should we? can we? how? *Archives of Internal Medicine, 159*(8), 800-808.

Schumm, K., & Lam, T. B. (2008). Types of urethral catheters for management of short-term voiding problems in hospitalised adults. *Cochrane Database of Systematic Reviews*, (2), 004013.

Schwartz, D. S., & Barone, J. E. (2006). Correlation of urinalysis and dipstick results with catheter-associated urinary tract infections in surgical ICU patients. *Intensive Care Medicine*, *32*(11), 1797-1801.

Scotia, N. (2010). Explaining odds ratios. *J can Acad Child Adolesc Psychiatry, 19*, 227.

Scott, B. (2010). Clinical and cost effectiveness of urethral catheterisation: A review. *Journal of Perioperative Practice, 20*(7), 235-240.

Scott, R. D. (2009). *The direct medical costs of healthcare-associated infections in US hospitals and the benefits of prevention* Division of Healthcare Quality Promotion National Center for Preparedness, Detection, and Control of Infectious Diseases, Centers for Disease Control and Prevention.

Seago, J. A. (2001). Nurse staffing, models of care delivery, and interventions. *Making Health Care Safer: A Critical Analysis of Patient Safety Practices*, *40*, 423.

Seidl, T., & Kriegel, H. (1998). Optimal multi-step k-nearest neighbor search. Paper presented at the *ACM SIGMOD Record, , 27*(2) 154-165.

Seymour, C. (2006). Audit of catheter-associated UTI using silver alloy-coated foley catheters. *British Journal of Nursing, 15*(11), 598-603.

Shang, J., Stone, P., & Larson, E. (2015). Studies on nurse staffing and health care–associated infection: Methodologic challenges and potential solutions. *American Journal of Infection Control, 43*(6), 581-588.

Shapiro, M., Simchen, E., Izraeli, S., & Sacks, T. G. (1984). A multivariate analysis of risk factors for acquiring bacteriuria in patients with indwelling urinary catheters for longer than 24 hours. *Infection Control : IC, 5*(11), 525-532.

Sheerin, N. S. (2011). Urinary tract infection. *Medicine, 39*(7), 384-389.

Shuman, E. K., & Chenoweth, C. E. (2010). Recognition and prevention of healthcare-associated urinary tract infections in the intensive care unit. *Critical Care Medicine, 38*(8 Suppl), S373-9. doi:10.1097/CCM.0b013e3181e6ce8f; 10.1097/CCM.0b013e3181e6ce8f

Southern, D. A., Quan, H., & Ghali, W. A. (2004). Comparison of the elixhauser and Charlson/Deyo methods of comorbidity measurement in administrative data. *Medical Care, 42*(4), 355-360. doi:00005650-200404000-00008 [pii]

Srinivasan, A., Karchmer, T., Richards, A., Song, X., & Perl, T. M. (2006). A prospective trial of a novel, silicone-based, silver-coated foley catheter for the prevention of nosocomial urinary tract infections. *Infection Control & Hospital Epidemiology, 27*(1), 38-43.

Stamm, W. E. (1975). Guidelines for prevention of catheter-associated urinary tract infections. *Annals of Internal Medicine, 82*(3), 386-390.

Stamm, W. E., & Hooton, T. M. (1993). Management of urinary tract infections in adults. *The New England Journal of Medicine, 329*(18), 1328-1334. doi:10.1056/NEJM199310283291808

Stamm, W. S. (1991). Catheter-associated urinary tract infections: Epidemiology, pathogenesis, and prevention. *The American Journal of Medicine, 91*(3), S65-S71. doi:10.1016/0002-9343(91)90345-X

Stark, R. S., & Dennis M. (1984). Bacteriuria in the catheterized patient. *The New England Journal of Medicine, 311*(9), 560-564. doi:10.1056/NEJM198408303110903

Stone, P. W. (2009). Economic burden of healthcare-associated infections: An american perspective. *Expert Review of Pharmacoeconomics & Outcomes Research, 9*(5), 417-422.

Stone, P. W., Pogorzelska, M., Kunches, L., & Hirschhorn, L. R. (2008). Hospital staffing and health care-associated infections: A systematic review of the literature.

Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America, 47(7), 937-944. doi:10.1086/591696 [doi]

Stovall, R. T., Haenal, J. B., Jenkins, T. C., Jurkovich, G. J., Pieracci, F. M., Biffl, W. L., Barnett, C. C., Johnson, J. L., Bensard, D.D., Moore, E. E., & Burlew, C.C. (2013). A negative urinalysis rules out catheter-associated urinary tract infection in trauma patients in the intensive care unit. *Journal of the American College of Surgeons, 217*(1), 162-166.

Sujijantararat, R., Booth, R. Z., & Davis, L. L. (2005). Nosocomial urinary tract infection: Nursing-sensitive quality indicator in a thai hospital. *Journal of Nursing Care Quality*, *20*(2), 134-139.

Tambyah, P. A. (2015). Prevention and control of catheter associated urinary tract infection. *Journal of Microbiology, Immunology and Infection, 48*(2), S16.

Tambyah, P. A., & Maki, D. G. (2000). Catheter-associated urinary tract infection is rarely symptomatic: A prospective study of 1497 catheterized patients. *Archives of Internal Medicine, 160*(5), 678-682.

Tambyah, P. A., Halvorson, K. T., & Maki, D. G. (1999). A prospective study of pathogenesis of catheter-associated urinary tract infections. *Mayo Clinic Proceedings, 74*(2), 131-136.

Tambyah, P. A., Knasinski, V., & Maki, D. G. (2002). The direct costs of nosocomial catheter-associated urinary tract infection in the era of managed care. *Infection Control & Hospital Epidemiology, 23*(1), 27-31.

Tambyah, P. A., & Maki, D. G. (2000). The relationship between pyuria and infection in patients with indwelling urinary catheters: A prospective study of 761 patients. *Archives of Internal Medicine, 160*(5), 673-677.

Tambyah, P. A., & Oon, J. (2012). Catheter-associated urinary tract infection. *Current Opinion in Infectious Diseases, 25*(4), 365-370. doi:10.1097/QCO.0b013e32835565cc; 10.1097/QCO.0b013e32835565cc

Tenke, P., Kovacs, B., Johansen, T., Matsumoto, T., Tambyah, P., & Naber, k. (2008). European and asian guidelines on management and prevention of catheter-associated urinary tract infections. *International Journal of Antimicrobial Agents, 31*, 68-78. doi:10.1016/j.ijantimicag.2007.07.033

Thibon, P., Le Coutour, X., Leroyer, R., & Fabry, J. (2000). Randomized multi-centre trial of the effects of a catheter coated with hydrogel and silver salts on the incidence of hospital-acquired urinary tract infections. *Journal of Hospital Infection, 45*(2), 117-124.

Tissot, E., Limat, S., Cornette, C., & Capellier, G. (2001). Risk factors for catheter-associated bacteriuria in a medical intensive care unit. *European Journal of Clinical Microbiology & Infectious Diseases, 20*(4), 260-262.

Tissot, E., Woronoff-Lemsi, M. C., Cornette, C., Plesiat, P., Jacquet, M., & Capellier, G. (2001). Cost-effectiveness of urinary dipsticks to screen asymptomatic catheter-associated urinary infections in an intensive care unit. *Intensive Care Medicine, 27*(12), 1842-1847.

Topal, J., Conklin, S., Camp, K., Morris, V., Balcezak, T., & Herbert, P. (2005). Prevention of nosocomial catheter-associated urinary tract infections through computerized feedback to physicians and a nurse-directed protocol. *American Journal of Medical Quality, 20*(3), 121-126.

Trautner, B. W., & Darouiche, R. O. (2004). Role of biofilm in catheter-associated urinary tract infection. *American Journal of Infection Control, 32*(3), 177-183.

Trautner, B. W., Hull, R. A., & Darouiche, R. O. (2005). Prevention of catheter-associated urinary tract infection. *Current Opinion in Infectious Diseases, 18*(1), 37-41.

Tsuchida, T., Makimoto, K., Ohsako, S., Fujino, M., Kaneda, M., Miyazaki, T., Fujiwara, F., & Sugimoto, T. (2008). Relationship between catheter care and catheter-associated urinary tract infection at japanese general hospitals: A prospective observational study. *International Journal of Nursing Studies, 45*(3), 352-361.

