# Statistical Modeling Approaches and User-Centered Design for Nursing Decision Support Tools Predicting In-Hospital Cardiopulmonary Arrest

By

Alvin Dean Jeffery

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
in
Nursing Science

May, 2017

Nashville, Tennessee

Approved:
Lorraine C. Mion, PhD, RN, FAAN
Mary S. Dietrich, PhD
Betsy Kennedy, PhD, RN
Laurie L. Novak, PhD

To my husband, Jamey, for his unwavering support during all life’s adventures

## ACKNOWLEDGEMENTS

I would like to thank my entire dissertation committee for sharing their wisdom and experience with me as I transition to the role of a scientist. Mary Dietrich, Betsy Kennedy, and Laurie Novak are such brilliant and kind people, and I don't believe a better dissertation committee has ever existed! I especially want to thank Lorraine Mion, my adviser since the beginning of the program, for her patience, mentorship, and commitment to my success.

Others who directly assisted in this dissertation work include Daniel Fabbri and Joseph Coco from Vanderbilt University's Department of Biomedical Informatics - they were essential in gathering data for the statistical prediction model study. I also acknowledge and appreciate the time and input of the nurses who served as my research participants.

I'd like to thank the Agency for Healthcare Research and Quality for allowing us to modify their video vignette for one of my studies. I'd also like to thank Dr. Sally Miller and the Vanderbilt University School of Nursing Simulation Center staff for allowing me to use their laboratory space and provide personnel support, and Vanderbilt University's Center for Research and Innovation in Systems Safety for input on prototype design.

The VA Quality Scholars faculty and fellows with whom I have worked over the past 3 years have been crucial in shaping my research and career trajectories. The mentoring sessions, work-in-progress presentations, and the opportunity to regularly engage in scholarly work and discussion have been invaluable.

Finally, this work would not have been possible without the financial support of Vanderbilt University's CTSA award No. UL1TR000445 from the National Center for Advancing Translational Sciences as well as resources and the use of facilities at the VA Tennessee Valley Healthcare System. Notably, the contents of this work are solely the responsibility of the authors and do not necessarily represent official views of the National Center for Advancing Translational Sciences, the National Institutes of Health, the Department of Veterans Affairs, or the United States Government.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

## LIST OF ABBREVIATIONS

AUC Area Under the Curve

CDS Clinical Decision Support

CPA Cardiopulmonary Arrest

CPT Current Procedural Terminology

EHR Electronic Health Record

ICD International Classification of Diseases

ICU Intensive Care Unit

PB-CDS Probability-Based Clinical Decision Support

ROC Receiver Operating Characteristic

RRT Rapid Response Team

# Chapter 1

# INTRODUCTION

## 1.1 Overview

This doctoral research explored strategies for the design and statistical development of probability-based nursing decision support tools within the clinical context of in-hospital cardiopulmonary arrest (IHCPA). IHCPA remains a harmful and costly event, and recent attempts to assist with early recognition via probability-based clinical decision support (PB-CDS) tools have fallen short of improving patient outcomes. These shortcomings are due, in part, to the complex nature of PB-CDS tools with inadequate attention paid to important design elements during the early stages of the tools' construction.[1] Failure to improve patient outcomes may also be a condition of the PB-CDS tools' underlying statistical assumptions. Thus, this paucity of evidence provided an opportunity to examine aspects of PB-CDS tools that influence clinician's decision making which in turn could impact patient outcomes.

## 1.2 Significance

With widespread implementation of electronic health records (EHR) in the last decade,[2] the emergence of extremely large datasets and the accompanying growth in statistical processing capabilities (referred to as *big data*) have provided researchers and clinicians the ability to answer new questions.[3] Predictive analytics are one application gaining popularity for addressing patient care challenges.[4] The purpose of predictive analytics is to collect and analyze big data in real-time while providing end-users with a probability of a particular outcome (e.g., hospital readmission, acute decompensation, or adverse drug events).[4] Although statistical outputs of these predictive analytic models can be highly accurate, nurses' perceptions and information display preferences of this new information are relatively unknown.

Using predictive analytics to influence clinical decision-making is a relatively new phenomenon; therefore, an initial step was to study a clinical event where a predicted likelihood is provided to the clinician at a point in time when prompt action would be warranted. In addition to facilitating real-time feedback to the clinician, using an outcome where the anticipated and actual events occur close together minimizes the potential influence of additional variables (e.g., other clinicians' actions or non-hospital factors) on explaining the connection between the probability and actual occurrence of events. Therefore, using currently available data to predict events likely to occur within 24-48 hours is ideal for the in-hospital nurse. The risk of in-hospital cardiopulmonary arrests (IHCPA) meets this criterion and served as the outcome for these studies. With 209,000 patients experiencing an IHCPA every year in the U.S.[5] and approximately one-third having clinically significant disability (10% with severe disability) following the event,[6] this is a clinically significant outcome that, if improved, could have an impact on thousands of lives.

### 1.2.1 Significance of Predictive Analytics

#### *1.2.1.1 History of Predictive Analytics*

Clinicians have always used current and historical patient data to predict future events. As technology advanced, applying objective methods to diagnostic and prognostic activities via the use of laboratory studies[7] became possible. The results of these studies, along with patient interviewing and physical assessments, have been used to identify diseases and disorders as well as potential treatment options and prognoses. Many believe the use of predictive analytics is the next step in expanding the clinician's toolkit because it provides a new dimension of information that can easily be analyzed from available data.[8,4,9] Most healthcare prediction model strategies have leveraged classification approaches where the outcome is a binary event,[10] but the optimal statistical approaches to embed within decision support tools and assist clinicians with recognition are still being identified[11] and need further exploration.

#### 1.2.1.2 Benefits of Predictive Analytics

A benefit of predictive analytics (also referred to as clinical prediction models) is the ability to aggregate several cues into one cue for the clinician to assess.[12] Instead of clinicians exposed to dozens or even hundreds of relevant variables, the computer can aggregate this information into clinically relevant information for clinicians. This has the potential to impact prioritization of care, allocation of organizational resources, and education of less experienced clinicians. Recently graduated clinicians, in particular, do not frequently have sufficient experiences to develop what many call an intuition or gut feeling.[13] Pattern recognition is a decision-making style used by many experts.[14] Predictive analytics performs the pattern matching for the user and has the potential to identify patterns of which clinicians are not yet aware.

#### 1.2.1.3 Shortcomings of Predictive Analytics

The addition of another technological tool to inform clinicians' decision making may not improve *patient* outcomes. The increasing complexity of the healthcare environment, paired with the rapid collection and availability of information, has made decision-making in today's clinical environment more challenging than it was previously. The increasing frequency of new technology incorporated into the clinical setting is one factor that has contributed to this complexity.[15,16] Although many technologies are developed to aid clinicians in identifying important changes in patient conditions (i.e., enhance the signal), the wide variety of user interfaces and poor interoperability may impede clinicians' ability to cognitively process and appropriately use the data in caring for patients (i.e., too much noise).[17]

Furthermore, relatively few nursing-focused decision support studies have been published,[18,19,1] and some evidence suggests nurses and physicians require distinct informatics interventions to support their unique decision-making styles.[20] This warrants further exploration within the field of nursing decision support, including whether different nursing roles (e.g., bedside nurses, charge nurses) use different decision-making styles and need separate decision support tools. **How technology influences clinical decision making (part of a field known as cognitive informatics[21]) and the potential impact of incorporating predictive modeling into the clinical environment needs to be explored.**

### 1.2.2 Significance of In-Hospital Cardiopulmonary Arrest

According to a national United States database (Get With the Guidelines-Resuscitation), approximately 209,000 people are treated for in-hospital cardiac arrest every year.[5] Data in 2011 revealed survival rates for in-hospital arrests was 23% for adults and 37% for children.[22] The cost of surviving an IHCPA is estimated to be at least $63,000 (excluding long-term care facility needs),[23] which given a roughly 30% survival rate,[22] has crude cost estimates of $4 billion per year in the U.S.

In-hospital cardiopulmonary arrests occurring outside of the intensive care unit (ICU) are of particular interest because these in-hospital events might be preventable.[24] The use of telemetry monitoring has been shown to be beneficial in survival of patients having out-of-ICU arrests,[25] and registry information available in 2012 revealed that approximately 87% of adults and 90% of children were being monitored before an IHCPA occurred.[22] Therefore, it is logical that use of already-available monitoring equipment could promote early identification of patients at risk for IHCPA. Given that being in an ICU during the onset of the IHCPA increases likelihood of survival,[26,27] a reduction in mortality might be possible if high-risk, non-ICU patients could be both identified and transferred before the IHCPA occurs. If real-time telemetry data were combined with additional data in the EHR (e.g., laboratory values and physical assessments), the use of predictive analytics could provide information about the likelihood of this important patient outcome.

## 1.3 Related Work

While the number of publications discussing predictive analytics continues to increase, most are focused on development of the statistical models, the variables included in the models, and their statistical performance.[28,10,29,30,31] These factors are necessary but insufficient to influence patient outcomes because a change in clinician behavior is also required for patient care to be impacted. In healthcare, risk scores comprised primarily of less than a dozen variables for outcomes such as pressure ulcers,[32] cardiopulmonary arrests,[33] and falls[34] have been developed and widely used. Although many published reports describe the statistical performance of healthcare predictive analytic models,[28,10] including those that predict CPAs,[35,36,37,38,39,40] these models have not significantly impacted patient outcomes other than a modest improvement in length of stay.[41]

Several studies have been limited by their lack of measuring whether clinicians decide to take action once the prediction model identifies a likely IHCPA[42,1] (i.e., treatment fidelity), and most studies progress directly from model development to implementation in the clinical environment without adequate preliminary testing. It has also been recognized that beginning with user-centered, iterative design changes in a controlled environment is safer than exposing patients and clinicians to potentially high-risk interventions.[8]

This dissertation work took a "small ball" approach[43] to developing an information resource and explored the user interface in a simulated environment. This approach permits assessment of clinician preferences as well as minor modifications of predictive analytics tools to be made before significant resources have been spent on tool implementation. Big data applications for healthcare are continuously emerging;[4,3] however, the analysts capable of developing big data models and outputs might not have the necessary expertise in clinical decision-making or human-computer interaction essential for integrating this information into clinicians' workflows.[44,45] The integration of big data outputs and clinician decision-making processes has not yet been fully addressed; thus, this dissertation work (a) explored nurses' information gathering processes and problem-solving strategies that may be augmented by a predictive analytics tool, (b) compared four analytical strategies for development and validation of a predictive analytics tool, and (c) explored information preferences among key nurse roles (bedside nurse, charge nurse, and rapid response team nurse) in the design of a predictive analytics tool.

## 1.4 Specific Aims

This dissertation research was conducted as three smaller studies: (1) qualitative interviews to explore nurses' information gathering, problem-solving strategies, and perceptions of risk, (2) quantitative development of IHCPA prediction models, and (3) participatory design sessions to identify and compare nurses' information preferences in the design of PB-CDS tools.

The specific aims were to:

1. **Explore hospital nurses' information gathering processes, problem-solving strategies, and perceptions of risk.**
   (a) Explore bedside nurses' recognition processes that precede either: (i) activation of a rapid response [medical emergency] team or (ii) IHCPA of a patient.
   (b) Explore charge nurses' information gathering processes and problem-solving strategies for assigning nurses to patients (particularly, those at risk for rapid deterioration).
   (c) Describe bedside nurses' and charge nurses' perceptions of, and ability to express, the following concepts pertaining to rapid deterioration: (i) risk, (ii) uncertainty, and (iii) probability.
2. **Compare the performance of logistic regression, survival analysis, random**

**forest, and random survival forest in the development and validation of a dynamic predictive model for in-hospital cardiopulmonary arrest based on patient characteristics.**

3. **Describe the similarities/differences of information preferences among bedside nurses, charge nurses, and rapid response team nurses during the design of an IHCPA decision support tool.**

#### 1.4.1 Aim 1: Nurses' Information Gathering Processes, Problem-Solving Strategies, and Perceptions of Risk

To understand how nurses gather information, individual interviews with bedside nurses were conducted to explore events preceding the decision to escalate care (e.g., rapid response team [RRT] activation) for a patient deemed at risk for IHCPA. Although barriers to initiation of a RRT are well identified in the literature,[46] little is known about individual facilitators and activities that promote its early activation. A specific aim (1a) was to explore bedside nurses' recognition processes that precede either: (a) activation of a rapid response [medical emergency] team or (b) IHCPA of a patient.

While the benefit of accurately predicting an IHCPA might assist bedside nurses in determining if/when to escalate individual patient care, other factors could also influence this decision, notably the number of patients to whom the nurse is assigned and the acuity of those patients. Although bedside nurses spend the most amount of time with patients, they have relatively little influence on the allocation of staff that could influence the disposition of high-risk patients. Therefore, charge nurses were also included in this study. Charge nurses are responsible for reviewing patient census, available staff, and assigning patients to nurses based on the match of nurses' experience with patient acuity. Thus, a specific aim (1b) was to explore charge nurses' information gathering processes and problem-solving strategies for assigning nurses to patients (particularly, those at risk for rapid deterioration).

Because nurses typically have little or no experience with predictive analytics for adverse patient events, they could not directly be asked about this topic. Therefore, perceptions of the concepts of risk, uncertainty, and probability were explored along with how participants expressed these concepts to others in both oral or written forms. This provided a proxy for understanding the potential implications of integrating predictive analytics into patient care processes. The final specific aim (1c) was to describe bedside nurses' and charge nurses' perceptions of, and ability to express, the following concepts pertaining to rapid deterioration: (a) risk, (b) uncertainty, and (c) probability.

### 1.4.2 Aim 2: Comparison of Modeling Strategies for Development and Validation of IHCPA Prediction Models

The optimal statistical approaches to embed within decision support tools and assist clinicians with recognition are still being identified.[11] Most statistical approaches are simply classification models that attempt to identify how likely it is an event will occur.[10] Researchers studying this phenomenon have focused on increasingly accurate models, but accuracy is not the only important feature of a statistical method's performance. Prediction model interpretability (e.g., single probability versus probability trends over time) is an important feature in which additional research might produce a better understanding of how to successfully implement models into the clinical environment. For hospital-based nurses, identifying when an event is likely to occur (or at least monitoring trends over time) might be equally important to the simple prediction of whether an event will occur at all. Four different prediction model strategies comprising classification and time-to-event outcomes along with traditional statistical approaches and machine learning methods were compared. The models' accuracy and discrimination were examined for statistical comparison. The models' alert frequency and numeric output were examined for clinical impact and interpretability, respectively. This aim was to compare the performance of logistic regression, survival analysis, random forest, and random survival forest in the development and validation of a dynamic predictive model for in-hospital cardiopulmonary arrest based on patient characteristics.

### 1.4.3 Aim 3: Nurses' Information Preferences for the Design of an IHCPA Decision Support Tool

Input from clinician users during the design phase of decision support tool user interfaces has been reported to increase the likelihood of tool adoption.[47] Co-creation of a decision support tool via participatory design can be beneficial because the active involvement of participants helps identify important design concepts that groups of participants (i.e., researchers and end-users) might not identify in isolation. Many researchers have developed decision support tools for nurses that are probability-based and/or cardiopulmonary arrest focused, but to our knowledge, the participatory design method has never been published for this use.

Just as the information needs of nurses differ from physicians,[20] it was hypothesized that the needs of various specialties of nurses might differ, given their different settings and work. C. M. Johnson and J. P. Turley[20] demonstrated that nurses tend to benefit from information displays focused on trends and the recall of relevant patient information while physicians benefit from displays that promote inference for decision-making. This study was designed to explore whether similar findings might hold true when exploring preferences among bedside nurses, charge nurses, and rapid response team nurses during the design of an IHCPA decision support tool.

## 1.5 Conceptual Framework

Works from human factors engineering (Carayon),[48] clinical decision support system rule development (Brokel),[49] and information technology acceptance theories (Venkatesh)[50] were used to develop a conceptual model (see Figure 1.1) for guiding the studies. The primary variables include a focus on the **Technology Characteristics** (e.g., user interface) and **Scientific Evidence** (e.g., predictor variable selection) of a **Clinical Decision Support System** (specifically, the processing and output of data) and its resulting impact on **Clinician Behaviors**, which mediate Patient Outcomes. It is also recognized that Organizational Characteristics (e.g., culture, capital resources) influence the Clinical Decision Support System and Clinician Behaviors, while Clinician Characteristics (e.g., role, experience level, education) influence both a Clinical Decision Support System's Data Input (i.e., documentation) as well as Clinician Behaviors. Environmental Characteristics (e.g., noise level, lighting) and Patient Characteristics (e.g., acuity, non-modifiable risk factors) can also influence data input, clinician behaviors, and patient outcomes.

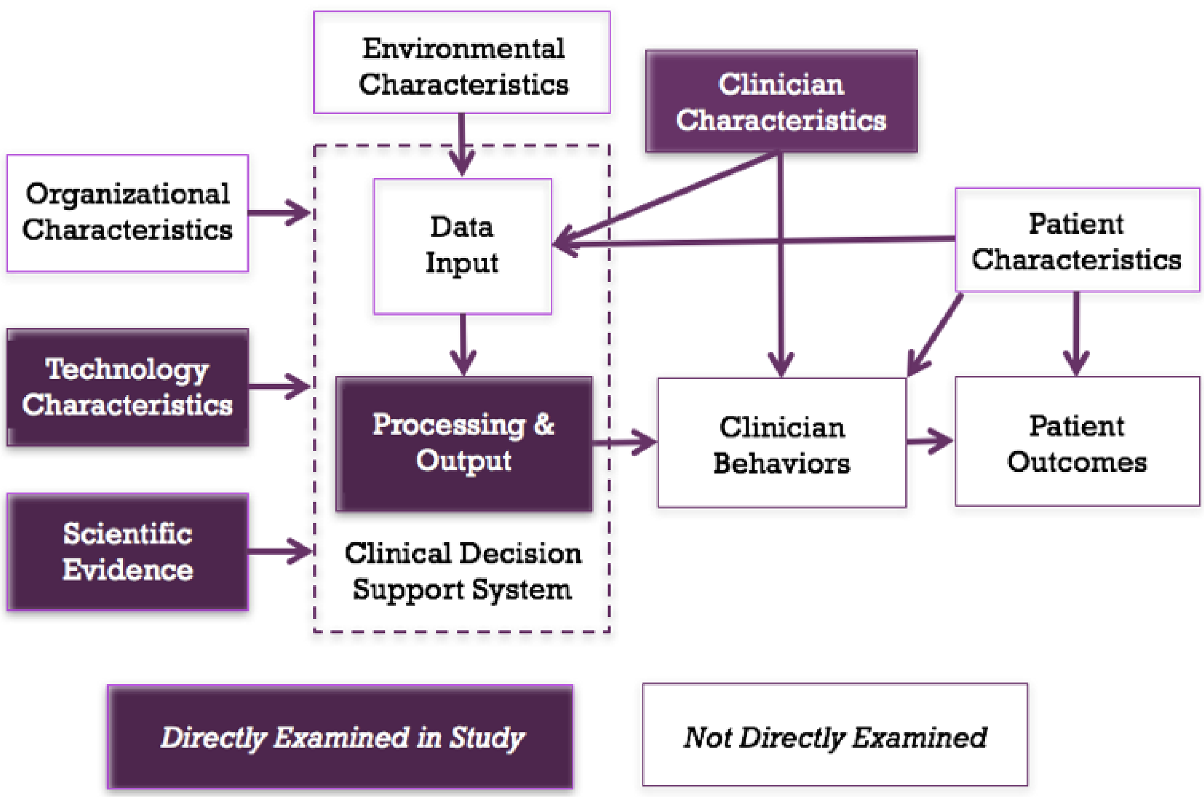



Figure 1.1: Conceptual framework.

