Essays on Risk Indicators and Assessment: Theoretical, Empirical and Engineering Approaches

Nasibeh Azadeh-Fard

Dissertation submitted to the faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Industrial and Systems Engineering

Jaime A. Camelio, Chair

Navid Ghaffarzadegan, Co-Chair

Brian M. Kleiner

Konstantinos P. Triantis

December 8, 2015

Blacksburg, Virginia

Keywords: Risk Assessment, Risk Indicators, Early Warning Systems, Healthcare, Preventability, Predictability

Essays on Risk Indicators and Assessment: Theoretical, Empirical and Engineering Approaches

Nasibeh Azadeh-Fard

ABSTRACT

Risk indicators are metrics that are widely used in risk management to indicate how risky an activity is. Among different types of risk indicators, early warning systems are designed to help decision makers predict and be prepared for catastrophic events. Especially, in complex systems where outcomes are often difficult to predict, early warnings can help decision makers manage possible risks and take a proactive approach. Early prediction of catastrophic events and outcomes are at the heart of risk management, and help decision makers take appropriate actions in order to mitigate possible effects of such events. For example, physicians would like to prevent any adverse events for their patients and like to use all pieces of information that help accurate early diagnosis and interventions.

In this research, first we study risk assessment for occupational injuries using accident severity grade as an early warning indicator. We develop a new severity scoring system which considers multiple injury severity factors, and can be used as a part of a novel three-dimensional risk assessment matrix which includes an incident's severity, frequency, and preventability. Then we study the predictability of health outcome based on early risk indicators. A systems model of patient health outcomes and hospital length of stay is presented based on initial health risk and physician assessment of risk. The model elaborates on the interdependent effects of hospital service and a physician's subjective risk assessment on length of stay and mortality. Finally, we extend our research to study

the predictive power of early warning systems and prognostic risk indicators in predicting different outcomes in health such as mortality, disease diagnosis, adverse outcomes, care intensity, and survival. This study provides a theoretical framework on why risk indicators can or cannot predict healthcare outcomes, and how better predictors can be designed. Overall, these three essays shed light on complexities of risk assessments, especially in health domain, and in the contexts where individuals continuously observe and react to the risk indicators. Furthermore, our multi-method research approach provides new insights into improving the design and use of the risk measures.

Dedication

I dedicate this dissertation to:

My father and my mother, Rajab Azadeh Fard and Faezeh Alizad Ashrafi, who always emphasized the importance of higher education, and have made sacrifices to ensure that I achieve my goals.

My husband, Ehsan Rashedi, who has been a great supporter for the past 12 amazing years, and has given me the inspiration to set high goals and the confidence to pursue them.

My sister, Nafiseh Azadeh Fard, who has encouraged and supported me with her unconditional love, and has been my emotional anchor in the life.

Acknowledgements

This dissertation is the result of many experiences that I have encountered at Virginia Tech from individuals who I wish to acknowledge. First and foremost, I would like to sincerely and gratefully thank my advisors, Dr. Jaime Camelio and Dr. Navid Ghaffarzadegan, for their guidance, support, motivation, and advice during my graduate studies at Virginia Tech. Dr. Camelio has been always supportive and encouraged me toward my goals. He gave me the freedom to explore new ideas and pursue my research interests, and he has provided insightful discussions about my research. I am also very grateful to Dr. Ghaffarzadegan for all the help and academic advice through every step of my research, for our weekly research meetings and conversations, and for reviewing my work. For me he is truly an example of a passionate advisor who enthusiastically cares about his student's research. They both set an example of a great advisor and mentor that I would like to follow.

I would also like to thank my other committee members, Dr. Brian Kleiner and Dr. Konstantinos Triantis, for serving on my committee and for their valuable comments and suggestions on my research. This work would not have been accomplished without their valuable inputs. Dr. Triantis's class on performance and productivity measurement was a great opportunity and has opened new research ideas for me.

During my PhD, I had the privilege of working with different faculty members in Grado Department of Industrial and Systems Engineering. I want to thank Dr. Ebru Bish, Dr. Doug Bish, and Dr. Ran Jin. Although for a short time but I really enjoyed working on research projects with them. I am also thankful to Dr. Maury Nussbaum, Dr. Kimberly Ellis, Dr. Barbara Fraticelli, Dr. Michael Taaffe, and Dr. William Woodall for what I have

learned in their classes. I thank Dr. Natalie Cherbaka for her kindness, help, and my wonderful GTA experience with her. My special thanks to Dr. Eileen Van Aken for giving me the opportunity to teach the *Discrete Event Simulation* class. Thank you to Dr. Joel Nachlas, with whom I worked with as faculty advisor of INFORMS student chapter for two years during my volunteer work as chapter's president and treasurer. I am also appreciative to the staff in ISE department, Hannah Parks, Rhonda Hawley, Angie Parvin, Akiko Nakata, Sandy Dalton, Dot Cupp, Nicole Conner, and Will Vest for their kindness and always being available to help.