Turney, P. D. (1995). Cost-sensitive classification: Empirical evaluation of a hybrid genetic decision tree induction algorithm. *Journal of Artificial Intelligence Research, 2*, 369-409.

Umscheid, C. A., Umscheid, C. A., Doshi, J. A., Agarwal, R., Williams, K., & Brennan, P. J. (2011). Estimating the proportion of healthcare-associated infections that are reasonably preventable and the related mortality and costs. *Infection Control & Hospital Epidemiology, 32*(02), 101-114.

Unruh, L. (2003). Licensed nurse staffing and adverse events in hospitals. *Medical Care, 41*(1), 142-152.

van der Kooi, T. I., de Boer, A. S., Mannien, J., Wille, J. C., Beaumont, M. T., Mooi, B. W., & van den Hof, S. (2007). Incidence and risk factors of device-associated infections and associated mortality at the intensive care in the dutch surveillance system. *Intensive Care Medicine, 33*(2), 271-278.

Verleyen, P., De Ridder, D., Van Poppel, H., & Baert, L. (1999). Clinical application of the bardex IC foley catheter. *European Urology, 36*(3), 240-246.

Wagenlehner, F. M., Cek, M., Naber, K. G., Kiyota, H., & Bjerklund-Johansen, T. E. (2012). Epidemiology, treatment and prevention of healthcare-associated urinary tract infections. *World Journal of Urology, 30*(1), 59-67. doi:10.1007/s00345-011-0757-1; 10.1007/s00345-011-0757-1

Wang, F., Xing, T., Li, J., He, Y., Bai, M., & Wang, N. (2013). Survey on hospital-acquired urinary tract infection in neurological intensive care unit. *APMIS, 121*(3), 197-201.

Warren, J. (2001). Catheter-associated urinary tract infections. *International Journal of Antimicrobial Agents, 17*(4), 299-303. doi:10.1016/S0924-8579(00)00359-9

Watzlaf, V. J., Zeng, X., Jarymowycz, C., & Firouzan, P. A. (2004). Standards for the content of the electronic health record. *Perspectives in Health Information Management / AHIMA, American Health Information Management Association, 1*, 1.

Weaver, C. A., Delaney, C. W., Weber, P., & Carr, R. L. (2010). Nursing informatics for the 21st century: An international look at practice, education, and EHR trends. HIMSS.

Weiskopf, N. G., & Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment: Enabling reuse for clinical research. *Journal of the American Medical Informatics Association : JAMIA, 20*(1), 144-151. doi:10.1136/amiajnl-2011-000681 [doi]

Westra, B., Bliss, D., Savik, K., Hou, Y., & Borchert, A. (2013). Effectiveness of wound, ostomy, and continence nurses on agency-level wound and incontinence outcomes in home care. *Journal of Wound, Ostomy, and Continence Nursing, 40*(1), 25-53. doi:10.1097/WON.0b013e31827bcc4f

Westra, B., Dey, S., Fang, G., Steinbach, M., Kumar, V., Oancea, C., . . . Dierich, M. (2011). Interpretable predictive models for knowledge discovery from home-care electronic health records. *Journal of Healthcare Engineering, 2*(1), 55-74.

Wilkerson, B. L. (2011). Specialty nurse certification effects patient outcomes. *Plastic Surgical Nursing : Official Journal of the American Society of Plastic and Reconstructive Surgical Nurses, 31*(2), 57-59. doi:10.1097/PSN.0b013e31821b5a9c [doi]

Willson, M., Wilde, M., Webb, M. L., Thompson, D., Parker, D., Harwood, J., . . . Gray, M. (2009). Nursing interventions to reduce the risk of catheter-associated urinary tract infection: Part 2: Staff education, monitoring, and care techniques. *Journal of Wound, Ostomy, and Continence Nursing, 36*(2), 137-154. doi:10.1097/01.WON.0000347655.56851.04

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques*. Burlington, MA: Elsevier.

Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied Artificial Intelligence, 17*(5-6), 375-381.

Zhao, C. M., & Luan, J. (2006). Data mining: Going beyond traditional statistics. *New Directions for Institutional Research, 2006*(131), 7-16.

Zhu, W., Zeng, N., & Wang, N. (2010). Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS® implementations. *NESUG Proceedings: Health Care and Life Sciences, Baltimore, Maryland,* , 1-9.