## 1.6 Dissertation Chapters

Consistent with the Specific Aims, the following three chapters of the dissertation describe: (1) the exploration of nurses' perceived information gathering behaviors related to IHCPA, (2) a comparison of statistical modeling strategies for IHCPA, and (3) the identification of design elements important for PB-CDS tools. The final chapter is a summary of my research trajectory given the results of the dissertation research.

Chapter 2

# A QUALITATIVE EXPLORATION OF BIG DATA APPLICATIONS FOR NURSING DECISION SUPPORT TOOLS

This chapter describes work done to describe nurses' perceptions of risk-related terms and identify current information gathering behaviors related to recognition of IHCPA. The results were used as preliminary work for the qualitative study reported in Chapter 4.

## 2.1 Background and Significance

With the last decade of widespread implementation of electronic health records (EHR), the rapid analysis of large datasets from a variety of sources using complex computational methods has opened the door to exploration of many new research and clinical questions. Although determining the most important questions to ask of this big data is a frequent topic of discussion,[51,3] predicting outcomes, such as high-cost patients, readmissions, triage, acute decompensation, adverse events, and treatment optimization, through the application of predictive analytics has been gaining popularity.[4]

The statistical outputs of these predictive models can be highly accurate, but little is known about how nurses perceive this type of information and how they might act upon it. Nurses are experienced using information about a patient's history and current condition (e.g., physical assessments, laboratory values), but information predicting a *future* outcome for a specific patient is a novel addition to his/her toolkit. Assuming we can leverage the big data at our disposal, the simple addition of another technological tool may or may not improve care delivery. While many technologies are developed to aid nurses in identifying important changes in patient conditions (i.e., enhance the signal), the wide variety of simultaneously available user interfaces and poor interoperability impedes nurses' ability to cognitively process and appropriately use the data in caring for patients (i.e., too much noise).[17] In contrast to the study of physicians and their workflows incorporating decision support, little attention has been placed on the nursing population.

Because the potential impact of predictive analytics on nurses' clinical decision-making is a newer concept to consider, it is proposed that the initial step in investigating the usefulness of those analytics would include the study of a clinical event in which ideally a probability score could be presented close to the time when the actual event might occur. This approach facilitates real-time feedback and minimizes the influence of additional variables (e.g., other clinicians' actions or non-hospital factors) on the connection between the probability of events and the actual occurrence of events. Using available data to predict events likely to occur within the subsequent 24-48 hours would be ideal for the in-hospital nurse. The risk of cardiopulmonary arrest (CPA) meets these criteria and served as the context for this study. With over 200,000 people treated for in-hospital cardiac arrest every year[5] and survival rates ranging from 23 to 37%,[22] this is a clinically important adverse event that might be preventable.[24]

## 2.2 Objectives

This study began the work of incorporating predictive analytics into clinical workflows by exploring nurses' current activities involved in problem recognition, information gathering, and problem solving. Given the novelty of producing risk data for nurses, it was important to not only understand their use of probability-based information but also their perceptions and use of the terms *probability*, *risk*, and *uncertainty*. The aim of this study was to identify key concepts in those three areas to determine where probability-based tools might fit within the "orienting frames" nurses use for organization and conduct of their work.[52] Beginning with a qualitative approach was thought to provide the most suitable insight into how we can successfully deploy big data applications such as predictive analytics into the clinical environment.

## 2.3 Methods

### 2.3.1 Design

We used a qualitative description approach[53,54] for data collection and analysis to understand participants' information gathering behaviors and term perceptions. Naturalistic inquiry incorporating design research concepts yielded rich description of participants' experiences gathering information and interacting with probability-based data. The naturalistic inquiry paradigm influenced our attempts to understand participants' experiences while minimizing speculation, and describing the context and nature of nurses' work is consistent with design research. Design Research methods[47,55] can answer questions such as: (a) What non-patient data are available within the healthcare system? (b) How/when do nurses think and act on the topic of interest? (c) What do different nursing roles need? and (d) How do we incorporate our results into the workflow?

### 2.3.2 Setting and Participants

The study took place at a single academic medical center in an urban city in the mid-South region of the U.S. Participant recruitment involved e-mail and printed flyers along with face-to-face discussions with nurse leaders. Inclusion criteria comprised: (a) bedside nurses who cared for a patient requiring activation of a rapid response team or CPA resuscitation efforts within the last 6 months for a duration of at least 2 hours before the event, and (b) charge nurses who assigned nurses to patients at least twice per week over the last 6 months. Our purposive sample consisted of 18 nurses. One-on-one interviews were conducted with ten direct care bedside nurses and five charge nurses. Bedside (i.e., direct care) nurses were included because they are the clinicians who spend the most time with patients and thus were appropriate to evaluate predicted versus actual CPAs. Charge nurses, who are responsible for determining nurse-patient assignments, were included because decision-making with predictive analytics might also serve organizational leaders who manage resources. One focus group of three charge nurses (rather than individual interviews) was also conducted because we assumed task recall for the complex activity of assignment making would be easier by hearing cues from others performing similar work.

### 2.3.3 Data Collection Procedures

#### *2.3.3.1 Variables of Interest*

Data collection focused on three major areas of interest: (1) information gathering behaviors of participants related to deteriorating patients, (2) potential benefit of clinical prediction models for deteriorating patients, and (3) definitions of risk, uncertainty, and probability. These focus areas guided the development of a semi-structured interview guide, which was pilot-tested with practicing nurses and graduate nursing students before using.

Regarding information gathering behaviors, we used questions focused on the materials accessed (e.g., electronic health record), activities performed (e.g., obtaining vital signs, performing physical assessments), and people contacted (e.g., other clinicians) during participants' efforts to gather information about a deteriorating patient. To prevent speculative findings, participants need to have experienced the phenomenon of interest. If the study setting does not allow participants to be observed experiencing the phenomenon, their input in interviews must be based on an actual experience. Therefore, to determine how prediction models might be perceived within clinical environments, we used weather scenarios as a proxy. Given that participants had prior exposure to probability-based weather information (but not clinical prediction tools), weather scenarios served as a surrogate to understand how they might use and interpret probability-based decision support tools. Scenarios and images were designed to elicit responses from participants that were similar to those we expected nurses might use when inferring a patient's likelihood of experiencing clinical deterioration from a predictive tool. Scenarios included weather forecasting examples of daily summaries, hourly trends, images, numbers, and one scenario of a discrepancy in predictions between two different forecasting websites for the same geographic location (see Figure 2.1 and Figure 2.2). To obtain perceptions of the terms *probability*, *risk*, and *uncertainty*, we simply asked participants to provide us with a definition and example.

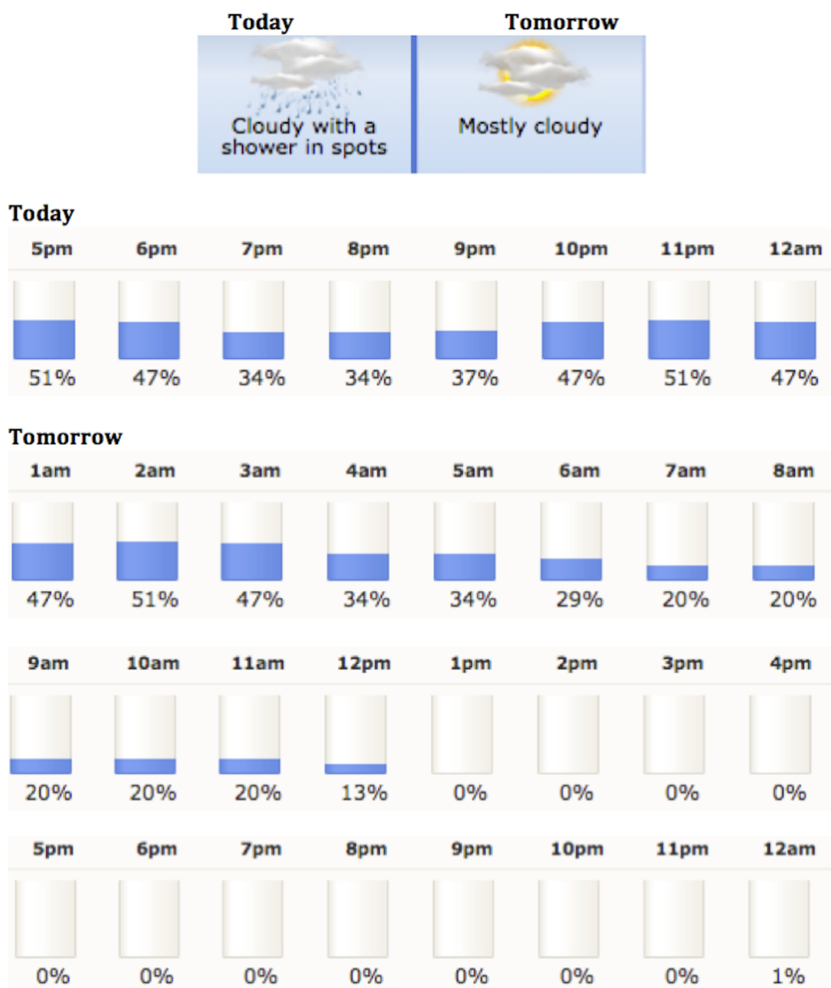



Figure 2.1: First example of weather scenario provided to participants. This scenario depicts the chance of precipitation for the next 30 hours from a general overview and hourly probabilities.[56]

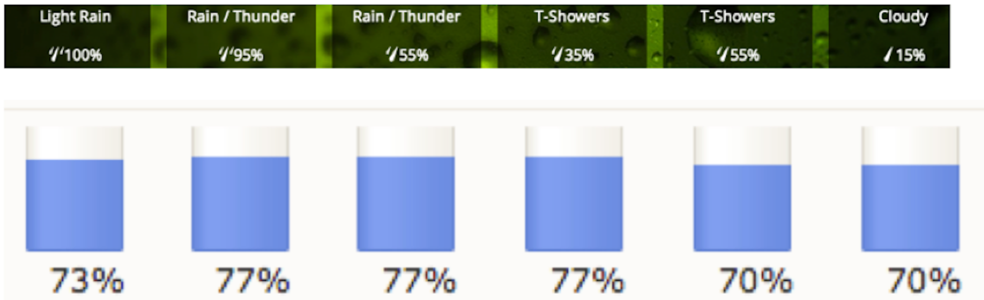



Figure 2.2: Second example of weather scenario provided to participants. This depicts a scenario where two different weather sources forecast discrepant precipitation probabilities for the same location.[57,56]

#### *2.3.3.2 Interview and Focus Group Methods*

The Institutional Review Board approved the study, and participants provided informed consent. We asked charge nurse participants to write responses to guided questions on a worksheet *before* attending the interviews as a memory aid to minimize recall bias[58] for the complex task of assignment-making. These questions were intended to elicit thoughts while making assignments on a clinical shift, and we discussed these written responses during the interviews and focus group.

Following researcher introductions, we used the pilot-tested, semi-structured interview guide for data collection with all study participants. Individual interviews lasted approximately 60 minutes, and the focus group lasted 120 minutes concluding when participants had no further information to add in response to questions and probes. All participant discussions were audio recorded and occurred in a private room at the medical center where participants were employed. Bedside nurses received a $30 gift card for participation, and charge nurses, because of the pre-interview worksheet activity requirement, received a $50 gift card for participation. Researchers collected field notes during the discussions with participants and during debriefing discussions with each other.

#### *2.3.3.3 Moderators*

The principal investigator (AJ), a male registered nurse and doctoral candidate with some qualitative research experience, was present for all interviews and the focus group. A co-investigator (LN), a female doctoral-prepared organizational anthropologist with expertise in qualitative research methodology, was present for three individual interviews and the focus group. LN's presence for individual interviews focused on evaluating AJ's moderating skills and immersing herself in a sample of the data while her presence in the focus group permitted the two-moderator approach. Consistent with a two-moderator approach to focus groups, the principal investigator (AJ) was able to be the subject matter expert while the co-investigator (LN) served as the focus group conduct specialist.[59] The moderators had no supervisory relationship with the participants.

### 2.3.4 Data Analysis

Thematic data analysis involved coding all transcribed discussions and written statements by two coders (AJ and LN). To develop the codebook, coders jointly applied codes to two interviews, discussing differences until consensus was reached on the set of codes and their definitions. The online qualitative analysis software Dedoose[60] was used to apply codes, group themes, look for patterns, and compare findings with existing literature. A Key Concepts analytic framework guided identification of factors most important to the study's objectives.[59] Methodologists have recommended that this surface-level interpretation results in findings "closer to the data as given" (p. 78).[61]

## 2.4 Results

Eighteen participants (see Table 2.1) from 15 interviews and 1 focus group produced 14.5 hours of recorded audio, 525 pages of transcripts, and many hand-written notes. Findings are presented in alignment with the three areas of interest: Information Gathering Behaviors, Potential Benefit of Clinical Prediction Models, and Perceptions of Probability, Risk, and Uncertainty.

Table 2.1: Descriptive Statistics of Participants (n=18).

| Demographic Variable | | n |
|---|---|---|
| Education∗ | Associate | 2 |
| | Bachelor | 10 |
| | Master | 3 |
| Age | Under 25 | 4 |
| | 26-35 | 8 |
| | 36-45 | 2 |
| | 46-55 | 3 |
| | 56-65 | 1 |
| Nursing Experience (Years) | <2 | 3 |
| | 5-Feb | 4 |
| | 10-Jun | 5 |
| | 15-Nov | 2 |
| | >15 | 4 |
| Patient Population | Adult | 9 |
| | Pediatric/Neonatal | 9 |
| Care Setting | Emergency Department | 4 |
| | Intensive Care Unit | 7 |
| | Ward/Floor | 7 |

∗ *Education variable not available for focus group participants*

### 2.4.1 Information Gathering Behaviors

Participants reported collecting information from many sources when evaluating whether a patient was at risk for CPA. We categorize these sources as: Patient (e.g., physical assessment, past medical history), Other People (e.g., family members, physicians), and Technology (e.g., electronic health record, vital sign monitor). The process by which they gathered information differed by role (i.e., bedside nurse or charge nurse).

#### *2.4.1.1 Bedside Nurses*

Bedside nurses working in the emergency department reported gathering initial information from the patient while nurses in the inpatient setting gathered initial information from the previous shift's nurse. Initial patient information quickly resulted in mental classification of the patient's clinical status such as "previously healthy" versus "multiple comorbidities", "high-risk" versus "low-risk" surgery, or "stable" versus "unstable" during previous shift. One participant noted:

> *I call it a drive-by assessment. It's when you like, if you were to come in the ER and if you walk in. . . , that gives me a few minutes to go in there and see, because you're probably stable because you're walking into the room.*

The initial mental classifications appeared to result in a baseline assessment against which future information was integrated. A different participant commented on this phenomenon by stating:

> *When you have a patient, either whether it's multiple days or there are multiple admissions. . . it's always beneficial to the nurse, I believe, because you kind of see them through their, at their baseline, you see what's going on with them from a day to day basis and you can evaluate kind of what's going on with that patient based on what you've seen previously.*

After compiling the initial information, bedside-nurse participants communicated with the patient (and/or family or caregivers), conducted physical assessments, obtained vital signs, and reviewed the patient's history, all in no particular or consistent order. When participants discovered discrepancies among information sources, for example a physical assessment finding inconsistent with the patient's history, they sought additional information or communicated those inconsistencies with others, such as the charge nurse, physician, or rapid response team. Bedside nurses provided insight on *what*, *where*, and *when* nurses were charting in the EHR. For example, most participants noted that if the change in patient status was capable of harming the patient, documentation of the nurses' findings did not occur in real-time but rather after the decisions and actions to safely manage the patients were finished. One participant noted:

> *. . . when he was probably decompensated I probably didn't chart as often because I was doing. . . procedures with the patient I guess. . . if there's a procedure at the bedside, we actually like paper chart.*

In addition to identifying which elements were not charted in real-time, participants were helpful in identifying additional variables for prediction model development versus those that are not beneficial for real-time decision support algorithms. For example:

> *I usually document any notifications, any provider notifications, and . . . who we spoke with, what the concern was, and what the result of the notification was. . . I'm sure early on I had documented some of my concerns [in real time], but then probably not in real time after that.*

#### *2.4.1.2 Charge Nurses*

When working with individual patients, charge nurses (who had previously worked as, or were currently working in the capacity of, a bedside nurse) expressed the same information gathering behaviors as bedside nurses. However, in their role of making the patient care assignments for the unit, charge nurses gathered information differently. Charge nurses reported needing less individual patient detail than bedside nurses because they had many patients' needs to consider, balanced with available nursing personnel. Themes surrounding the number of available clinicians (including nurses, nursing assistants, and physicians, both on their own unit and on other units) surfaced frequently. Charge nurses tended to gather information from the bedside nurse more frequently than any other source, even though charge nurses did report talking to physicians during patient changes and reviewing the patient's medical record upon admission. Charge nurses in the focus group summarized by stating:

> *Participant #1: I go a lot with the intuition of the bedside staff. Most of the time, the attending physician, I don't trust most of the residents, but our attending physicians, our fellows, and the experienced nurses, especially the ones that I've personally worked with for a long time and trust, I would trump a nurse's intuition over every other kind of objective data.*
>
> *Participant #2: I agree with that totally.*

Charge nurses also strongly considered individual staff members' experience, expertise, and preferences along with individual patients' needs and their relationship with nurses.

> *I knew that I needed, and this is our expression, I needed a strong nurse for this assignment because it could either end up in. . . , withdrawal, you know bereavement or escalating care. . . , even more so than what we already had.*

### 2.4.2 Potential Benefit of Clinical Prediction Models

Three major themes and four minor themes emerged in participant responses to the weather scenarios (see Table 2.2). Major themes included attempts to find information from additional sources during uncertainty (i.e., triangulation), always being prepared for the worst-case scenario regardless of predictions, and the desire to review more detailed projections (e.g., viewing hourly predicted values rather than a daily summary and knowing the source of the information). Regarding this latter theme, participants also noted that consistency (i.e., hour-to-hour stability) and extreme values (i.e., 0% or 100%) provide a sense of confidence or certainty in the outcome. In contrast, certainty appeared to diminish with temporally distal predictions such as those more than 24 hours from the current time. Minor themes that surfaced among some participants included: compromising when faced with discrepant values, a deference to personal preference to simplify decision-making, building a story to accompany the empirical data, and the ability of prior beliefs to supersede new evidence.

Table 2.2: Major and Minor Themes Resulting from Weather Scenarios.

| **Major Themes** | **Description** |
|---|---|
| #1: Triangulation | When faced with uncertainty, participants sought information from different sources (e.g., reviewing a weather map in addition to probabilities), facilitating discovery of a "true" state. |
| #2: Always be Prepared | Some participants never trusted the predictions because they have found them to be wrong too many times. Others seemed slightly risk averse. In both cases, they preferred to bring an umbrella or raincoat in the event of an undesirable outcome. |
| #3: Desire for Detail | Almost all participants preferred an hourly forecast view. This level of detail provided insight into which activities can be performed by the participant and provided a trend by which one can make more-informed predictions. |
| #3a: Consistency and Extreme Values | Consistency and extreme values (e.g., repeated hours of 0% probability) promoted comfort, certainty, and confidence. |
| #3b: The only certainty is uncertainty. | Predicting soon is, at best, uncertain — predicting the far future, one might as well not look, that is unknown. |
| **Minor Themes** | |
| #1 Compromising | When faced with the discrepant values, participants assumed the true value was somewhere in the middle. |
| #2 Deference to Preference | Similar to the cognitive heuristic of simplifying decision making, participants noted that when faced with discrepancies, they would defer either to the outcome they prefer (i.e., no rain) or to the outcome suggested by their most trusted source. |
| #3 Story Building | Giving a story or personality to the data (i.e., devising a rationale for why something might be displayed as it is). |
| #4 Prior beliefs can supersede new evidence. | Belief that context (e.g., geographic location) was so important that no new evidence would make someone come up with a different conclusion. |

Of note, there were a few divergent views among our participants. With respect to Major Theme #3 regarding a desire for detail, two participants noted that too much detail (i.e., hour-by-hour predictions) showed the model was "too confident" and therefore could not be trusted. One of those participants noted:

> *You've got to apply it across the whole day because you can't say that the wind is not gonna pick up at that moment and bring this 55% worth of showers into my next hour.*

This opinion was in contrast to that of another participant who felt some hourly variation *increased* her belief in the model. One of the focus group participants noted that abrupt changes made him disbelieve a prediction.