I would also like to thank the staff in Lewis Gale Medical Center in Salem, VA who gave me the opportunity to collect the data that I needed for my research.

Many thanks to all my friends and colleagues in ISE department for the great moments and wonderful friendship. Thanks to Fatemeh Hashemi, Nevin Mutlu, Hadi El-Amine, Maryam Andalib, Anna Schuh, Hussein Tarhini, Esra Agca, Prasanna Ragavan, Alba Rojas Cordova, Babak Badri Koohi, Alireza Sedighi, Arash Baghaei Lakeh, Daniel Steeneck, Wenmeng Tian, Devita McCullough, Corinna Zhang, Mohammed Shafae, Ahmed El-Habashy, Greg Purdy, Christina Knopf and many others. A special thanks goes to my office mate, Niyousha Hosseinichimeh, who was a true and kind friend and has been supportive in every way.

Throughout the experience and my stay here in Blacksburg, I was nurtured by love and friendship of many Iranian friends. Thanks to Maryam Kamali, Hoda Koushyar, Atieh Vahidmanesh, Reyhaneh Ojani, Sadeh Jafari, Leyla Nazhandali, Azadeh Aryan, Fatemeh Zarghami, Hasti Ardekani, Atieh Haghdoust, Maliheh Borhani, Sahar Sadeghi, Zahra

Mashhadi, Somayeh Badieyan, Baharak Javidi, Fatemeh Yazdandoust, and their families for all the wonderful gatherings, trips, and memories.

I would like to express my special thanks to my family for their continuous support and encouragement. First and foremost, my sincere and heartiest thanks go to my love, Ehsan Rashedi, who is my best friend, soul-mate, wonderful colleague, and amazing husband. Ehsan has been a true and great supporter for the past 12 years, and always encouraged me to continue. He is the one who gave me the inspiration and the confidence to pursue PhD. Without his constant support this would have not been accomplished. I also thank his family; my father and mother-in-law, my sister-in-law, brother-in-law, and their families for their kind support.

I am also very thankful to my sister, Nafiseh Azadeh Fard, and brother-in-law, Hamid Sadeghi, for always being there for my parents during these years that I was far away from them. Without their help this could have not been possible. My sister has been my best friend all my life and I thank her for all her encouragement and kind support.

Finally, I want to thank my parents, Rajab Azadeh Fard and Faezeh Alizad Ashrafi, for their endless sacrifice, unconditional love, support, and encouragement. I appreciate all they taught me on the values of ethical behavior, helping people, and hard work. They always emphasized on the values of higher education, and supported me and Ehsan unconditionally to ensure that we achieve this goal. I would not have made it this far without them.

Table of Contents

List of Figures

List of Tables

1. Introduction

Design and utilization of appropriate risk metrics are crucial in accurate risk assessment as well as taking actions to mitigate potential undesired consequences. Risk assessment is "a systematic process for describing and quantifying the risks associated with hazardous substances, processes, actions, or events" [1]. Accurate risk assessments help to improve awareness of hazards and risks, and determine if existing control measures are adequate or if more should be done. Furthermore, risk assessment can prevent catastrophic events when it is performed at the design or planning stage. Risk assessment methods are also used to prioritize hazards and control measures, which help better resource allocation and crisis management.

There are various examples of utilizing risk assessment techniques in different industries. For example, risk assessment has been used frequently to estimate the probabilities of possible consequences of accidents at a nuclear power plant [2]. Or in food industry, risk assessment can be used to quantify and describe the risks involved with diseases that have been linked to consumption of poultry products [3]. More examples of risk assessment techniques exist in healthcare and occupational safety. These techniques have been used to determine and assess the risk of specific diseases, e.g., cardiovascular disease, for different patient groups, risk assessment methods have been used [4], and in occupational health and safety to prevent injuries that may happen in the workplace [5, 6].

A wide range of methods exists for design and implementation of risk metrics, Different organizations may use different indicators. Subjective risk indicators, i.e., person's perception of the likelihood of a risky event, are also common and extensively used in different industries to

quantify the risk of hazard. However, unlike the objective measures, all subjective metrics are prone to individual biases and mistakes since one cannot precisely calculate the actual probability.