With the respect to Minor Theme #4 regarding prior beliefs, one participant expressed that context (e.g., geographic location) was so important that no new evidence would make her come up with a different conclusion.

> *Interviewer [I]: . . . What would something like this [see Figure 2.1] mean to you if you were planning something, this is today, this is tomorrow, if you were planning something tomorrow outside, would you feel comfortable?*
>
> *Participant [P]: Nashville, Florida or the tropics, give me a location.*
>
> *I: Why does location matter?*
>
> *P: Because if you're in the tropics it's going to rain whether you think it's going to or not. . . And if it's Florida or California it'll probably rain for 2 hours in the afternoon and you'll be good for the rest of the day. And if it's Jamaica and the Dominican it always says it's probably gonna rain and it never does 'cause it's gorgeous, so it really just depends on where we are. . .*
>
> *I: So there's a lot of context or prior knowledge that . . .*

*P: Yeah. Yeah. I need more information.*

Another participant made a similar comment that because she grew up in the area, her opinions on weather patterns were more trustworthy than the meteorologist's predictions. Similarly, one participant stated she would not look for more information from different sources and did not care about the reputation of a source - only which sources appear more accurate based on her past experience.

As it related to the clinical intent of these questions, some participants expressed reservations about the ability of a prediction model to accurately predict cardiopulmonary arrest. One participant noted:

> *An arrest is so multifactorial and I think that, I don't know. I would have a hard time seeing; could be if any, something that could pre-, predict it with the kind of accuracy beyond just this is a higher-risk patient because they've had a more complex surgery.*

Another participant stated:

> *How... are you collecting your data? How are you presenting it? Pretty and green doesn't make a difference if you're wrong, if you're guessing, but neither does basic and blue, so it depends. At some point in time I just need a yes or no. [laughs]*

### 2.4.3 Perceptions of Probability, Risk, and Uncertainty

Regarding perceptions of the words *probability*, *risk*, and *uncertainty*, participants were inconsistent in their use of these words and frequently used the words interchangeably. Those words did not appear to have a consistent, intrinsic meaning between or within participants (see Table 2.3). For example, one participant noted:

> *I mean, I guess you could group them all kind of under the same classification if you're like, there's a probability of risk and then you can have a probability of uncertainty. And I think risk and uncertainty are pretty similar because when you're uncertain there's always a risk.*

Another participant noted, *"...the risk of going to the unit is 100%, the probability of going to the intensive care is 100%, and the, I mean, you hear what I'm saying?"*

Table 2.3: Themes Related to Perceptions of Risk, Uncertainty, and Probability.

| **Theme** | **Description** |
|---|---|
| #1: Probability = "probably" | Participants used "probability" to imply that an event/outcome will "probably" occur (i.e., more like to occur than not to occur). |
| #2: Risk is related to a patient's history and might assist with prioritizing. | Participants were able to mention some types or classifications of patients that could automatically be labeled as "high risk," which implied greater attention should be focused on that patient. |
| #3 Risk is related to harm. | Risk was the term most related to the concept of patient harm. |
| #4 Uncertainty = "unexpected, unpredictable, unknown." | Participants noted that one can never fully expect or predict what will happen with a hospitalized patient. |

#### *2.4.3.1 Probability*

Several participants related the term "probability" to mathematics; however, participants' examples did not adhere to traditional rules of mathematics. The majority held a view summarized by a participant who stated:

> *I really don't think in terms of like probability that something bad is going to happen... Probability is a more statistical term that I tend to stay away from.*

#### *2.4.3.2 Risk*

Participants related *risk* to a patient's medical history and associated it strongly with harm. For example, one participant noted:

> *The risk of hypotension in a heart failure patient is high because... we give them very strong diuretics.*

Other participants indicated that risk stratification helped them prioritize the needs of their patients, both for the purpose of care planning and interpersonal communication with other clinicians.

#### *2.4.3.3 Uncertainty*

Regarding *uncertainty*, a few participants implied that clinicians can never fully expect or predict what will happen with a hospitalized patient. One participant described uncertainty as *"... an over-looming thing with everybody."* Another participant commented on the nurses' role in managing the tension between statistical data and the lived experience of patients and families, noting:

> *as far as statistics, there's some legitimacy to them, but they're never, they should never dictate your care that you provide or the information you give a family. You can talk statistics with them, but you know what, allow them to keep their hope and say, "Look statistically speaking these are the things that happen." You could have them mentally prepare themselves for those things, but also let them know, you know what? Things can get better too ... things can be better than what statistically speaking and you know, you can have your hope and you can continue to push.*

Finally, a few participants noted that uncertainty was present when one does not know why something is occurring (e.g., a deteriorating patient with no rationale for that state). One participant described uncertainty as "insecurity" or a lack of confidence in the sense that a novice nurse might not know if she should share concerns with a physician.

## 2.5 Discussion

We have described nurses' perceived workflows for gathering information related to clinical deterioration and nurses' beliefs related to probability-based information. An initial qualitative approach provided a rich foundation for understanding behavior change (i.e. work processes) in response to predictive analytics to support design and evaluation. Although the *content* of information gathered by nurses was similar, a consistent *temporal pattern* in their information gathering was lacking. The lack of a predictable temporal pattern for information gathering has important implications for the design of decision support tools, similar to the non-linear workflows described in bar code medication administration.[62] Given that nurses gather a significant amount of information outside of the electronic health record (EHR) and that EHR documentation of clinical deterioration tends to occur after problem recognition and management, a predictive tool for this outcome likely belongs outside the EHR. Furthermore, brief and informal patient assessments (referred to by the participants as a "drive-by") are not routinely considered important enough to document in the EHR, even though they contain information that influences a nurse's anticipated trajectory of patient status.

Our finding that information gathering behaviors appeared to differ between bedside nurses and charge nurses suggests recommendations concerning information provision and decision support may not be transferable to all nurses but rather might require unique designs based on roles. To our knowledge, no studies have explored clinical decision support tool specifications that differ by nursing role. Our findings suggest that charge nurses might be key players in effectively incorporating decision support systems predicting negative patient outcomes because of their expressed preference for high-level overviews of patient status. Probability-based decisions support tools provide numerical outputs that aggregate information from multiple sources. Therefore, if summarized probability information could be provided for a charge nurse's patients in one medium (e.g., ranked in order), such a display could provide assistance to charge nurses' decisions in patient care assignments.

Discrepancies, whether in patient assessments or weather scenarios, were important in initiating the process of gathering additional information. This finding aligns with decision theory principles that one option for responding to uncertainty is the gathering of additional information in an effort to reduce uncertainty.[63] Participants expressed concerns surrounding a prediction model displaying results contradicting their clinical gestalt, but our findings suggest those discrepancies might simply result in seeking additional information.

Few studies have evaluated the direct impact of predictive analytics on the decisions nurses make,[1] yet several researchers have explored nurses' cognitive work in the context of the recognition and response to a deteriorating patient. For example, J. S. Braaten[64] conducted a cognitive work analysis with 12 nurses and reported, similar to us, that nurses gather information from a variety of sources and that they preferred the collection of objective clinical criteria to justify rapid response team activation. Additionally as part of the systematic review of deterioration detection among ward patients conducted by M. Odell, et al.,[65] the roles of *intuition via pattern recognition*, *patients and families expressing concerns*, and *detecting abnormalities during routine care* were identified as information gathering behaviors. Finally, similar findings by J. Cioffi[66] suggested that baseline mental stratification of patients and prior beliefs superseded new evidence.

Clinical prediction models provide outputs in the form of statistical probabilities that require analytical decision-making skills for interpretation. C. G. Parker[67] found that nurses who use analytical decision-making skills call rapid response teams more frequently than nurses who primarily use intuitive decision-making. However, if nurses tend to use intuitive decision-making skills frequently,[68] the design and implementation of probability-based CDS tools should appreciate nurses' propensity for intuitive decision-making while facilitating the analytical counterpart when prompting action. Given that several of our participants did not prefer statistical probabilities for decision-making, providing predictive model information will be an important consideration in future work.

The weak agreement of definitions for *risk*, *uncertainty*, and *probability* likely resulted from scarce decision theory knowledge among participants. This topic is not routinely covered in nursing curricula. Furthermore, many possible actions exist for nurses during a changing patient condition in which risk or uncertainty are present. The unique arrangements of actions and goals chosen by individual nurses might result in unique definitions. A few participants' association of these terms with statistical probabilities and their statements regarding infrequent use is consistent with Kahneman's[69] view that humans do not think probabilistically.

Several participants mentioned the role of context as influential in the extent to which nurses would accept the validity of a prediction model. This can have important implications for CDS tools (see Table 2.4) because the use of predictive analytics without an awareness and appreciation for the context of its application could lead to both: (a) identification of correlation without an understanding of causation that could thwart future work and (b) unsuccessful implementation of clinical tools developed from analytical methods. We propose that engaging nurses in the design of analytics solutions is one of the best solutions for these challenges. Specifically, Design Research methods can be both a starting point and an implementation aid for big data applications. Strengths of our study included the variety of both nurses and settings, as well as the use of a proxy situation to elicit responses that were more likely to represent actual behaviors and decrease speculation. Limitations of our study included the single research site, small focus group size, and dependence on participant recall and perceptions. There is a need in future studies for documentation of what actually occurs rather than asking participants to recall what they did, similar to a recent report of critical care information gathering behaviors.[70] Additional nursing roles, such as advanced practice nurses, also need exploring, similar to the work of S. Weber.[71] We plan to examine nurses' use of prediction models in simulated and real-world settings in the near future, and findings from our study will guide the design and implementation of big data applications into the clinical arena. Prior to embarking on large-scale experimental studies however, further qualitative investigations of technology's influence on clinical decision-making and predictive modeling's impact on information technologies are needed.

Table 2.4: Implications of Findings for Probability-Based Decision Support Tools.

| **Finding** | **Implication for Decision Support Tools** |
|---|---|
| Desire for detail | Provide trends. Identify sources of data. |
| Discrepancies promote information-gathering | Even if the tool does not align with clinical gestalt, it might still be helpful in determining if a patient has a problem. |
| Deference to trusted source | If decision support tools provide discrepant findings, they are less likely to prompt behavior because participants noted they are more likely to go to their trusted source when faced with uncertainty. |

## 2.6 Conclusions

An increased understanding of nurses' perceived work practices and perceptions of risk will enhance the development of effective probability-based clinical decision support tools. Understanding nurses' cognitive work contributes to improved design and implementation of predictive tools in the clinical setting and informs important expansions of nursing educational curricula. While these recommendations may seem intuitive to clinicians, these processes are not as well known to informaticians creating decision support systems intended for the clinical arena. Documentation of clinicians' work practices will enable more opportunities for implementing decision support systems into their cognitive and physical workflows.

Chapter 3

# ACCURACY AND INTERPRETABILITY OF IN-HOSPITAL CLINICAL DETERIORATION PREDICTION MODELS

This chapter describes a quantitative study comparing the performance of several modeling strategies in the development of a clinical prediction model for IHCPA. The results will lead to future studies examining differences in nurses' abilities to interpret the outputs of the differing statistical strategies.

## 3.1 Introduction

Widespread implementation of rapid response teams and early warning scoring systems throughout hospitals has resulted in debatable improvements in clinical deterioration outcomes.[72,73] Even if one believes these early warning systems and rapid response teams improve patient outcomes, the incidence of in-hospital clinical deterioration remains high and continues to be associated with low survival rates.[74,22,5] Given that the prevention of adverse outcomes will depend on early recognition followed by appropriate management, tools to aid these processes are needed. Clinical prediction models, especially those incorporated into decision support tools that automatically retrieve data from electronic health records, are becoming increasingly popular and might be able to assist in the early identification of clinical deterioration.[72,35,36,37,38,39,75,76,77,40,78]

Optimal statistical approaches to embed within decision support tools and assist clinicians with recognition are still being identified. Most statistical approaches are simply classification models that attempt to identify how likely it is an event will occur. Researchers studying this phenomenon have focused on increasingly accurate models, but accuracy is not the only important feature of a statistical method's performance. For example, a model resulting in a single probability as opposed to probability trends over time might yield weaker models for implementation into the clinical environment. For nurses, especially those in a hospital, identifying when an event is likely to occur (or at least monitoring trends over time) might be equally important to the classification outcome of whether an event will occur at any point.

We compared two *traditional statistical modeling* strategies (logistic regression and Cox proportional hazards regression) and two related *machine learning* strategies (random forest and random survival forest) for in-hospital cardiopulmonary arrest (CPA). The focus was on both their statistical performance and clinical interpretability. We selected these four strategies based on their common occurrence in the scientific literature and because two of the strategies (logistic regression and random forest) predict a binary outcome while the other two strategies (Cox proportional hazards regression and random survival forest) predict a time-to-event outcome (see Table 3.1). The traditional statistical strategies leverage regression methods for classification and survival analyses while the machine learning strategies average the results of many decision trees that have been created by splitting a random selection of predictor variables in each tree.[79] Each of the four approaches was evaluated for model accuracy and discrimination, for expected number of alarms at select thresholds, and for differences in model outputs with respect to *what* was being predicted.

Table 3.1: Comparison of Analytical Approaches to Predicting In-Hospital Cardiopulmonary Arrest.

| | **Approach** | |
|---|---|---|
| **Purpose** | *Statistical* | *Machine Learning* |
| *Classification*<br><br>Predicts whether an event will occur | Logistic Regression | Random Forest |
| *Survival/Time-to-Event*<br><br>Predicts how likely an event is at each time point | Cox Proportional Hazards Regression | Random Survival Forest |

*Note:* Our chosen statistical approaches leverage regression methods. Our chosen machine learning approaches average the results of many decision trees that have been created by splitting a random selection of predictor variables in each tree.

## 3.2 Methods

### 3.2.1 Design & Setting

For this retrospective cohort study, we collected data from a de-identified copy of the electronic health records from adults (aged greater than 18 years old) at a large, urban academic medical center from 2006 to 2015. A start date of 2006 accounted for organizational policy changes related to rapid response team changes, which could have influenced the outcome of interest. The Vanderbilt University Institutional Review Board approved the study.

### 3.2.2 Variables

We defined the outcome of interest (dependent variable) using CPT code 92950 (i.e., Cardiopulmonary Resuscitation), and a review of the literature guided our selection of candidate predictor variables, which comprised demographics, vital signs, laboratory values, and ICD-9 codes upon hospital admission. To identify time-to-event outcomes (i.e., event day for cases and length of stay for controls) for the survival analysis approaches, CPT codes were used in this data source as the most accurate method for identifying an exact date of care provided. To identify a hospitalization day, we required a patient to meet one of the following criteria: (a) one of approximately 50 hospitalization CPT codes, (b) a Braden assessment, or (c) a complete blood count or basic metabolic profile specimen collection CPT code. The length of a hospitalization course was then constructed by combining all sequential dates in which one of the aforementioned criteria was met. For patients with an emergency department visit CPT code on the day before a hospitalization course, the emergency department visit date served as the first hospitalization day.

### 3.2.3 Sample

Patients were excluded if they received cardiopulmonary resuscitation on the same day as an emergency department visit or on the first day of their hospitalization (see Figure 3.1). Among eligible patients with multiple cardiopulmonary resuscitation events, we retained the encounter with the earliest event. Control patients who did not experience a cardiopulmonary arrest were selected from all hospitalized patients who never had a documented CPT Code 92950. For control patients with multiple hospitalizations during the study period, we retained the encounter with the least amount of missing data.

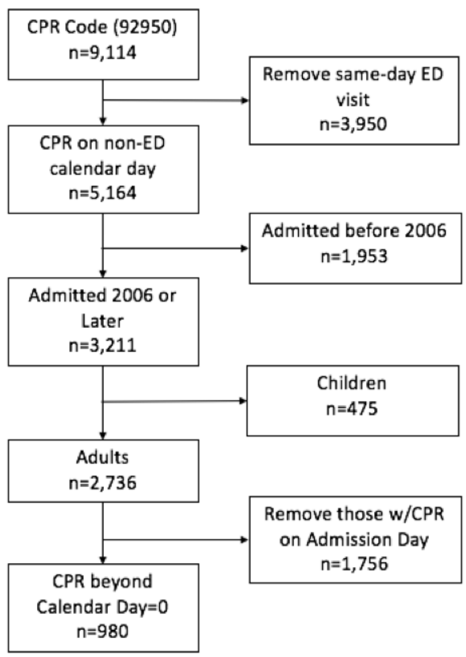



Figure 3.1: Selection of patients who experienced cardiopulmonary resuscitation (CPR).

### 3.2.4 Data Analysis: Pre-Processing

We began pre-processing by exploring extreme values, patterns of missingness, and collinear associations. Physiologically-implausible values (e.g., serum sodium $< 100$ mEq/L, pulse rate $> 240$ beats per minute) were re-coded as missing. We began with 60 candidate predictor variables and then removed 10 variables because they: (a) were missing in more than 80%

of patients [e.g., blood gas values], (b) were highly collinear with another variable [based on Spearman's rho $> 0.4$ or if values could be predicted by other values in a regression model with $> 90\%$ of the variance explained], or (c) had indeterminate time stamps where the first value could not be distinguished from latter values [e.g., blood pressure]. A full list of candidate predictor variables and rationale for exclusion is provided in Appendix A (see Table 3.4). Characteristics of patients for all the final predictor variables are found in Table 3.2.

Table 3.2: Descriptive statistics comparing patients who did and did not receive cardiopulmonary arrest. $_a$ $b$ $_c$ represent the lower quartile $a$, the median $b$, and the upper quartile $c$ for continuous variables. $x \pm s$ represents $\bar{X} \pm 1$ SD. Numbers after proportions are frequencies. Tests used: [1]Wilcoxon test; [2]Pearson test. Proc = procedural code. Dx = diagnostic code.

| | Controls $N = 168177$ | Cases $N = 980$ | |
|---|---|---|---|
| Age | 39 55 69 (54 ±19) | 48 61 71 (59 ±17) | <1e-04[1] |
| Gender | | | <1e-04[2] |
| female | 50% (84148) | 40% ( 393) | |
| male | 50% (83626) | 60% ( 586) | |
| Respirations | 16.0 18.0 20.0 (18.6 ± 4.2) | 16.0 18.0 22.0 (20.2 ± 6.2) | <1e-04[1] |
| Pulse | 77 90 104 ( 92 ± 21) | 80 96 113 ( 97 ± 22) | <1e-04[1] |
| BMI | 23.7 27.7 32.9 (29.2 ± 8.6) | 23.4 27.4 33.4 (29.6 ± 9.1) | 0.92[1] |
| Calcium | 8.40 8.90 9.30 (8.83 ±0.81) | 8.10 8.70 9.20 (8.62 ±0.96) | <1e-04[1] |
| Anion Gap | 7.0 9.0 11.0 ( 9.1 ± 3.5) | 8.0 10.0 13.0 (10.5 ± 4.4) | <1e-04[1] |
| Glucose | 95 111 139 (129 ± 69) | 102 123 166 (148 ± 87) | <1e-04[1] |
| Creatinine | 0.75 0.92 1.20 (1.25 ±1.36) | 0.89 1.23 1.94 (1.96 ±2.15) | <1e-04[1] |
| Serum CO2 | 22.0 25.0 27.0 (24.4 ± 3.9) | 21.0 24.0 27.0 (23.8 ± 5.2) | <1e-04[1] |
| Sodium | 136.0 138.0 140.0 (137.8 ± 4.0) | 134.0 137.0 140.0 (136.9 ± 5.6) | <1e-04[1] |
| Potassium | 3.60 3.90 4.30 (3.96 ±0.62) | 3.60 4.10 4.60 (4.15 ±0.81) | <1e-04[1] |
| Platelets | 172 223 283 (235 ±107) | 149 211 275 (226 ±131) | <1e-04[1] |
| WBC | 6.9 9.3 12.9 (10.7 ± 8.4) | 7.6 11.0 16.2 (13.7 ±15.8) | <1e-04[1] |
| RDW | 13.1 13.8 15.1 (14.4 ± 2.1) | 13.9 15.2 16.9 (15.7 ± 2.5) | <1e-04[1] |
| Hemoglobin | 10.6 12.4 14.0 (12.3 ± 2.4) | 9.7 11.4 13.3 (11.5 ± 2.6) | <1e-04[1] |
| Proc: Urinary System | 0.00 0.00 0.00 (0.01 ±0.12) | 0.00 0.00 0.00 (0.02 ±0.23) | 0.013[1] |
| Proc: Integumentary System | 0.00 0.00 0.00 (0.03 ±0.18) | 0.00 0.00 0.00 (0.09 ±0.67) | 0.078[1] |
| Proc: Respiratory System | 0.00 0.00 0.00 (0.02 ±0.18) | 0.00 0.00 0.00 (0.12 ±0.52) | <1e-04[1] |
| Proc: Nose, Mouth, and Pharynx | 0.00 0.00 0.00 (0.04 ±0.23) | 0.00 0.00 0.00 (0.10 ±0.55) | <1e-04[1] |
| Proc: Nervous System | 0.00 0.00 0.00 (0.06 ±0.35) | 0.00 0.00 0.00 (0.18 ±0.73) | <1e-04[1] |
| Proc: Musculoskeletal System | 0.00 0.00 0.00 (0.23 ±0.55) | 0.00 0.00 1.00 (0.37 ±0.75) | <1e-04[1] |
| Proc: Male Genital System | 0.00 0.00 0.00 (0.00 ±0.04) | 0.00 0.00 0.00 (0.00 ±0.00) | 0.36[1] |
| Proc: Hemic and Lymphatic System | 0.00 0.00 0.00 (0.00 ±0.07) | 0.00 0.00 0.00 (0.03 ±0.25) | <1e-04[1] |
| Proc: Female Genital System | 0.00 0.00 0.00 (0.01 ±0.09) | 0.00 0.00 0.00 (0.00 ±0.10) | 0.41[1] |
| Proc: Eye | 0.00 0.00 0.00 (0.00 ±0.09) | 0.00 0.00 0.00 (0.01 ±0.12) | 0.68[1] |
| Proc: Endocrine System | 0.00 0.00 0.00 (0.00 ±0.05) | 0.00 0.00 0.00 (0.00 ±0.07) | 0.86[1] |
| Proc: Ear | 0.00 0.00 0.00 (0.00 ±0.04) | 0.00 0.00 0.00 (0.00 ±0.00) | 0.31[1] |
| Proc: Digestive System | 0.00 0.00 0.00 (0.04 ±0.30) | 0.00 0.00 0.00 (0.12 ±0.75) | 0.001[1] |
| Proc: Diagnostic and Therapeutic | 0.00 0.00 0.00 (0.21 ±0.63) | 0.00 0.00 2.00 (1.04 ±1.84) | <1e-04[1] |
| Proc: Cardiovascular System | 0.00 0.00 0.00 (0.12 ±0.51) | 0.00 0.00 0.00 (0.61 ±1.42) | <1e-04[1] |
| Dx: Blood and Blood-Forming Organs | 0.00 0.00 0.00 (0.29 ±0.81) | 0.00 0.00 2.00 (1.12 ±1.68) | <1e-04[1] |

Table 3.2: *(continued)*

| | Controls $N = 168177$ | Cases $N = 980$ | |
|---|---|---|---|
| Dx: Circulatory System | 0.0 1.0 2.0 (1.8 ±3.1) | 0.0 4.0 9.0 (5.7 ±6.2) | <1e-04[1] |
| Dx: Congenitall Anomalies | 0.00 0.00 0.00 (0.04 ±0.34) | 0.00 0.00 0.00 (0.11 ±0.68) | <1e-04[1] |
| Dx: Digestive System | 0.00 0.00 0.00 (0.49 ±1.36) | 0.00 0.00 2.00 (1.12 ±2.14) | <1e-04[1] |
| Dx: Endocrine, Nutritional, Metabolic, Immunity | 0.00 0.00 1.00 (0.83 ±1.54) | 0.00 2.00 5.00 (2.69 ±2.83) | <1e-04[1] |
| Dx: Genitourinary System | 0.00 0.00 0.00 (0.37 ±1.04) | 0.00 1.00 3.00 (1.55 ±2.01) | <1e-04[1] |
| Dx: Infectious and Parasitic Diseases | 0.00 0.00 0.00 (0.16 ±0.59) | 0.00 0.00 1.00 (0.84 ±1.46) | <1e-04[1] |
| Dx: Injury and Poisoning | 0.0 0.0 1.0 (1.5 ±4.4) | 0.0 1.0 3.0 (3.7 ±8.2) | <1e-04[1] |
| Dx: Mental Disorders | 0.00 0.00 0.00 (0.53 ±1.46) | 0.00 0.00 1.00 (0.53 ±1.15) | 0.041[1] |
| Dx: Musculoskeletal System and Connective Tissue | 0.00 0.00 0.00 (0.35 ±0.98) | 0.00 0.00 0.00 (0.54 ±1.43) | 0.002[1] |
| Dx: Neoplasms | 0.00 0.00 0.00 (0.29 ±1.41) | 0.00 0.00 0.00 (0.68 ±1.94) | <1e-04[1] |
| Dx: Nervous System and Sense Organs | 0.00 0.00 0.00 (0.36 ±0.95) | 0.00 0.00 1.00 (0.78 ±1.62) | <1e-04[1] |
| Dx: Nonspecific Abnormal Findings | 0.00 0.00 0.00 (0.15 ±0.47) | 0.00 0.00 0.00 (0.27 ±0.65) | <1e-04[1] |
| Dx: Pregnancy, Childbirth, and the Puerperium | 0.00 0.00 0.00 (0.22 ±1.42) | 0.00 0.00 0.00 (0.03 ±0.54) | <1e-04[1] |
| Dx: Respiratory System | 0.00 0.00 1.00 (0.61 ±1.32) | 0.00 2.00 6.00 (3.54 ±3.57) | <1e-04[1] |
| Dx: Skin and Subcutaneous Tissue | 0.00 0.00 0.00 (0.12 ±0.61) | 0.00 0.00 0.00 (0.29 ±1.02) | <1e-04[1] |
| Dx: Symptoms | 0.0 1.0 2.0 (1.5 ±2.1) | 1.0 3.0 5.0 (3.3 ±3.2) | <1e-04[1] |
| Dx: Ill-Defined-Unknown Causes Morbidity-Mortality | 0.00 0.00 0.00 (0.05 ±0.25) | 0.00 0.00 0.00 (0.19 ±0.55) | <1e-04[1] |
| Dx: Supplemental V-Codes | 0.0 1.0 2.0 (1.4 ±2.0) | 1.0 2.0 5.0 (3.3 ±3.2) | <1e-04[1] |

Due to the high amount of unexplainable missing data (approximately 40% for lab values and 60% for vital signs) and lack of definitive guidelines on how to handle that magnitude of missing data,[12] we separately performed a statistical simulation study using 10,000,000 patients. We replicated distributions and associations from the empirical data to create a population, imposed several causes of missing data (i.e., completely at random, at random, and not at random), and then tested three imputation approaches to identify which method was most accurate under the missing data assumptions. Imputation approaches included missing-assumed-normal (similar to a median imputation), multiple imputation *without* the outcome, and multiple imputation *with* the outcome. The best approach to handle the high amount of missing data under the majority of assumptions was multiple imputation with the outcome using chained equations with predicted mean matching; therefore, we used that approach for our study.

#### 3.2.5 Data Analysis: Model Development

The first available measure of each of the predictor variables on the first day of hospitalization was included in each of the four methods: logistic regression, Cox regression, random forest, and random survival forest. Logistic regression is frequently used in biomedical studies as a multivariate regression model with a binary outcome. Cox proportional hazards regression is another commonly used multivariate regression model, but the outcome is time-to-event and allows for censoring.[79] Machine learning approaches for both binary classification and time-to-event include random forests and random survival forests, respectively.[79,80] Both random forest approaches build classification (and time-to-event, in our case) trees each comprising a random sample of predictor variables. Trees are split into branches based on cut points that optimize differences between the two new branches. After multiple trees are built, the predictions are averaged to develop a forest.

Logistic and Cox regression models were fit flexibly using restricted cubic splines and no interaction effects between variables. Consistent with multiple imputation, we fit these models to multiple imputed datasets and pooled coefficients and performance metrics across model fits. We performed post-hoc analyses of residuals and influential observations and found that reducing the number of knots in the restricted cubic splines from 5 to 3 helped the models meet assumptions. We assessed calibration and performed internal validation using the bootstrap of the last imputed dataset from the multiple imputation process. Random forest and random survival forest methods were trained using 50% of the data from last imputed dataset from the multiple imputation process. We reserved another 25% of the data for testing and the final 25% of the data for validation.

Statistical model performance comparisons were conducted via area-under-the-curve scores and F1 scores, along with receiver operating characteristic curves and precision-recall curves, respectively. Clinical impact and interpretability comparisons were conducted via positive prediction rate (number of patients triggering an alarm), recall (true positive rate), and graphical representations of model predictions from pooled and individual patients. The validation data set held out during the machine learning approaches (25% of the original data) served as the data for direct comparison of the models' expected future performance. Rather than using imputed values from the multiple imputation process, we performed median imputation for missing values to create a dataset with greater similarity to the clinical environment where multiple imputation is not easily feasible. Appendix A contains a visual representation of data used for imputation, development, and validation (see Figure 3.6). All analyses were performed using R, version 3.3.1.[81] The specific R packages used along with the mathematical formulas used to compare the models are available in Appendix A.

## 3.3 Findings

### 3.3.1 Statistical Performance

From a statistical perspective, all models performed similarly based on area under the receiver operating characteristic curve (AUC) but differed with respect to harmonic mean of recall and precision (F1 score). AUC values of the 4 models ranged from 0.847 to 0.861 suggesting good, consistent performance, yet F1 scores ranged from 0.170 to 0.325 suggesting poor and variable performance (see Table 3.3 and Figure 3.2). The order of most important variables changed with each model, but ICD codes associated with the Respiratory, Circulatory, Genitourinary, Endocrine, and Symptom-based diagnoses along with Diagnostic and Therapeutic procedures were within top ten most important variables in all four of the models. Variable importance rankings were similar between logistic regression and the random forest; however, between the survival approaches, the ICD codes were more influential in the Cox regression model while the clinical variables were more influential in the random survival forest. Additional details of the variable importance differences between models can be found in Appendix A (see Figure 3.7).

Table 3.3: Performance of Statistical Modeling Approaches.

| **Strategy** | **AUC** | **F1 Score** |
|---|---|---|
| Logistic Regression | 0.851 | 0.273 |
| Cox Proportional Hazards | 0.854 | 0.284 |
| Random Forest | 0.861 | 0.325 |
| Random Survival Forest | 0.847 | 0.170 |

*Note:* AUC = area under the (receiver operating characteristic) curve; F1 score = harmonic mean of recall and precision

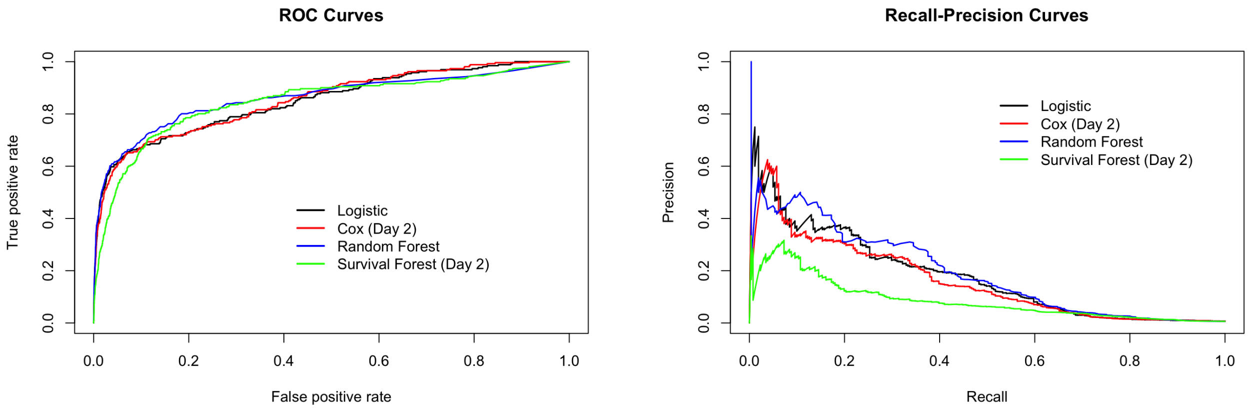



Figure 3.2: ROC curves (left) and Recall-Precision curves (right) for logistic regression, Cox proportional hazards regression, random forest, and random survival forest approaches. *Note:* Evaluation of survival approaches is provided at Time = Day 2 because that was the median time point.

### 3.3.2 Clinical Impact Performance

From a clinical impact perspective, the random forest and random survival forest identified more patients than logistic and Cox regression models at the same thresholds for CPA event probabilities ranging from 0.006 (actual event rate) to twenty times the event rate at 0.12 (see Figure 3.3). Similarly, the random forest and random survival forest had higher sensitivity rates at these same thresholds. With respect to the display of predictions that can be provided to clinicians, logistic regression and random forest models can provide a point estimate probability while the Cox regression and random survival forest models can provide probabilities that change at future time points. Figure 3.4 illustrates the estimated probability of a CPA event produced by all four models for two different patients – the average patient obtained by median values for all variables and an "ill" patient with several abnormal values. The random survival forest curve for the "ill" patient illustrates the most drastic change in predicted probability.

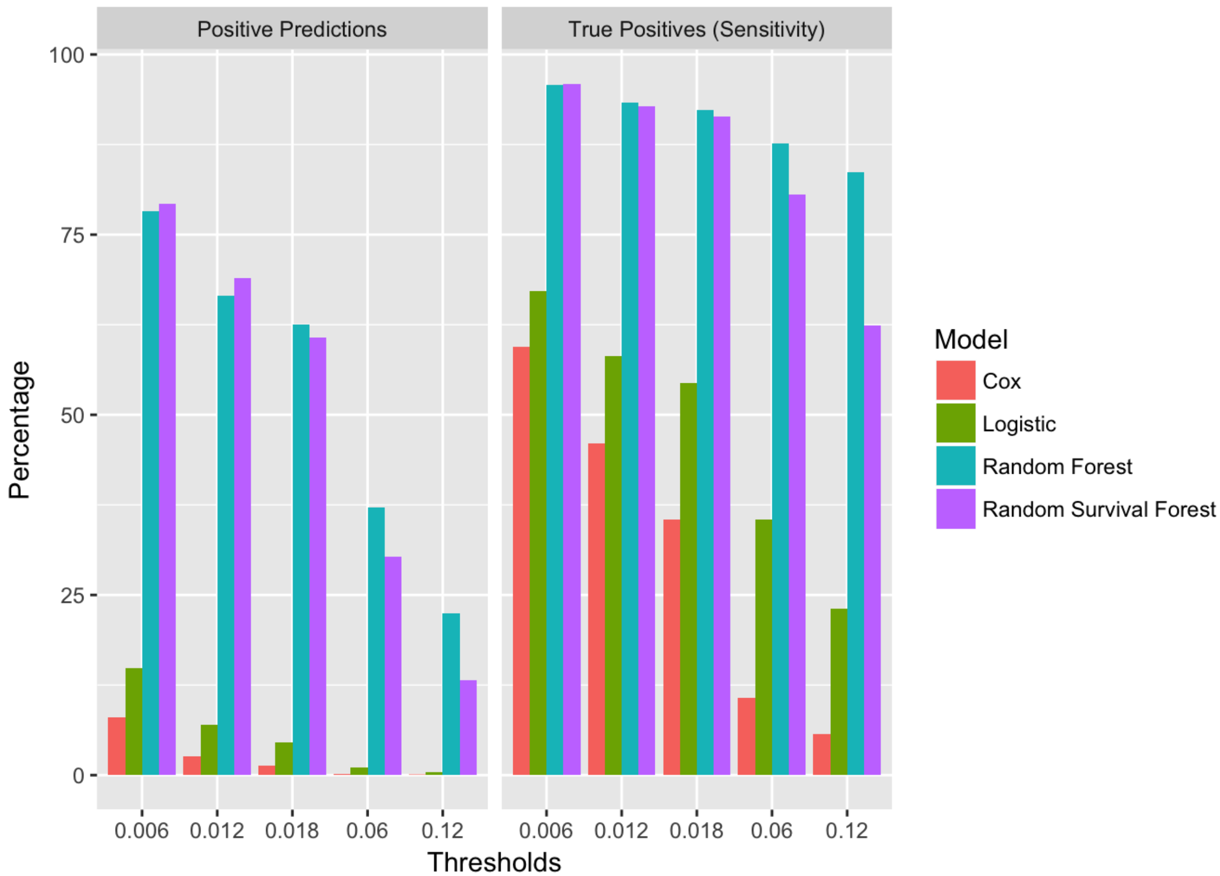
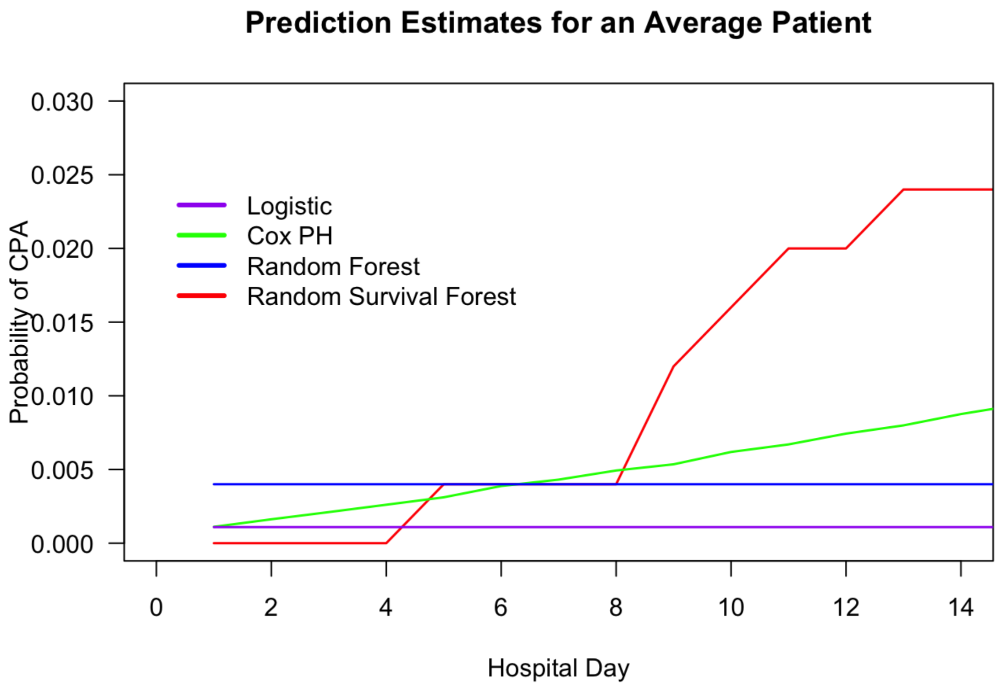



Figure 3.3: Comparison of positive prediction rate and sensitivity among all models at thresholds comprising the event rate in this dataset (0.006) and several of its multiples.