In this dissertation, we tackle the big question of *how risk assessment techniques can be improved in complex healthcare and occupational contexts.* In other words, our research objective is to investigate the power of risk indicators in predicting the actual outcome, and to help improve the existing risk assessment processes. We narrow down this big objective into three more specific complementary studies and 1) offer, a new objective risk indicator for evaluating the risk of occupational injuries, 2) assess a common risk indicator in healthcare systems using a novel dataset that we gathered from a large hospital, and 3) examine the predictive power of early warning systems under different conditions and for different purposes. Our focus is on human health using data from an occupational safety context and health service domain. More information about these three studies are provided in the following.

1.1 Risk Indicators of Occupational Injuries

Occupational accidents are responsible for many fatal and nonfatal injuries in the United States. The Bureau of Labor Statistics reports that about 3 million nonfatal workplace injuries and illnesses occurred in 2012 [7], while fatal injuries in the same period were reported to have a rate of 3.2 cases out of every 100,000 full-time workers [8]. The National Institute for Occupational Safety and Health (NIOSH) states that one of its main missions is to reduce work-related injuries, illnesses, and death through periodic surveillance [9], in an effort to increase the utility of information gathered from different stakeholders regarding injuries and hazards in the workplace.

The relative severity of an accident is a common surveillance metric that is used to determine the magnitude of an incident [10]. An accident's severity is often defined based on the number of lost or restricted workdays resulting from the incident [10, 11]. However, the definition of severity, e.g., OSHA's Severity Rate formula, used by current risk management tools does not consider important employee and workplace factors, such as age, gender, and weather with potential significant impacts on accident severity.

The need to formally monitor an accident severity metric regarding to possible worker and workplace characteristics has been brought to the forefront in the literature [12]. Moreover, many authors have studied the effect of different factors on occupational injuries [13-15]. Therefore, introducing a new severity scoring system that considers multiple influential factors will improve the risk assessment of occupational accidents.

1.2 Risk Indicators in Healthcare

Different methods have been developed to identify the areas of actual or potential risk in patients, and to reduce the incidence of adverse events and their consequences in hospitals. Ideally, we would like to predict health risks early during hospital admission. An early warning system is a measurement tool to assess patients' health risk objectively and quickly determine the degree of illness. They aim at reducing the risk of sudden, life-threatening events for patients with the help of hospital rapid response teams [16]. These systems help nurses or patient family members call for a designated group of healthcare professionals to a patient's bedside to react immediately to a deteriorating condition.

The main use of these measures is to provide early warnings to health providers to spur quick preventive reactions. Some have argued that these measures have much more to offer, such as

helping to predict patients' length of stay (LOS) in hospitals or health outcomes such as the chance of in-hospital mortality [17-19]. Given the importance of LOS and mortality in assessing healthcare quality, resource allocations, and costs [20], the hope is that early warning scores can make it possible to predict and improve hospital utilization and outcomes.

1.3 Predictive Power of Early Warning Systems in Healthcare

Early warnings in healthcare are aimed at providing timely predictions of high consequent and catastrophic events [17, 21]. They are often used to prevent catastrophic events such as intensive care unit (ICU) admission and in-hospital mortality [22]. Especially, in complex systems where outcomes are often difficult to predict, early warnings can help decision makers manage possible risks and take a proactive approach. All decision makers would like to receive early and trustable warnings, however, the question is that with the growing complexities and uncertainties in healthcare domain, how useful these early warning systems are, or under which conditions they are better predictors of outcomes? The concern is whether early warning systems and prognostic indicators are sensitive and specific enough to predict health outcome.

1.4 Research Approach

Pursuing our main research question needs a portfolio of approaches which encompass theory building, empirical analysis, and design. This objective was followed in three complimentary essays, as following: First, given the importance of subjective risk assessments, the risk assessment of occupational injuries using accident severity grade as an early warning indicator was studied, and a framework to translate subjective assessments to a more objective analysis was

provided. These techniques involved risk assessment to identify potential hazards and the expected severity of injuries that may result from these hazards, usually based on the severity of similar past injuries. A new severity scoring system was introduced which considers multiple injury severity factors, and was used as part of a novel three-dimensional risk assessment matrix which includes an incident's severity, frequency, and preventability.

In the second study, a systems model of patient health outcomes and LOS was presented based on initial health risk and physician assessment of risk. The model elaborates on the interdependent effects of hospital service and a physician's subjective risk assessment on LOS and mortality. The model was used to offer hypotheses about the predictive power of early warnings that were empirically tested by analyzing a detailed dataset of 1,031 patients admitted to a large hospital in the southeastern United States.