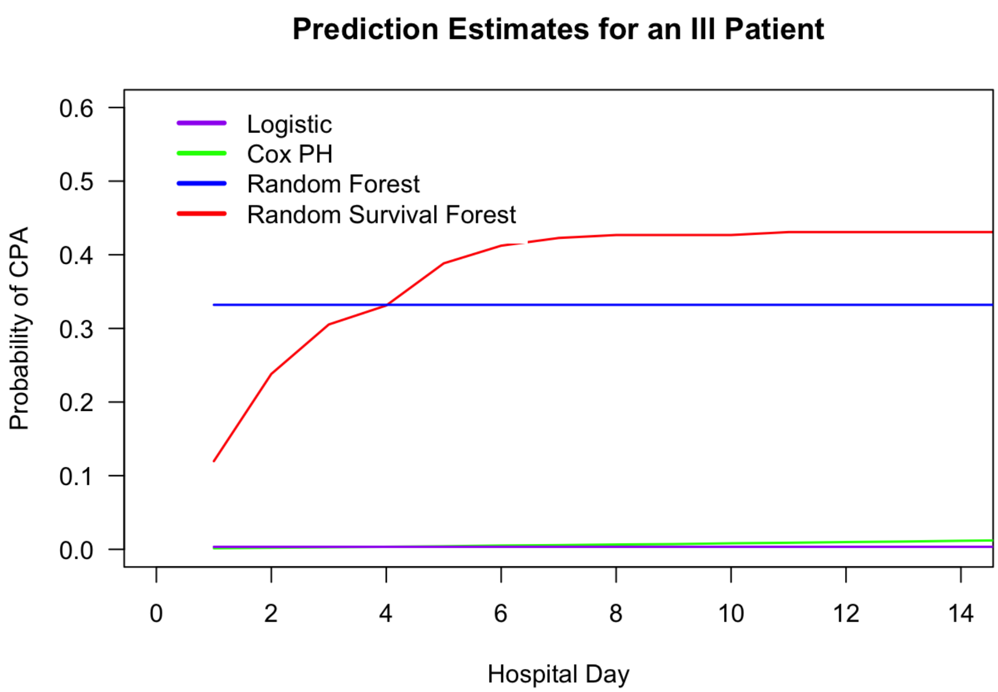




Figure 3.4: Comparison of estimated CPA event probability from 2 fictitious patients. Top: average patient defined as all model variables' values set at the median value. Bottom: ill patient characterized by several abnormal values (i.e., creatinine = 2, glucose = 300, potassium = 5, sodium = 150, hemoglobin = 7, red cell distribution width = 20, respiratory rate = 24, pulse = 115, and age = 80). *Note:* The average patient (top figure) has a different y-axis scale than the ill patient (bottom figure).

Because the random survival forest predictions showed the largest variability across time points for the ill patient, we explored whether the random survival forest demonstrated a similar degree of variability in predicted probabilities among all patients in our available data. We averaged the random survival forest prediction curves for all individuals in the data and compared these against the average Cox regression model predictions for the same individuals. Figure 3.5 shows the day-to-day changes in predicted probabilities for the random survival forest curves as much larger between CPA-positive versus CPA-negative patients when compared to the same two groups with the Cox regression model.

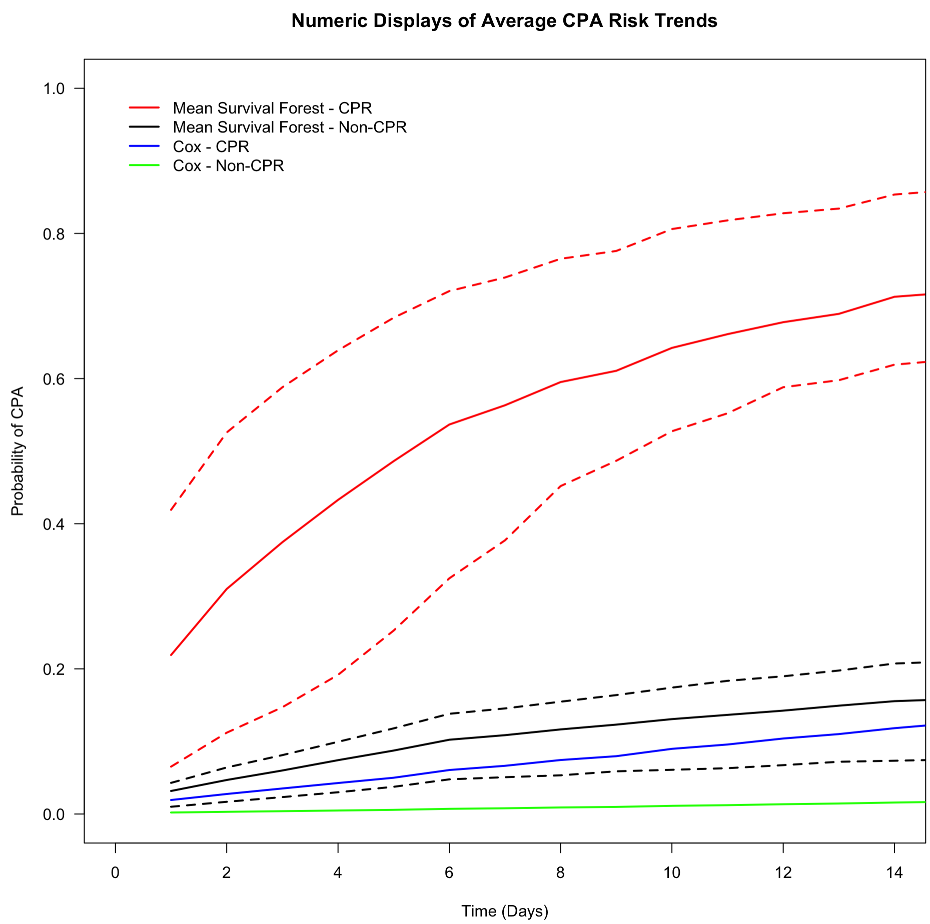



Figure 3.5: Summary curves for all predicted patients, stratified by those with CPA versus those without CPA. (- - - Dashed lines indicate 1st and 3rd quartiles of the random survival forest)

## 3.4 Discussion

Using a large dataset, we directly compared regression modeling and machine learning techniques for predicting in-hospital CPA. The approaches produced similar AUC values ranging 0.85-0.86, which are comparable to the findings of other researchers. A recent systematic review of early warning system scores for in-hospital clinical deterioration found most AUCs in the range of 0.74-0.86 for CPA.[73] These moderately large AUCs should not be surprising given the low event rate of CPA. Others' efforts to directly compare classification modeling strategies for CPA (i.e., logistic regression versus machine learning methods) have recently been published,[11] and their findings differed slightly from our findings in that the random forest approach outperformed logistic regression with respect to AUC (0.801 versus 0.770). Their study also found respiratory rate, heart rate, and age were the 3 most important predictor variables while we found several laboratory values to be the most important clinical variables in our models. Of note, they used a composite outcome of non-intensive-care unit CPA, unexpected intensive care unit transfer, and death rather than a single endpoint of CPA.

Conversely, the statistical performance of all modeling approaches was more dissimilar for recall and precision with F1 scores of 0.17-0.33. The two regression models (i.e., Cox proportional hazards for *time-to-event* outcomes and logistic for *classification* outcomes) performed similarly with F1 scores of 0.28 and 0.27, respectively. In contrast, the *time-to-event* machine learning approach (i.e., random survival forest) performed worse than the *classification* machine learning approach (i.e., random forest) with F1 scores of 0.17 and 0.33, respectively. Unfortunately, we were not able to compare our F1 scores with others' work because these metrics are not frequently reported in CPA prediction literature. With rare events, comparing precision (i.e., positive predictive value) is preferable to specificity due to precision's insensitivity to event rate.

The potential clinical influence of the models with respect to number of alarms varied as well. At all thresholds, machine learning approaches produced more clinical alarms than regression approaches (see Figure 3.3). This finding was accompanied by the benefit of increased sensitivity but could contribute to clinicians' alert fatigue if too many alarms are generated. Increased thresholds decrease the positive prediction rate and recall (sensitivity) while increasing precision (positive predictive value). In our study, increases in precision were noted at increasingly higher thresholds but eventually returned to zero in 3 of the 4 approaches (see Figure 3.2). The random forest model did not exhibit the same behavior, and in fact, precision reaches 1 at the most extreme threshold before returning to values comparable with other approaches. For clinical environments where precision is valued more than recall (i.e., where certainty in a positive prediction is more important than a false negative), the random forest approach could be more appropriate.

In terms of clinical interpretability, the prediction trends of time-to-event models might be more likely to influence clinicians' decisions. Time-to-event models produce trajectory curves that align more closely with the underlying deterioration phenomenon than a single probability that is expressed as a straight line on a graph (see Figure 3.4). The display of graphical probability trends offers a potential solution to alarm fatigue that might result from simple numerical cutoffs. While there does not appear to be a single superior approach at this time, given that the random forest *machine learning* methods have several advantages (i.e., fewer assumptions and increased variability in prediction trends) over the *traditional statistical* regression models and the time-to-event models allow prediction trends, the random survival forest might provide the best option for further model development work for in-hospital CPA. Future research looking at what is most likely to influence clinicians' decisions would be helpful.

### 3.4.1 Strengths & Limitations

We leveraged robust prediction model methods, including flexible regression models and newer machine learning methods. Random forests have the benefit of fewer predictor variable assumptions than traditional modeling strategies (e.g., linearity, interaction effects) and minimal overfitting compared to simple classification-and-regression trees. Use of the survival models is accompanied by the benefit of knowing *when* an event is going to occur. In contrast to non-critical events, such as 30-day readmission rates or pressure ulcers, greater precision of probability estimates for CPAs is of more value to in-hospital nurses.

We used a single outcome of CPA, but there is some evidence that using a composite measure (e.g., CPA, intensive care unit transfer, and mortality) increases statistical power.[82] We only included data available upon admission even though we expect that adding more values as they become available will increase the predictive accuracy of the model. Several additional approaches exist for repeated-measures data (e.g., mixed-effects regression, time-vary covariate survival models, and discrete-time survival models). These repeated-measures methods should continue to be explored despite our finding that a single-time model performed similar to multi-time models with respect to AUC.

The amount of missing data limits the trustworthiness and clinical applicability of the models. We found no evidence that patient characteristics influenced missing data patterns, and thus, we assumed data were missing completely at random. The cause of missing data likely resulted from data manipulation during transfer from electronic health records to the research database. The use of multiple imputation with chained equations and predicted mean matching, especially in such a large sample, produced results that were very similar to population/true values in our statistical simulations. Although having more non-missing values would have been preferred, missing data within clinical records is a commonly-encountered problem, and we cannot simply ignore data that are present by discarding variables with excessive missingness.

### 3.4.2 Future Directions

Future work should focus on obtaining datasets with less missing data and including additional variables that might predict CPA (e.g., mental status scales[83]). The field of predictive analytics for in-hospital CPA continues to expand, as noted by people publishing prospective protocols[84] and testing additional statistical methods.[85,86,87] We excluded patients who experienced a CPA on their first day of care because we anticipate different statistical strategies and model variables would be necessary to represent the phenomenon occurring earlier in a patient's hospitalization (e.g., using only emergency room triage data). Although we reported the heuristic advantage of noting trend line displays, we should compare whether trends or point estimates are more likely to influence nurses' behavior (within the larger context of usability studies).[88] We provided information on variable importance; however, these findings could be due to the amount of missing data, and the importance ordering should be re-visited in future studies.

## 3.5 Conclusion

As we embark on the continued development of probability-based clinical decision support tools for recognizing clinical deterioration, we must use the most appropriate statistical methods to model the underlying phenomenon. Improvements in accuracy are only one aspect of building decision support tools that are beneficial to clinicians; potential clinical impact (e.g., prediction format or number of alarms) is also an important consideration as we consider usefulness for bedside nurses. If we expect clinicians to incorporate these tools into their clinical workflows, we must be cognizant of both of these issues. Finally, given the potential impact of decision support interventions on workflow, nurses' roles, and patient outcomes, we advocate for increased collaboration between nurse scientists and biomedical informatics researchers to develop decision support tools that influence nursing work.

## 3.6 Appendix A

This supplementary information is intended to provide additional detail on the data pre-processing and data analysis as a way of increasing transparency and facilitating reproducibility.

Table 3.4: List of Candidate Predictor Variables (n=60) Initially Considered and Rationale for Exclusion of Variables (n=10) Not in Final Models. *Blood Gas Panel comprised pH, $pCO_2$, base excess, $pO_2$, lactic acid, and methemoglobin. *Note:* Temperature and pulse oximetry (variables frequently collected for hospitalized patients) were not available in the data set used for this study. All laboratory values were obtained from serum collections. Raw ICD-9 codes were collapsed into 19 diagnostic categories and 16 procedural categories.

| **Variable** | **Included in Final Models?** | **Reason for Exclusion** |
|---|---|---|
| Age | Yes | |
| Gender | Yes | |
| Race | No | Small sample in some categories resulted in a singular matrix during model fits |
| Ethnicity | No | Small sample in some categories resulted in a singular matrix during model fits |
| Body Mass Index | Yes | |
| Heart Rate | Yes | |
| Respiratory Rate | Yes | |
| Blood Pressure | No | Data source listed all timestamps at 00:00, so we were unable to determine first value |
| Sodium | Yes | |
| Potassium | Yes | |
| Chloride | No | Could be predicted by other variables in a regression model with $R^2$ >0.9 |
| Glucose | Yes | |
| Blood Urea Nitrogen | No | Collinear with Creatinine (Spearman's rho $\sim$ 0.4) |
| Creatinine | Yes | |
| Anion Gap | Yes | |
| Calcium | Yes | |
| Carbon Dioxide | Yes | |
| White Blood Cell Count | Yes | |
| Red Blood Cell Count | No | Collinear with Hemoglobin (Spearman's rho $\sim$ 0.8) |
| Hemoglobin | Yes | |
| Platelet Count | Yes | |
| Red Cell Distribution Width | Yes | |
| Blood Gas Panel* | No | Missing in >80% of patients |
| Braden Score | No | Missing in >80% of patients |
| ICD-9 Codes | Most | The Obstetrical procedure category was removed because it resulted in a singular matrix during model fits |
| CPT Codes | No | Only used for outcome variables |

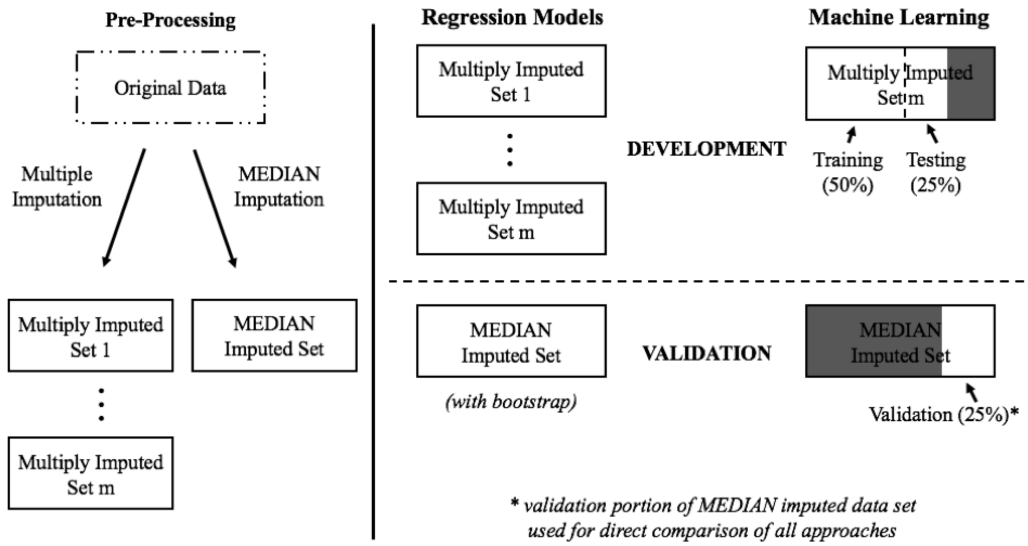



Figure 3.6: Datasets used for model training development and validation.

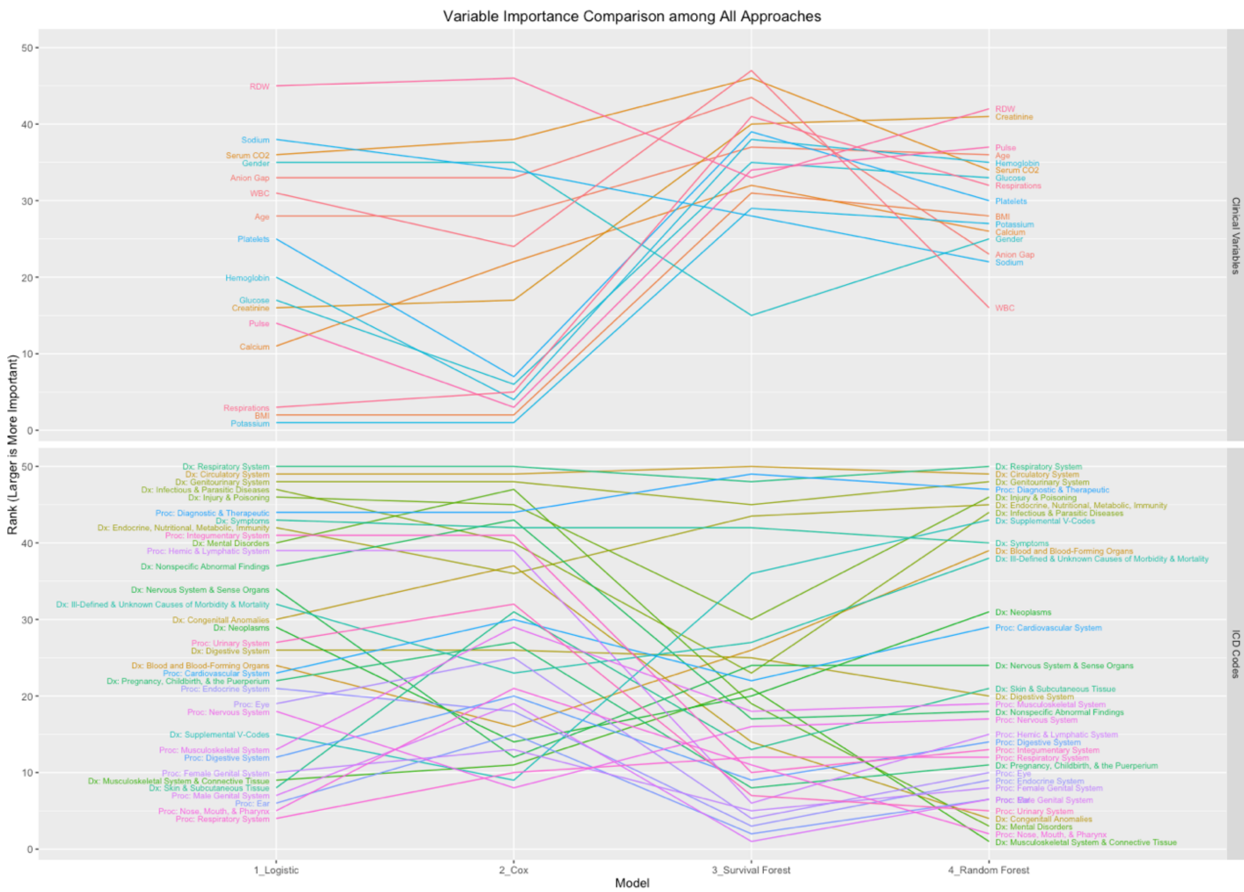



Figure 3.7: Comparison of variable importance rankings among modeling strategies.

### Data Analysis Software

R packages used for analysis include:

rms[89]

Hmisc[90]

ggplot2[91]

data.table[92]

dplyr[93]

tidyr[94]

knitr[95]

ROCR[96]

directlabels[97]

pROC[98]

randomForest[99]

randomForestSRC[100]

caret[101]

ggRandomForests[102]

**Formulas**

In all modeling approaches, the predicted cardiopulmonary arrest event $E$ is said to occur if the probability estimate $\hat{Y}$ meets or exceeds the threshold $c$, set at the event rate (0.006) and several of its multiples.

$$E = \begin{cases} 1, & \text{if } \hat{Y} \geq \text{c} \in \{0.006, 0.012, 0.018, 0.06, 0.12\} \\ 0, & \text{otherwise} \end{cases}$$

This formulation creates a binary classification for direct comparison of predicted events $E$ with actual events $A$ in a sample of $n$ patients with the following metrics:[103,104]

$$\text{Sensitivity (Recall, True Positive Rate)} = \frac{\sum(E = 1|A = 1)}{\sum A} \tag{3.1}$$

$$\text{Positive Prediction Rate} = \frac{\sum E}{n} \tag{3.2}$$

$$\text{Positive Predictive Value (Precision)} = \frac{\sum(E = 1|A = 1)}{\sum E} \tag{3.3}$$

$$\text{False Positive Rate} = \frac{\sum(E = 1|A = 0)}{\sum n - A} \tag{3.4}$$

$$\text{F1 Score} = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3.5}$$

The area under the receiver operating characteristic curve metric $AUC$ was calculated with a trapezoidal approximation using a plot comparing the false positive rate $FPR$ to the true positive rate $TPR$ at each unique predicted probability $i$ in $\{\hat{Y}\}$.

$$AUC = \sum_{i \in \{2,3,\ldots|\hat{Y}|\}} \frac{1}{2}(FPR_i - FPR_{i-1})(TPR_i + TPR_{i-1}) \tag{3.6}$$

*Logistic Regression*

Probability estimates for logistic regression models given a vector of coefficients $\beta$ and new data $X$ are calculated by:[105]

$$\hat{Y} = \frac{1}{1 + \exp^{-X\beta}} \tag{3.7}$$

*Cox Proportional Hazards Regression*

Probability estimates for Cox proportional hazards regression models require a specification of the time $t$ to which a survival probability at that time point $\hat{S}_t$ is calculated. Along with the vector of coefficients $\beta$ and the new data $X$, the formulation is:[105]

$$\begin{aligned} \hat{Y}_t &= 1 - \hat{S}_t \\ &= 1 - S_0(t)^{\exp(X\beta)} \end{aligned} \tag{3.8}$$

In this study, $t = 2$ was used for comparisons because that was the median time to both the event and censoring.

*Random Forests*

For each of the R trees $T_r$ and new data $X$, the event probability $\hat{Y}$ becomes:[79]

$$\hat{Y} = \frac{1}{R} \sum_{r=1}^{R} T_r(x) \tag{3.9}$$

*Random Survival Forests*

Similar to the Cox proportional hazards regression model, we must specify a time $t$ at which to calculate a survival probability $\hat{S}_t$. For each of the R trees $T_r$ and new data $X$, the event probability $\hat{Y}$ becomes:[79]

$$\begin{aligned} \hat{Y}_t &= 1 - \hat{S}_t \\ &= 1 - \frac{1}{R}\sum_{r=1}^{R} T_{r,t}(x) \end{aligned} \tag{3.10}$$

Once again, $t = 2$ was used, as it was the median time.

Chapter 4

# PARTICIPATORY DESIGN OF PROBABILITY-BASED DECISION SUPPORT TOOLS FOR IN-HOSPITAL NURSES

This chapter describes a qualitative study using a participatory design method in a simulation laboratory to identify the important design elements of probability-based clinical decision support tools. The results of this study will lead to future prototype refinement for the design and development of probability-based clinical decision-support tools.

## 4.1 Introduction

Sound clinical decision-making depends on one's ability to access, process, and use the array of information at one's disposal. Growing complexity in healthcare has made decision-making in today's clinical environment more challenging than ever. Increased introduction of information-providing technology into the clinical setting has added to this complexity and influences the decision-making of clinicians.[17] Clinical decision support (CDS) tools are intended to assist decision-making, but the rapidity of technological advancements has outpaced our knowledge of tool use, design display, and decision-making influence in the clinical environment.[21]

Probability-based clinical decision support (PB-CDS) tools (referred to by some as predictive analytics) are relatively new phenomena for influencing clinical decision-making. To facilitate the study of these tools' benefits, an initial approach would be to study a clinical situation where clinicians' prompt decision and actions are warranted. Additionally, the prototypic situation would include an outcome where the anticipated and actual events occur close together in order to minimize the potential influence of additional variables (e.g., other clinicians' actions or non-hospital factors) on weakening the temporal connection between the probability and actual occurrence of events. Therefore, using currently available data to predict events likely to occur within 24-48 hours would be ideal. The risk of cardiopulmonary arrest meets this criterion and served as the clinical situation for this study. Cardiopulmonary arrests are an exemplar by which to study PB-CDS phenomena because they are common (approximately 209,000 hospitalized patients in the United States every year[5]) and have substantial associated mortality (survival rates are only 23-37%[22]). Cardiopulmonary arrests occurring outside of the intensive care unit (ICU) are of particular interest because these in-hospital events might be preventable, or at least survivable with early intervention.[22,25,24]

Published reports demonstrating the accuracy of predictive analytic models in healthcare exist,[28,10] and many of these have been developed to predict cardiopulmonary arrests.[35,36,37,38,39,40] Most of these reports, however, are focused on how the statistical models were developed, the variables included in the models, and how accurate the models are.[28,10,29,30,31] These factors are necessary but insufficient to influence patient outcomes because a change in clinician behavior is also required for patient care to be impacted. Some studies have simply provided clinicians with information from the predictive model[42] while others have attempted to automatically initiate an intervention such as a rapid response team.[41] Even though studies examining the impact of predictive models on identification and management of patients preceding cardiopulmonary arrest have demonstrated high accuracy,[42] especially when compared to traditional scoring systems,[39] the CDS tools lack demonstrable benefit on patient outcomes outside of modest improvements in length of stay.[42,41] One reason for this inefficacy could be that most studies progress directly from model development to implementation in the clinical environment without adequate preliminary testing.[88,106]

Our study took what Friedman calls a "small ball" approach[43] to developing an information resource and challenges previous research approaches by exploring the user interface in a simulated environment before introducing the tool in clinical practice. This approach permits assessment of clinician preferences as well as modifications of the PB-CDS tool before significant resources have been spent. The overall objective of this study was to perform laboratory development and testing of a PB-CDS tool for cardiopulmonary arrest identification. In this paper we report our findings on the similarities/differences among information preferences of bedside nurses, charge nurses, and rapid response team nurses during the design phase.

## 4.2 Methods

### 4.2.1 Design

Three participatory designs were used in this study. This type of design is a qualitative method that engages participants as co-investigators in the design process; it is used widely in technology and engineering fields but only recently applied in healthcare informatics research.[55,107,108] Three phases of activities comprise a participatory design study: Priming, Designing, and Debriefing. The Priming activity helps participants understand the intended tasks and context surrounding the study purpose while preparing them to become active participants in the next phase. The Designing activity is the more active portion of the study where all participants (i.e., researchers, designers, intended end-users) co-create design elements of the tool. The Debriefing activity allows participants to describe their experience of creating and reflect on the words and actions of others. The final product is a report, possibly with prototypes.[109,55] Our study was reviewed and approved by the Vanderbilt University Institutional Review Board.

### 4.2.2 Participants and Setting

The participants in this study were nurses working in an adult teaching hospital, a pediatric teaching hospital, and an adult federal hospital. Inclusion criteria comprised bedside nurses and charge nurses working in non-critical care, inpatient departments (e.g., medical wards, surgical wards) with either adults or children from an academic medical center and a federal hospital in a large urban city in the mid-South region of the U.S. We also included nurses working in intensive care units who responded to rapid response team calls. Participants received a $75 gift card for their participation. Data collection occurred in the Vanderbilt University School of Nursing Simulation Center, which houses high and low fidelity simulation manikins, specially trained personnel for operating the manikins, several patient rooms that mimic a hospital unit, and a large, open space for small group work. High-fidelity manikins were capable of connecting to continuous telemetry monitoring, receiving general physical assessments (e.g., chest rise and palpable pulses), and communicating with participants.

### 4.2.3 Participatory Design Sessions

Three participatory design sessions were conducted. Each 2-hour session comprised 5-10 end-users currently working as either bedside nurses, charge nurses, or rapid response team nurses. Facilitated by at least two of the researchers, each session contained a Priming activity (approximately 20 minutes in length), a Designing activity (approximately 60 minutes), and a Debriefing activity (approximately 30 minutes).

#### *4.2.3.1 Priming Activity*

Each session began with a Priming activity. During this time, we gathered all participants into a conference room and watched an 8-minute video vignette in which a patient experienced clinical deterioration warranting activation of a rapid response team. The vignette scenario was modified from a video created by the Agency for Healthcare Research and Quality,[110] was validated for content by researchers (n=5) and subject matter experts (n=7), and is available from the author upon request. During the Priming activity, we instructed the nurses to take notes on what they observed/remembered from the vignette along with additional information they would have requested in a real scenario. Following the video, we collected those notes to include in data analysis.

#### *4.2.3.2 Designing Activity*

The Designing activity engaged the nurses in the hands-on creation of a physical representation of an electronic CDS tool using paper, colored pencils, scissors, rulers, and adhesive note paper. Bedside nurses were physically located near patient manikins in the simulation laboratory, charge nurses were located farther from the manikins but within eyesight of bedside nurses, and rapid response team nurses remained in the nearby conference room where the Priming activity occurred. The simulation laboratory included three low-fidelity manikins and one high-fidelity manikin, the latter representing the patient described in the Priming activity vignette and operated by a laboratory staff member. A brief narrative of each patient's history and physical assessment along with vital signs, laboratory values, and the numerical result of a fictitious PB-CDS tool (e.g., 1.3%, 56.8%) were available at each bedside. We provided an abbreviated overview of all patients (including the results of a fictitious CDS tool) to charge nurses. We gave no patient information to the rapid response team. Researchers interacted with all nurse end-user participants throughout the Designing activity, and examples of previously published cardiopulmonary arrest CDS tools were available to assist with brainstorming.

Approximately halfway through the Designing activity, the high-fidelity manikin experienced a deteriorating condition, and researchers encouraged the group of bedside nurses to ask for help from the group of charge nurses and rapid response team nurses. The first participatory design session did not produce the degree of CDS tool-focused design data we expected, so during the second and third sessions, we attempted to induce cognitive dissonance with one of the fictitious results by providing a very high numerical result to a patient whose history and physical assessment suggested a very low probability of cardiopulmonary arrest.

#### *4.2.3.3 Debriefing Activity*

After the Designing activity, all participants returned to an adjacent conference room for the Debriefing activity. During the Debriefing activity, researchers used semi-structured, open-ended questions to ask nurse end-users to share their designs and provide rationale for each of the chosen design elements. We audio-recorded the debriefing conversations, took notes of the discussion, and captured photos of physical artifacts. In all participatory design sessions following the first, we shared concepts and photos from previous sessions with participants, offering the opportunity for convergence of designs.

### 4.2.4 Analysis

Consistent with usability testing principles, the research team leveraged theme-based content analysis, ongoing aggregation of results, as well as discussion and deliberation of nurse end-user comments and artifacts.[111] We used a computer-based qualitative data analysis software program (Dedoose[60]) to facilitate deliberation among researchers. After a preliminary analysis was conducted, we collaborated with human-computer interaction and design experts to provide heuristic insights into design elements. We synthesized all recommendations, developed a low-fidelity prototype, and shared the prototype with 14 of the nurse end-users who participated in the workshops. We used this final step as a form of "member checking"[112] to ensure participants felt their preferences were appropriately converted into the prototype design.

## 4.3 Results

Five bedside nurses, nine charge nurses, and six rapid response team nurses (n=20) attended the sessions from 14 unique units. In addition to several minor themes identified in the Priming activity notes, three major themes and several considerations for design elements of a PB-CDS tool surfaced.

### 4.3.1 Themes

End-user notes taken during the Priming activity described a need for communication, bedside nurse autonomy, attention to the patient's physical assessment, review of historical vital signs and laboratory values, timing of treatments, and standardization of actions. Three major themes emerged from the Designing and Debriefing activities and represent participants' goals for the CDS tool.

#### *4.3.1.1 Goal #1: Communication of Patient Status*

First, participants reported they wanted a CDS tool that "paints a picture" or "tells the story" of the patient condition over time. They requested the ability for individual users to select which variables become visible and layer those variables' trends for hypothesis generation and succinct communication. For example, the electronic health record could provide a visual depiction of heart rate values layered over the probability-based cardiopulmonary arrest summary value. One participant noted:

> *I like the idea that you could see the trending vital signs during that rapid response call, like we started here and this is where we're going, so you can easily see at a glance. Have things go up or down. We mentioned seeing the interventions, like a little tab, where you just tap - "Look, IV fluids given and who did an EKG" ... Timing, to go with it, so you can see where it goes and all that's trending. That way anybody that walks into the room, they can easily see what's going without asking a bunch of questions, repeating the story every time...*

#### *4.3.1.2 Goal #2: Promotion of Autonomy*

The concepts of empowerment and autonomy surfaced in the second goal. If a CDS tool is designed well, the tool could empower nurses and promote their autonomy in advocating for the patient and contributing to treatment decisions. As an objective assessment of the patient's condition, the CDS tool has the potential to provide participants with a structured method by which nurses' can garner support for their recommendations.

#### *4.3.1.3 Goal #3: Consistency with Context*

In the third goal, nurses were in agreement that the model had to make sense, and the general perception was that probability-based models are more helpful for confirming what one already thinks rather than identifying unrecognized patient conditions. If the CDS tool provides results that are discrepant with what one thinks or does not appear to consider a patient's "context" or "baseline," the CDS tool prompts many questions, which has potential for both benefit and harm. To paraphrase several of the participants, one of the researchers noted in a post-workshop discussion:

> *. . . changes in the number need interpretation. Why or what contributed to a rapid change? . . . a slow steady trend also needs interpretation. . . Don't let a tool overtake critical thinking. It's all about the trends and the baseline.*

### 4.3.2 Design Elements

A list of design elements requested by nurse end-users accompanied by heuristic recommendations are provided in Table 4.1. Participants frequently expressed a desire for the ability to visualize the temporal trend of the predicted probability of the outcome along with user-selected overlapping depictions of vital signs, laboratory values, and outcome-related treatments and interventions. Charge nurses and rapid response team nurses had a strong request for only viewing a ranked order of the highest risk patients. Less notable but fairly commonly-heard requests included alerts only for values exceeding an absolute threshold or high degree of change, a green/yellow/red color scheme, and the ability to view the tool on both a mobile device as well as a dashboard.

Table 4.1: Design Element Considerations for PB-CDS Tools

| Elements | Participant Preferences | Heuristic Evaluation |
|---|---|---|
| **Trends** | Desire current value in addition to historical trends • Want to know when previous values where acquired • Trends should be linear (not circular) • Moving 12-hour window | Consider combinations of color-coding and ranking. |
| **Layers/Filters** | Ability to select which variables to include • Vital sign values most preferred (laboratory values mentioned far less than vital signs) • Defaults can be unit-specific or the variables that contribute "most" to a change or high probability • Ability to see any variables that contribute to the model as well as anything abnormal (even if it doesn't contribute to the statistical model) | |
| **Treatments & Interventions** | Want to see what's been done to mitigate event risk • Selected treatments should relate to problem being viewed (e.g., antibiotics for sepsis but not falls) | Might need to be unit-specific. |
| **Ranking** | Rank patients in descending order of probability • Might not work without a consideration of "context/baseline" • Would need different view for charge nurse vs. bedside nurse vs. RRT (prefer to see only those in one's care) | |
| **Alert Notification** | See/read *why* the prediction score changed (i.e., what individual value[s] changed) • Accompany alerts with a recommended action • Review tool at beginning of shift (e.g., during shift change) & then be notified of changes • Alerts for exceeding an absolute value threshold as well as percent change • Should be specific to unit/department | Consider building statistics for 12-24 hours early so that nurses are "helping the next shift out" as opposed to "depending on a statistical model to tell them how to do their job" |
| **Color Scheme** | Red/Yellow/Green acceptable if also including the actual number • Several requested flashing/blinking | Consider color-blind persons • Font size to represent magnitude • Flashing *not* recommended |
| **Medium** | Dashboard displays, especially for low-risk patients • Mobile-friendly option (e.g., cell phone) • Prefer information available at the bedside for RRT arrival (possibly something where RRT could obtain information while en route) • Ability to click elements or "zoom in" to see details | |
| **Communication** | Capturing data in real time from EHR • Ability to send screenshots to EHR/chart, RRT, provider, and/or charge nurse | |

*Notes:* EHR = electronic health records; RRT = rapid response team

Although not always focused solely on our CDS tools, participants gave additional recommendations for future technology development and identified potential barriers (see Table 4.2). The most prominent findings include ensuring the tool is readily available to all healthcare team members, balancing ease of information access with patient privacy, and being concerned about discrepancies in objective probabilities and subjective perceptions. Regarding this latter point, some participants expressed concern of the potential for over-reliance on CDS tools with a loss in critical thinking as these tools become more common.

Table 4.2: Recommendations for Future Technology Development & Potential Barriers

| | **Participant Request** | **Heuristic Perspective** |
|---|---|---|
| **Other Desired Features** | Voice-activation • Providing risk scores for multiple outcomes - could treat these as filters with selection of what one wants to see. • All clinicians should have opportunity to view • Live video stream once RRT activated • Creation of a summary paragraph of the problem (similar to a History & Physical Note) • Show how reliable prediction score is (e.g., confidence intervals) • A few participants mentioned wanting to know who was involved in the patient's support system (i.e., family) • "Start" button for when activating RRT- could more thoroughly record all that happens as well as provide recent history | Training will be important for clinicians to successfully access & use tool. Family involvement will be challenging to incorporate, but pediatric nurses consider this factor in decision-making. |
| **Barriers** | Ability to select which variables to include • Vital sign values most preferred (laboratory values mentioned far less than vital signs) • Defaults can be unit-specific or the variables that contribute "most" to a change or high probability • Ability to see any variables that contribute to the model as well as anything abnormal (even if it doesn't contribute to the statistical model) | |

*Notes:* RRT = rapid response team

### 4.3.3 Prototype Development

Figures 4.1, 4.2, and 4.3 illustrate screenshot examples of a prototype to represent the most salient preferences from participants. Consistent with requests for ranking, Figure 4.1 provides a prototype of what a charge nurse might use to review a list of all patients on that unit, ranked in descending order of risk to promote easy recognition of high-risk patients. In order to illustrate individual patient trends and accompanying "baseline," Figure 4.2 displays how all types of nurse end-users preferred to see an individual patient's risk. Combining the most salient themes of trend lines, filters, layers, and treatments, Figure 4.3 exemplifies several screenshots: (a) vital signs and laboratory values layered over a predicted probability of cardiopulmonary arrest, (b) additional detail of one vital sign selected, and (c) cardiopulmonary arrest-related interventions layered over the time period in which they occurred. Participatory design session end-users who reviewed the prototype did not recommend any changes for the current design; however, they did provide additional suggestions for future implementation including a desire for automaticity of data exchange with the electronic health record and individual configuration of all filters and layers.