In the third study, the first and second studies were extended to investigate the predictive power of early warning systems in the domain of healthcare. The available literature, published during the past 15 years, was systematically reviewed to assess the predictive power of early warning systems and prognostic risk indicators in predicting different outcomes in health such as mortality, disease diagnosis, adverse outcomes, care intensity, and survival.

Overall, this dissertation contribute to the theoretical and empirical foundations of risk assessment and especially early warning systems, offer a new way to engineer more powerful risk assessment tools, and brings more insights into predictive power of early warning systems in healthcare.

References

1. Covello, V.T. and M.W. Merkhoher, *Risk assessment methods: approaches for assessing health and environmental risks*. 2013: Springer Science & Business Media.

2. Keller, W. and M. Modarres, *A historical overview of probabilistic risk assessment development and its use in the nuclear power industry: a tribute to the late Professor Norman Carl Rasmussen.* Reliability Engineering & System Safety, 2005. **89**(3): p. 271-285.

3. Golden, N.J., et al., *Risk assessment for Clostridium perfringens in ready-to-eat and partially cooked meat and poultry products.* Journal of Food Protection®, 2009. **72**(7): p. 1376-1384.

4. Brindle, P., et al., *Accuracy and impact of risk assessment in the primary prevention of cardiovascular disease: a systematic review.* Heart, 2006. **92**(12): p. 1752-1759.

5. Fung, I.W., et al., *Developing a risk assessment model for construction safety*. International Journal of Project Management, 2010. **28**(6): p. 593-600.

6. Nelson, A., et al., *Development and evaluation of a multifaceted ergonomics program to prevent injuries associated with patient handling tasks.* International journal of nursing studies, 2006. **43**(6): p. 717-733.

7. Asadollahi, K., et al., *Prediction of hospital mortality from admission laboratory data and patient age: A simple model.* Emergency Medicine Australasia, 2011. **23**(3): p. 354-363.

8. Bureau of Labor Statistics. *Census of Fatal Occupational Injuries Summary*. 2012 [cited 2013; Available from: http://www.bls.gov/news.release/cfoi.nr0.htm.

9. Centers for Diesease Control and Prevention. *NIOSH Strategic Goals*. 2013; Available from: http://www.cdc.gov/niosh/programs/surv/goals.html.

10. Safety, O. and H. Administration, *OSHA technical manual: Section VII, chapter 1*. 1999.

11. OSHA, *FORMULAS for CALCULATING RATES, OSHA Recordable Incident Rate, Lost Time Case Rate, Lost Work Day Rate (LWD), DART Rate, Severity Rate*.

12. Schuh, A. and J.A. Camelio. *Including Accident Severity in Statistical Monitoring Systems for Occupational Safety*. in *Industrial and Systems Engineering Conference*. 2013.

13. Kines, P., *Occupational injury risk assessment using injury severity odds ratios: Male falls from heights in the Danish construction industry, 1993-1999.* Human and Ecological Risk Assessment, 2001. **7**(7): p. 1929-1943.

14. Messing, K., et al., *Be the fairest of them all: challenges and recommendations for the treatment of gender in occupational health research.* American Journal of Industrial Medicine, 2003. **43**(6): p. 618-629.

15. DeGroot, D.W., et al., *Epidemiology of US Army Cold Weather Injuries, 19801999.* Aviation, space, and environmental medicine, 2003. **74**(5): p. 564-570.

16. Institute for Healthcare Improvement. *Early Warning System: Scorecards That Save Lives*. Improvement Stories 2014 [cited 2014; Available from: http://www.ihi.org/resources/Pages/ImprovementStories/EarlyWarningSystemsScorecardsThatSaveLives.aspx.

17. Paterson, R., et al., *Prediction of in-hospital mortality and length of stay using an early warning scoring system: clinical audit.* Clinical Medicine, 2006. **6**(3): p. 281-284.

18. Burch, V., G. Tarr, and C. Morroni, *Modified early warning score predicts the need for hospital admission and inhospital mortality.* Emergency Medicine Journal, 2008. **25**(10): p. 674-678.

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

20. Brownell, M. and N. Roos, *Variation in length of stay as a measure of efficiency in Manitoba hospitals.* CMAJ: Canadian Medical Association Journal, 1995. **152**(5): p. 675.

21. Smith, G.B., et al., *The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death.* Resuscitation, 2013. **84**(4): p. 465-470.

22. McGaughey, J., et al., *Outreach and Early Warning Systems (EWS) for the prevention of intensive care admission and death of critically ill adult patients on general hospital wards.* Cochrane Database Syst Rev, 2007. **3**.