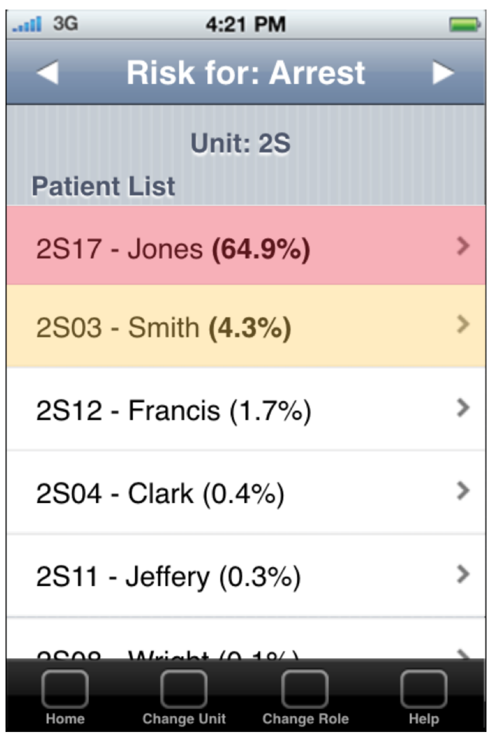



Figure 4.1: Prototype of charge nurse view of all patients on a unit, ranked in descending order of those most at risk for a cardiopulmonary arrest.

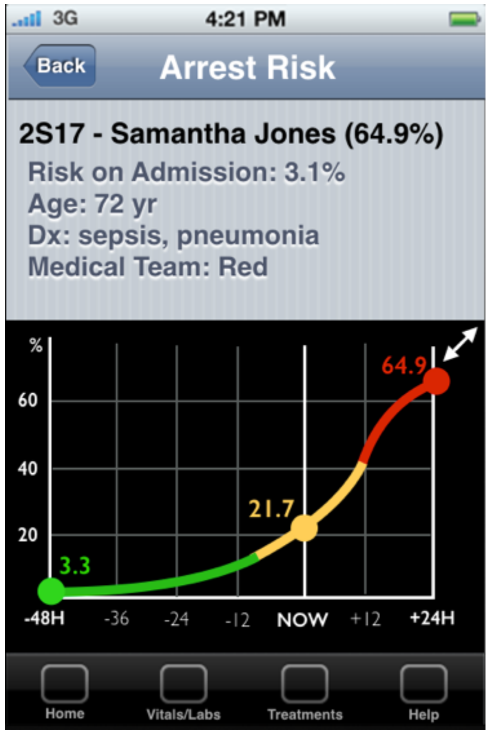



Figure 4.2: Prototype of individual patient view containing basic patient information accompanied by a 72-hour trend of predicted probability of cardiopulmonary arrest.

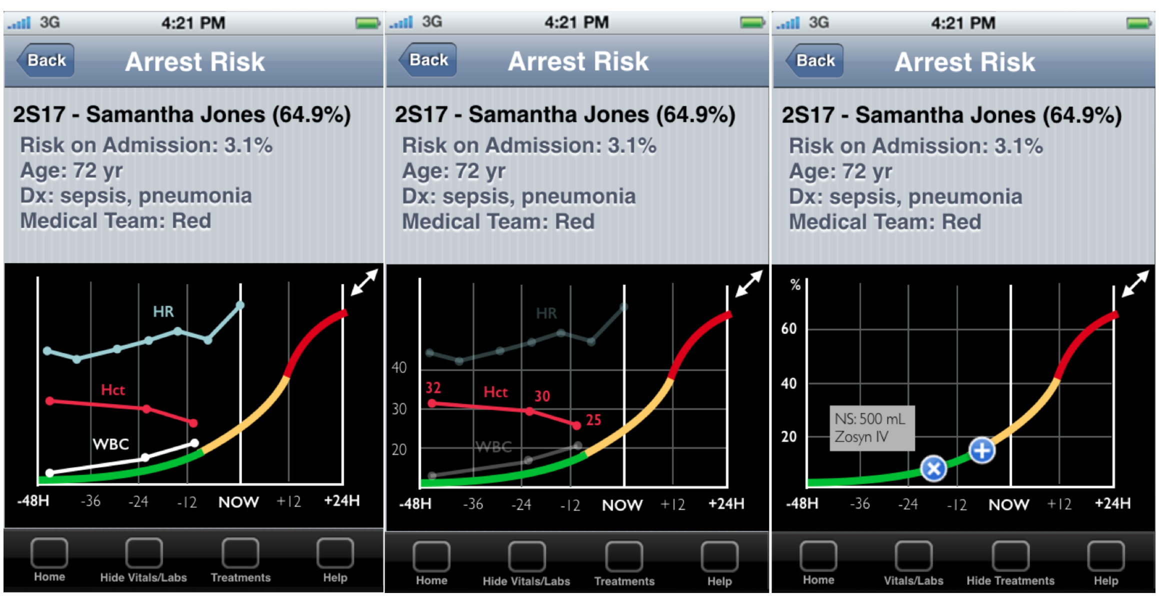



Figure 4.3: Prototype of applying filters and layers to predicted probability of cardiopulmonary arrest. *Left*: User-selected vital signs and laboratory values displayed. *Center*: Additional detail of one variable among user-selected vital signs and laboratory values. *Right*: Cardiopulmonary arrest-related preventive treatment displayed on the clinical shift in which they occurred.

## 4.4 Discussion

We used the participatory design method to identify important design elements and create a prototype for a PB-CDS tool that predicts cardiopulmonary arrests. Co-creation of the CDS tool via participatory design was beneficial because the active involvement of multiple stakeholders facilitated the identification of novel, integrative design concepts that groups of participants (i.e., researchers and end-users) might not have identified separately. The goals and associated recommended design elements can be used and extended by designers and developers to create PB-CDS tools for in-hospital nurses in the future.

Similar to how the information needs of nurses differ from physicians due to their different practice and diagnostic models,[113,20] we had hypothesized the needs of various specialties of participants would differ, given their different settings and work. Work performed by others has demonstrated that nurses benefit from information displays focused on trends and the recall of relevant patient information while physicians benefit from displays that promote inference for decision-making.[20] We did not find such distinct differences among our participants, even though we anticipated charge nurses and rapid response team nurses would prefer more inference-focused displays for their discrete decision-making activities. We found the desire for inference-support and hypothesis-generation assistance to be present across all participant roles. When considering this desire paired with participants' request for exploring a patient's "baseline" and "context," our findings appear consistent with the view that nurses' diagnostic reasoning skills are context-dependent in the social and humanistic domains.[113] Finally, a recent simulation study of nurses' acceptance of CDS suggestions found the primary reason to accept a suggestion resulted from the belief that it was "good for the patient,"[114] and we believe this supports our theme of "consistency with context" as a CDS tool goal.

### 4.4.1 Strengths and Limitations

Strengths of the study include placing participants in an environment that mimics real workflows, recruiting three unique roles of in-hospital nurses, and iterative design testing in collaboration with end-users, researchers, and human-computer interaction experts. We condensed preferred design elements into a mobile-phone based prototype due to participant's requests for a mobile-friendly tool; this will hopefully ease the transition to computer monitor-sized displays (in contrast to the removal of key elements during screen size reduction). When reviewing the Priming activity notes provided by end-user participants at the beginning of the sessions, we treated these as a type of needs assessment, and we believe our recommended design elements and prototype would contribute to meeting these needs. Limitations of the study include a convenience sample from only three hospitals located in two health care systems and the inability to determine if CDS tool-based information is capable of changing behavior.

### 4.4.2 Future Directions

Bedside nurses and charge nurses from the intensive care unit and emergency department were excluded because the workflows of these nurses are different from those of non-critical care, inpatient nurses. Future studies should include these settings for possible identification of unique design elements of CDS tools in these environments. The terms *baseline*, *context*, and *sick* surfaced frequently and are likely specific to healthcare clinicians and perhaps even nurses or practice specialties. Future studies should explore their meanings across settings and how information technology can provide information within this mental framework. Finally, formal usability testing is needed with more robust prototypes developed from our recommended design elements. Usability testing will be especially important when comparing some design elements head-to-head. As the designs become more robust and prepare for integration into a clinical setting, we plan to crosswalk our recommendations with the recently released international standards for nursing-process-focused CDS tools, which contain additional criteria necessary for optimal integration into workflows that support practice and advance the science.[115]

## 4.5 Conclusion

The information we gained about the preferred design elements of predictive analytics tools that support, rather than interrupt, nurses' cognitive workflows can benefit future studies in this field as well as nurses' practice. As these themes and elements undergo additional testing and refinement, we anticipate they can eventually serve as standards for developing PB-CDS tools that are more likely to influence clinician behavior and ultimately patient outcomes.

Chapter 5

# CONCLUSION

The goal of this initial inquiry was to identify strategies for designing and developing probability-based clinical decision support (PB-CDS) tools for nurses. Recognition of in-hospital cardiopulmonary arrest (IHCPA) served as the clinical context for the studies. Given the paucity of literature regarding optimal design and development strategies of these tools that could influence patient outcomes, these dissertation findings add to the corpus of knowledge and create a foundation for future research in this field. As more PB-CDS tools become available with the growth in predictive analytics, nurses (and other clinicians) are in need of tools that integrate well with their physical and cognitive workflows if they are to make the best use of these tools. If PB-CDS tools are simply inserted into the clinical environment without a consideration of key elements for success (e.g., real-time data entry practices, current information-gathering behaviors, perceptions and use of probability-based results), they might not be able to produce the desired change in clinician behaviors. It stands to reason that without these behavioral changes, it becomes challenging to improve patient outcomes. In this work's qualitative interviews, nurses' current information-gathering behaviors were identified, and their perception and use of probability-based information was explored. The statistical prediction study compared classification and time-to-event approaches to developing prediction models for IHCPA, which will inform future evaluation studies of PB-CDS tool interpretability. The participatory design sessions captured key design elements for PB-CDS tools that could continue to be explored in development and evaluation of more robust tools. Given the findings and limitations of these studies, several research gaps have surfaced to be explored in the future.

## 5.1 Gaps to Address

### 5.1.1 Design of PB-CDS Tools

The qualitative study of information-gathering behaviors was limited by the inherent challenge of participant recall, and the qualitative study that leveraged participatory design was limited by the sterility of its simulated environment. Both of these limitations were expected and acceptable given that design research tends to be formative, rather than summative, in nature.[116] These approaches were beneficial given the novel nature of this work (i.e., existing tools were not already available for study). Once PB-CDS tools are developed, summative methods (e.g., direct observation of actual use and usability surveys) will be more helpful in evaluation. Future work will focus on evaluating nurses' routine work in simulated environments where they are provided with varying PB-CDS tool designs. Examining responses to those various designs would permit a comparison of which design elements contribute to improved usability and are more likely to alter their practice behaviors. This incremental build on tool design affords the opportunity of making iterative changes before significant resources are spent on a final product.

Following any iterative changes, usability in real-world scenarios will be conducted. In clinical environments (as in simulated environments), formal usability can be conducted with methods such as usability surveys, direct observation, "think aloud" activities, and cognitive walkthroughs. Non-simulated environments are accompanied by additional confounders but provide the best context for evaluating whether a design will perform as expected in its intended environment. As evidence accumulates for optimal designs of PB-CDS tools, synthesizing results and publishing recommendations will be helpful in guiding designers and programmers during future tool development.

### 5.1.2 Development of PB-CDS Statistical Models

The statistical study compared classification and time-to-events models for building prediction models for IHCPA, but a clinically useful model could not be developed due to the large amount of missing data. A more robust dataset with less missing data is essential for developing a clinical prediction model that has potential for use in a clinical environment. Once a more robust dataset is obtained and the statistical performance of the varying approaches compared again, a formal evaluation of clinicians' abilities to, and preferences for, interpretation will be conducted. A balance between high statistical performance and ease of interpretability will likely be needed for developing PB-CDS tools with the greatest potential for influencing clinicians' practice behaviors.

### 5.1.3 Implementation of PB-CDS Tools

This dissertation work has focused on tools that support decision-making in the sense of recognition, rather than action. Although *recognition* of a problem must precede its *management*, PB-CDS tools that do not provide clinicians with recommended actions will be of little value in the clinical environment. A simple solution is the incorporation of recommended actions identified by key stakeholders and subject matter experts. However, the clinical decision support field would benefit from the application of decision science methods to construct models providing individualized patient information on the benefit of each possible treatment. The conceptually and computationally complex problem of identifying the optimal treatments for individual patients is one of the next steps in the field of big data. Given that researchers can already create prediction models assigning baseline risk, future efforts should focus on assigning projected risks following diverse treatment options. If those projected risks could be reviewed along with each possible outcome's utility, this may enhance the opportunity for shared decision-making between clinicians and patients/families.

Another challenge of implementation efforts is an appreciation of the context into which new interventions are placed. Decision support interventions that comprise context-specific features and accommodate clinicians' competing priorities are more likely to be successful in today's complex and resource-constrained clinical environments. New interventions necessitate work environment and/or process changes, thus examining a number of covariates, including interpersonal communication, organizational culture, and limited resources, will increase the likelihood of PB-CDS tool adoption. Indeed, many informaticians and health services researchers have formally recognized these challenges within sociotechnical frameworks, and quality improvement groups are attempting to address these concerns.[88,117] A remaining research gap, however, is the synthesis of best practices for successful implementation of novel tools in varying contexts (an area where implementation science could be particularly helpful). For example, developing context-specific and context-agnostic implementation recommendations via the exploration of successful and unsuccessful implementation efforts across multiple settings would be invaluable.

#### 5.1.4 Evaluation of Patient Outcomes

Several methodological challenges for evaluating PB-CDS tools' influence on patient outcomes exist[1] and need to be addressed. First, the multilevel nature of information resource interventions will require cluster-randomized trials (as opposed to the more common patient-level randomization) to examine patient outcomes. Stepped wedge designs are particularly appealing for those interventions because they include a cluster design while permitting iterative changes in the intervention (and its implementation) during spread. Second, treatment fidelity has not been consistently reported in the literature, and novel methods might be needed to measure the degree to which PB-CDS tools are used by clinicians in the experimental arms of trials. Finally, in order to influence patient outcomes, interventions will require not only a helpful tool for problem recognition but also an effective treatment to implement. This synergy of timely recognition and appropriate management will create the best opportunity for improving patient outcomes.

## 5.2 Contributions to Science and Nursing

Strategies for designing, developing, and implementing PB-CDS tools for nurses have not been thoroughly explored or understood. Rigorous evaluation at each of these stages, and especially related to patient outcomes, is needed. This dissertation work has filled some of these gaps by surveying the current state of perceived information-gathering behaviors and preferred design elements for PB-CDS tools. The work also set the foundation for formal assessment of probability-based information interpretability in the context of clinical nursing work. Ultimately, this work contributes to assisting clinicians in obtaining the right information at the right time to help them make the best decisions with patients and their families.

REFERENCES


[1] A. D. Jeffery. Methodological challenges in examining the impact of healthcare predictive analytics on nursing-sensitive patient outcomes. *Comput Inform Nurs*, 33(6):258–264, 2015.

[2] U.S. Department of Health and Human Services. Health IT adoption rates. http://www.healthit.gov/policy-researchers-implementers/health-it-adoption-rates, January 2013.

[3] T. B. Murdoch and A. S. Detsky. The inevitable application of big data to health care. *JAMA*, 309(13):1351–2, 2013.

[4] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff (Millwood)*, 33(7):1123–31, 2014.

[5] R. M. Merchant, L. Yang, L. B. Becker, R. A. Berg, V. Nadkarni, G. Nichol, B. G. Carr, N. Mitra, S. M. Bradley, B. S. Abella, P. W. Groeneveld, and Investigators American Heart Association Get With The Guidelines-Resuscitation. Incidence of treated cardiac arrest in hospitalized patients in the united states. *Crit Care Med*, 39(11):2401–6, 2011.

[6] S. Girotra, B. K. Nallamothu, J. A. Spertus, Y. Li, H. M. Krumholz, P. S. Chan, and Investigators American Heart Association Get with the Guidelines-Resuscitation. Trends in survival after in-hospital cardiac arrest. *N Engl J Med*, 367(20):1912–20, 2012.

[7] D. Berger. A brief history of medical diagnosis and the birth of the clinical laboratory.

part 1–ancient times through the 19th century. *MLO Med Lab Obs*, 31(7):28–30, 32, 34–40, 1999.

[8] R. Amarasingham, R. E. Patzer, M. Huesch, N. Q. Nguyen, and B. Xie. Implementing electronic health care predictive analytics: considerations and challenges. *Health Aff (Millwood)*, 33(7):1148–54, 2014.

[9] A. R. Weil. Big data in health: a new era for research and patient care. *Health Aff (Millwood)*, 33(7):1110, 2014.

[10] G. S. Collins, J. A. de Groot, S. Dutton, O. Omar, M. Shanyinde, A. Tajar, M. Voysey, R. Wharton, L. M. Yu, K. G. Moons, and D. G. Altman. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol*, 14:40, 2014.

[11] M. M. Churpek, T. C. Yuen, C. Winslow, D. O. Meltzer, M. W. Kattan, and D. P. Edelson. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Crit Care Med*, 44(2):368–74, 2016.

[12] E. W. Steyerberg. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer, New York, NY, 2009.

[13] O. Neill ES, N. M. Dluhy, and E. Chin. Modelling novice clinical reasoning for a computerized decision support system. *J Adv Nurs*, 49(1):68–77, 2005.

[14] G. Klein and D. Klinger. Naturalistic decision making. *Human Systems IAC*, 11(3):16–19, 1991.

[15] T. R. Clancy. Technology and complexity: trouble brewing? *J Nurs Adm*, 40(6):247–9, 2010.

[16] M. Macdonnell and A. Darzi. A key to slower health spending growth worldwide will be unlocking innovation to reduce the labor-intensity of care. *Health Aff (Millwood)*, 32(4):653–60, 2013.

[17] S. Dekker. Complexity, signal detection, and the application of ergonomics: Reflections on a healthcare case study. *Applied Ergonomics*, 43(3):468–472, 2012.

[18] C. L. Fillmore, B. E. Bray, and K. Kawamoto. Systematic review of clinical decision support interventions with potential for inpatient cost reduction. *BMC Med Inform Decis Mak*, 13:135, 2013.

[19] M. W. Jaspers, M. Smeulers, H. Vermeulen, and L. W. Peute. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. *J Am Med Inform Assoc*, 18(3):327–34, 2011.

[20] C. M. Johnson and J. P. Turley. The significance of cognitive modeling in building healthcare interfaces. *Int J Med Inform*, 75(2):163–72, 2006.

[21] K. Mastrian and D. McGonigle. Cognitive informatics: an essential component of nursing technology design. *Nurs Outlook*, 56(6):332–3, 2008.

[22] A. S. Go, D. Mozaffarian, V. L. Roger, E. J. Benjamin, J. D. Berry, M. J. Blaha, S. Dai, E. S. Ford, C. S. Fox, S. Franco, H. J. Fullerton, C. Gillespie, S. M. Hailpern, J. A. Heit, V. J. Howard, M. D. Huffman, S. E. Judd, B. M. Kissela, S. J. Kittner, D. T. Lackland, J. H. Lichtman, L. D. Lisabeth, R. H. Mackey, D. J. Magid, G. M. Marcus, A. Marelli, D. B. Matchar, D. K. McGuire, E. R. Mohler, C. S. Moy, M. E. Mussolino, R. W. Neumar, G. Nichol, D. K. Pandey, N. P. Paynter, M. J. Reeves, P. D. Sorlie, J. Stein, A. Towfighi, T. N. Turan, S. S. Virani, N. D. Wong, D. Woo, M. B. Turner, Committee American Heart Association Statistics, and Subcommittee

Stroke Statistics. Heart disease and stroke statistics–2014 update: a report from the american heart association. *Circulation*, 129(3):e28–e292, 2014.

[23] R. M. Merchant, L. B. Becker, B. S. Abella, D. A. Asch, and P. W. Groeneveld. Cost-effectiveness of therapeutic hypothermia after cardiac arrest. *Circ Cardiovasc Qual Outcomes*, 2(5):421–8, 2009.

[24] T. J. Hodgetts, G. Kenward, I. Vlackonikolis, S. Payne, N. Castle, R. Crouch, N. Ineson, and L. Shaikh. Incidence, location and reasons for avoidable in-hospital cardiac arrest in a district general hospital. *Resuscitation*, 54(2):115–23, 2002.

[25] K. Cleverley, N. Mousavi, L. Stronger, K. Ann-Bordun, L. Hall, J. W. Tam, A. Tischenko, D. S. Jassal, and R. K. Philipp. The impact of telemetry on survival of in-hospital cardiac arrests in non-critical care patients. *Resuscitation*, 84(7):878–82, 2013.

[26] C. Sandroni, G. Ferro, S. Santangelo, F. Tortora, L. Mistura, F. Cavallaro, A. Caricato, and M. Antonelli. In-hospital cardiac arrest: survival depends mainly on the effectiveness of the emergency response. *Resuscitation*, 62(3):291–7, 2004.

[27] P. J. Sharek, L. M. Parast, K. Leong, J. Coombs, K. Earnest, J. Sullivan, L. R. Frankel, and S. J. Roth. Effect of a rapid response team on hospital-wide mortality and code rates outside the ICU in a children's hospital. *JAMA*, 298(19):2267–74, 2007.

[28] G. Collins and Y. Le Manach. Multivariable risk prediction models: it's all about the performance. *Anesthesiology*, 118(6):1252–3, 2013.

[29] G. S. Collins, O. Omar, M. Shanyinde, and L. M. Yu. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. *J Clin Epidemiol*, 66(3):268–77, 2013.

[30] R. G. Ettema, L. M. Peelen, M. J. Schuurmans, A. P. Nierich, C. J. Kalkman, and K. G. Moons. Prediction models for prolonged intensive care unit stay after cardiac surgery: systematic review and validation study. *Circulation*, 122(7):682–9, 2010.

[31] D. Kansagara, H. Englander, A. Salanitro, D. Kagen, C. Theobald, M. Freeman, and S. Kripalani. Risk prediction models for hospital readmission: a systematic review. *JAMA*, 306(15):1688–98, 2011.

[32] Jun Seongsook, Jeong Ihnsook, and Lee Younghee. Validity of pressure ulcer risk assessment scales; cubbin and jackson, braden, and douglas scale. *Int J Nurs Stud*, 41(2):199–204, 2004.

[33] H. Gao, A. McDonnell, D. A. Harrison, T. Moore, S. Adam, K. Daly, L. Esmonde, D. R. Goldhill, G. J. Parry, A. Rashidian, C. P. Subbe, and S. Harvey. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. *Intensive Care Med*, 33(4):667–79, 2007.

[34] D. Oliver, F. Daly, F. C. Martin, and M. E. McMurdo. Risk factors and risk assessment tools for falls in hospital in-patients: a systematic review. *Age Ageing*, 33(2):122–30, 2004.

[35] C. A. Alvarez, C. A. Clark, S. Zhang, E. A. Halm, J. J. Shannon, C. E. Girod, L. Cooper, and R. Amarasingham. Predicting out of intensive care unit cardiopulmonary arrest or death using electronic medical record data. *BMC Med Inform Decis Mak*, 13:28, 2013.

[36] M. M. Churpek, T. C. Yuen, S. Y. Park, R. Gibbons, and D. P. Edelson. Using electronic health record data to develop and validate a prediction model for adverse outcomes in the wards. *Crit Care Med*, 42(4):841–848, 2014.

[37] G. J. Escobar, J. C. LaGuardia, B. J. Turk, A. Ragins, P. Kipnis, and D. Draper. Early detection of impending physiologic deterioration among patients who are not

in intensive care: development of predictive models using data from an automated electronic medical record. *J Hosp Med*, 7(5):388–95, 2012.

[38] R. S. Evans, K. G. Kuttler, K. J. Simpson, S. Howe, P. F. Crossno, K. V. Johnson, M. N. Schreiner, J. F. Lloyd, W. H. Tettelbach, R. K. Keddington, A. Tanner, C. Wilde, and T. P. Clemmer. Automated detection of physiologic deterioration in hospitalized patients. *J Am Med Inform Assoc*, 22:350–360, 2014.

[39] G. D. Finlay, M. J. Rothman, and R. A. Smith. Measuring the modified early warning score and the rothman index: advantages of utilizing the electronic medical record in an early warning system. *J Hosp Med*, 9(2):116–9, 2014.

[40] L. L. Kirkland, M. Malinchoc, M. O'Byrne, J. T. Benson, D. T. Kashiwagi, M. C. Burton, P. Varkey, and T. I. Morgenthaler. A clinical deterioration prediction tool for internal medicine patients. *Am J Med Qual*, 28(2):135–42, 2013.

[41] M. H. Kollef, Y. Chen, K. Heard, G. N. LaRossa, C. Lu, N. R. Martin, N. Martin, S. T. Micek, and T. Bailey. A randomized trial of real-time automated clinical deterioration alerts sent to a rapid response team. *J Hosp Med*, 9(7):424–9, 2014.

[42] T. C. Bailey, Y. Chen, Y. Mao, C. Lu, G. Hackmann, S. T. Micek, K. M. Heard, K. M. Faulkner, and M. H. Kollef. A trial of a real-time alert for clinical deterioration in patients hospitalized on general medical wards. *J Hosp Med*, 8(5):236–242, 2013.

[43] C. P. Friedman. "Smallball" evaluation: A prescription for studying community-based information interventions. *J Med Libr Assoc*, 93(4 Suppl):S43–8, 2005.

[44] P. Carayon, R. Cartmill, M. A. Blosky, R. Brown, M. Hackenberg, P. Hoonakker, A. S. Hundt, E. Norfolk, T. B. Wetterneck, and J. M. Walker. ICU nurses' acceptance of electronic health records. *J Am Med Inform Assoc*, 18(6):812–9, 2011.

[45] E. O. Im and W. Chee. Nurses' acceptance of the decision support computer program for cancer pain management. *Comput Inform Nurs*, 24(2):95–104, 2006.

[46] B. D. Winters, S. J. Weaver, E. R. Pfoh, T. Yang, J. C. Pham, and S. M. Dy. Rapid-response systems as a patient safety strategy: a systematic review. *Ann Intern Med*, 158(5 Pt 2):417–25, 2013.

[47] M. W. Jaspers. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. *Int J Med Inform*, 78(5):340–53, 2009.

[48] P. Carayon, A. Schoofs Hundt, B. T. Karsh, A. P. Gurses, C. J. Alvarado, M. Smith, and P. Flatley Brennan. Work system design for patient safety: the seips model. *Qual Saf Health Care*, 15 Suppl 1:i50–8, 2006.

[49] J. M. Brokel, T. J. Schwichtenberg, D. S. Wakefield, M. M. Ward, M. G. Shaw, and J. M. Kramer. Evaluating clinical decision support rules as an intervention in clinician workflows with technology. *Comput Inform Nurs*, 29(1):36–42, 2011.

[50] V. Venkatesh, M. G. Morris, G. B. Davis, and F. D. Davis. User acceptance of information technology: Toward a unified view. *Manag Inf Syst Q*, 27(3):425–478, 2003.

[51] D. Fallik. For big data, big questions remain. *Health Aff (Millwood)*, 33(7):1111–4, 2014.

[52] B. Hazlehurst and C. McMullen. Orienting frames and private routines: the role of cultural process in critical care safety. *Int J Med Inform*, 76 Suppl 1:S129–35, 2007.

[53] H. Kim, J. S. Sefcik, and C. Bradway. Characteristics of qualitative descriptive studies: A systematic review. *Res Nurs Health*, 40(1):23–42, 2017.

[54] M. Sandelowski. Whatever happened to qualitative description? *Res Nurs Health*, 23(4):334–40, 2000.

[55] B. Reeder, R. A. Hills, A. M. Turner, and G. Demiris. Participatory design of an integrated information system design to support public health nurses and nurse managers. *Public Health Nurs*, 31(2):183–92, 2014.

[56] The Weather Channel. http://www.weather.com/, 2014.

[57] AccuWeather. http://www.accuweather.com/, 2014.

[58] K. F. Schulz and D. A. Grimes. Case-control studies: research in reverse. *Lancet*, 359(9304):431–4, 2002.

[59] R. A. Krueger and M. A. Casey. *Focus Groups: A Practical Guide for Applied Research*. SAGE, Los Angeles, CA, 2009.

[60] Dedoose. Homepage. http://www.dedoose.com, 2016.

[61] M. Sandelowski. What's in a name? qualitative description revisited. *Res Nurs Health*, 33(1):77–84, 2010.

[62] L. L. Novak, R. J. Holden, S. H. Anders, J. Y. Hong, and B. T. Karsh. Using a sociotechnical framework to understand adaptations in health IT implementation. *Int J Med Inform*, 82(12):e331–44, 2013.

[63] R. Lipshitz and O. Strauss. Coping with uncertainty: a naturalistic decision-making analysis. *Organ Behav Hum Decis Process*, 69(2):149–163, 1997.

[64] J. S. Braaten. Original research: hospital system barriers to rapid response team activation: a cognitive work analysis. *Am J Nurs*, 115(2):22–32; test 33; 47, 2015.

[65] M. Odell, C. Victor, and D. Oliver. Nurses' role in detecting deterioration in ward patients: systematic literature review. *J Adv Nurs*, 65(10):1992–2006, 2009.

[66] J. Cioffi. Nurses' experiences of making decisions to call emergency assistance to their patients. *J Adv Nurs*, 32(1):108–14, 2000.

[67] C. G. Parker. Decision-making models used by medical-surgical nurses to activate rapid response teams. *Medsurg Nurs*, 23(3):159–64, 2014.

[68] C. A. Tanner. Thinking like a nurse: A research-based model of clinical judgment in nursing. *J Nurs Educ*, 45(6):204–211, 2006.

[69] D. Kahneman. *Thinking, fast and slow.* Farrar, Straus and Giroux, New York, NY, 2013.

[70] M. C. Wright, S. Dunbar, B. C. Macpherson, E. W. Moretti, G. Del Fiol, J. Bolte, J. M. Taekman, and N. Segall. Toward designing information display to support critical care. a qualitative contextual evaluation and visioning effort. *Appl Clin Inform*, 7(4):912–929, 2016.

[71] S. Weber. A qualitative analysis of how advanced practice nurses use clinical decision support systems. *J Am Acad Nurse Pract*, 19(12):652–67, 2007.

[72] A. N. Makam, O. K. Nguyen, and A. D. Auerbach. Diagnostic accuracy and effectiveness of automated electronic sepsis alert systems: A systematic review. *J Hosp Med*, 10(6):396–402, 2015.

[73] M. E. Smith, J. C. Chiovaro, M. O'Neil, D. Kansagara, A. R. Quinones, M. Freeman, M. L. Motu'apuaka, and C. G. Slatore. Early warning system scores for clinical deterioration in hospitalized patients: a systematic review. *Ann Am Thorac Soc*, 11(9):1454–65, 2014.

[74] L. W. Andersen, K. M. Berg, M. Chase, M. N. Cocchi, J. Massaro, M. W. Donnino, and Investigators American Heart Association's Get With The Guidelines-Resuscitation. Acute respiratory compromise on inpatient wards in the united states: Incidence, outcomes, and factors associated with in-hospital mortality. *Resuscitation*, 105:123–129, 2016.

[75] S. Jarvis, C. Kovacs, J. Briggs, P. Meredith, P. E. Schmidt, P. I. Featherstone, D. R. Prytherch, and G. B. Smith. Aggregate national early warning score (NEWS) values are more important than high scores for a single vital signs parameter for discriminating the risk of adverse outcomes. *Resuscitation*, 87:75–80, 2015.

[76] S. Jarvis, C. Kovacs, J. Briggs, P. Meredith, P. E. Schmidt, P. I. Featherstone, D. R. Prytherch, and G. B. Smith. Can binary early warning scores perform as well as standard early warning scores for discriminating a patient's risk of cardiac arrest, death or unanticipated intensive care unit admission? *Resuscitation*, 93:46–52, 2015.

[77] M. A. Kang, M. M. Churpek, F. J. Zadravecz, R. Adhikari, N. M. Twu, and D. P. Edelson. Real-time risk prediction on the wards: a feasibility study. *Crit Care Med*, 44(8):1468–1473, 2016.

[78] T. J. Moss, D. E. Lake, J. F. Calland, K. B. Enfield, J. B. Delos, K. D. Fairchild, and J. R. Moorman. Signatures of subacute potentially catastrophic illness in the ICU: model development and validation. *Crit Care Med*, 44(9):1639–1648, 2016.

[79] T. Hastie, R. Tibshirani, and J. H. Friedman. *The elements of statistical learning : data mining, inference, and prediction.* Springer, New York, NY, 2nd edition, 2009.

[80] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S. Lauer. Random survival forests. *Ann Appl Stat*, 2(3):841–860, 2008.

[81] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2016.

[82] M. M. Churpek, T. C. Yuen, and D. P. Edelson. Predicting clinical deterioration in the hospital: the impact of outcome selection. *Resuscitation*, 84(5):564–8, 2013.

[83] F. J. Zadravecz, L. Tien, B. J. Robertson-Dick, T. C. Yuen, N. M. Twu, M. M. Churpek,

and D. P. Edelson. Comparison of mental-status scales for predicting mortality on the general wards. *J Hosp Med*, 10(10):658–63, 2015.

[84] M. Xu, B. Tam, L. Thabane, and A. Fox-Robichaud. A protocol for developing early warning score models from vital signs data in hospitals using ensembles of decision trees. *BMJ Open*, 5(9):e008699, 2015.

[85] K. L. Caballero Barajas and R. Akella. Dynamically modeling patient's health state from electronic medical records: a time series approach. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 69–78. ACM, 2015.

[86] M. M. Churpek, R. Adhikari, and D. P. Edelson. The value of vital sign trends for detecting clinical deterioration on the wards. *Resuscitation*, 102:1–5, 2016.

[87] P. Kipnis, B. J. Turk, D. A. Wulf, J. Carlos LaGuardia, X. Liu V, M. M. Churpek, S. Romero-Brufau, and G. J. Escobar. Development and validation of an electronic medical record-based alert score for detection of inpatient deterioration outside the ICU. *J Biomed Inform*, 2016.

[88] K. Dunn Lopez, S. M. Gephart, R. Raszewski, V. Sousa, L. E. Shehorn, and J. Abraham. Integrative review of clinical decision support for registered nurses in acute care settings. *J Am Med Inform Assoc*, 2016.

[89] F. E. Harrell Jr. *rms: Regression Modeling Strategies*, 2017. R package version 5.1-0.

[90] F. E. Harrell Jr and with contributions from C. Dupont. *Hmisc: Harrell Miscellaneous*, 2016. R package version 4.0-2.

[91] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.

[92] M. Dowle and A. Srinivasan. *data.table: Extension of ‘data.frame‘*, 2017. R package version 1.10.4.

[93] H. Wickham and R. Francois. *dplyr: A Grammar of Data Manipulation*, 2016. R package version 0.5.0.

[94] H. Wickham. *tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions*, 2017. R package version 0.6.1.

[95] Y. Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*, 2016. R package version 1.15.1.

[96] T. Sing, O. Sander, N. Beerenwinkel, and T. Lengauer. ROCR: visualizing classifier performance in R. *Bioinformatics*, 21(20):7881, 2005.

[97] T. D. Hocking. *directlabels: Direct Labels for Multicolor Plots*, 2015. R package version 2015.12.16.

[98] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J. C. Sanchez, and M. Müller. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12:77, 2011.

[99] Andy Liaw and Matthew Wiener. Classification and regression by randomforest. *R News*, 2(3):18–22, 2002.

[100] H. Ishwaran and U. B. Kogalur. *Random Forests for Survival, Regression and Classification (RF-SRC)*, 2016. R package version 2.4.1.

[101] M. Kuhn. Contributions from J. Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, the R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, C. Candan, and T. Hunt. *caret: Classification and Regression Training*, 2016. R package version 6.0-73.

[102] J. Ehrlinger. *ggRandomForests: Visually Exploring Random Forests*, 2016. R package version 2.0.1.

[103] X. Zhou, D. K. McClish, and N. A. Obuchowski. *Statistical methods in diagnostic medicine*. Wiley series in probability and statistics. Wiley, Hoboken, N.J., 2nd edition, 2011.

[104] L. A. Jeni, J. F. Cohn, and F. De La Torre. Facing imbalanced data recommendations for the use of performance metrics. *Int Conf Affect Comput Intell Interact Workshops*, pages 245–251, 2013.

[105] F. Harrell. *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer, New York, NY, 2nd edition, 2015.

[106] R. Amarasingham, R. E. Patzer, M. Huesch, N. Q. Nguyen, and B. Xie. Implementing electronic health care predictive analytics: considerations and challenges. *Health Aff (Millwood)*, 33(7):1148–54, 2014.

[107] K. A. Thursky and M. Mahemoff. User-centered design techniques for a computerised antibiotic decision support system in an intensive care unit. *Int J Med Inform*, 76(10):760–8, 2007.

[108] F. Ventura, I. Koinberg, R. Sawatzky, P. Karlsson, and J. Öhlén. Exploring the person-centeredness of an innovative e-supportive system aimed at person-centered care: prototype evaluation of the care expert. *Comput Inform Nurs*, 34(5):231–239, 2016.

[109] K. Bodker, F. Kensing, and J. Simonsen. *Participatory IT Design: Designing for Business and Workplace Realities*. MIT Press, Cambridge, MA, 2004.

[110] Agency for Healthcare Research and Quality. Training videos. http://www.ahrq.gov/teamstepps/rrs/videos/index.html, December 2012.

[111] Miami University of Ohio. Usability testing: developing useful and usable products. http://www.units.miamioh.edu/mtsc/usabilitytestingrevisedFINAL.pdf, 2004.

[112] M. B. Miles, A. M. Huberman, and J. Saldana. *Qualitative Daya Analysis: A Methods Sourcebook*. SAGE Publications, Inc., Los Angeles, CA, 3rd edition, 2014.

[113] D. Chiffi and R. Zanotti. Medical and nursing diagnoses: a critical comparison. *J Eval Clin Pract*, 21(1):1–6, 2015.

[114] V. E. Sousa, K. D. Lopez, A. Febretti, J. Stifter, Y. Yao, A. Johnson, D. J. Wilkie, and G. M. Keenan. Use of simulation to study nurses' acceptance and nonacceptance of clinical decision support suggestions. *Comput Inform Nurs*, 33(10):465–72, 2015.

[115] M. Müller-Staub, H. de Graaf-Waar, and W. Paans. An internationally consented standard for nursing process-clinical decision support systems in electronic health records. *Comput Inform Nurs*, 2016.

[116] A. Collins, D. Joseph, and K. Bielaczyc. Design research: theoretical and methodological issues. *J Learn Sci*, 13(1):15–42, 2004.

[117] H. Singh and D. F. Sittig. Measuring and improving patient safety through health information technology: the health IT safety framework. *BMJ Qual Saf*, 25(4):226–32, 2016.