2. Risk Assessment of Occupational Injuries Using Accident Severity Grade[1]

2.1 Abstract

Problem: In spite of recent efforts to improve occupational health and safety, many occupational accidents result in serious injury and death every year. Continued efforts are therefore necessary to improve current safety initiatives and reduce the frequency and severity of these incidents. To identify workplace hazards, many safety surveillance techniques have been used, including severity metrics to determine the significance of an accident. These techniques involve risk assessment to identify potential hazards and the expected severity of injuries which may result from these hazards, usually based on the severity of similar past injuries. However, these severity metrics do not consider important employee and workplace risk factors, such as age, gender, and weather, which may have significant impacts on accident severity. *Method:* A new severity scoring system is introduced which considers multiple injury severity factors, and is used as part of a novel three-dimensional risk assessment matrix which includes an incident's severity, frequency, and preventability. A case study using the proposed methodology with real data is presented. *Discussion:* The consideration of additional severity factors improves risk assessment and the estimation of injury severity. A three dimensional risk assessment matrix allows for the analysis of an incident's degree of preventability, frequency, and severity all at once. *Practical Applications:* This study demonstrates that organizations, industries, and regulatory bodies can improve workplace safety surveillance tools by incorporating this new severity metric in a three-dimensional risk assessment matrix.

---

1 The following manuscript by Nasibeh Azadeh-Fard, Anna Schuh, Ehsan Rashedi, and Jaime Camelio, has been published in Safety Science journal in July 2015.

Keywords: Injury severity, Occupational safety, Preventability, Risk assessment, Surveillance.

2.2 Introduction

Occupational accidents are responsible for many fatal and nonfatal injuries in the United States. The Bureau of Labor Statistics reports that about 3 million nonfatal workplace injuries and illnesses occurred in 2012 [1], while fatal injuries in the same period were reported to have a rate of 3.2 cases out of every 100,000 full-time workers [2]. The National Institute for Occupational Safety and Health (NIOSH) states that one of its main missions is to reduce work-related injuries, illnesses, and death through periodic surveillance [3], in an effort to increase the utility of information gathered from different stakeholders regarding injuries and hazards in the workplace.

2.2.1 Severity as a Surveillance Metric

The relative severity of an accident is a common surveillance metric that is used to determine the magnitude of an incident [4]. An accident's severity is often defined based on the number of lost or restricted workdays resulting from the incident [4, 5]. According to the dictionary of scientific and technical terms, an accident severity rate is defined as "the number of worker days lost due to a disabling accident per thousand worker-hours of exposure" [6]. Similarly, the Occupational Safety and Health Administration (OSHA) defines the Severity Rate (SR) of an incident as [4]:

$$SR = \frac{(Total\ number\ of\ lost\ or\ restricted\ workdays\ in\ the\ past\ 12\ months)*200{,}000}{Number\ of\ work\ hours\ during\ the\ past\ 12\ months} \tag{2.1}$$

The numerator in Eq. (2.1) represents the number of lost or restricted workdays in specific department in a 12-month period, multiplied by 200,000 to normalize the number of observed workers to a standard form of 100 employees working 50 weeks per year. SR is a generic metric that can be used in different industries and work environments to quantify injury severity.

Prevention and control of occupational injuries require information about the leading causes of incidents or risk factors. Literature has shown the causal role of work and environmental conditions in the occurrence of occupational accidents [7]. Further, employee factors such as age and gender may have effect on accident occurrence [8]. Notably, any correlation between the SR and employee factors such as age and gender is not considered in Eq. (2.1). Other commonly used severity metrics such as odds ratios [9] and the Injury Severity Score [10] do not include these factors either. Including the impact of these predictors as part of a regular injury surveillance can eventually help safety managers identify potential safety hazards before they lead to severe injuries [11]. For example, the most common finding in the literature associated with age is that accident severity tends to increase with age. In terms of gender, comparisons between male and female workers suggest that men tend to have a comparable average days away from work per injury to women, but higher rates of permanent disabilities and fatalities were observed [8]. Furthermore, it has been noted that the rate of serious injuries increases with age, however, the total number of injuries decreases for older ages [9, 12-14]; in recognizing this, it is clear that instituting age-specific injury prevention interventions may help reduce both the rates of serious injuries in older employees and the number of serious injuries in younger employees. Likewise, gender-specific practices and data collection has been shown to improve safety processes in workplaces [15], which may result in less injuries and fatalities.

In addition, it may be desirable to consider workplace factors such as the weather, job location, and the condition of work environment. For instance, cold weather has been shown to increase injury severity in the US Army [16]. Moreover, climate may also affect the slip and fall injuries among construction workers, and rainy weather has a significant effect on workplace fatalities in

this industry [17, 18]. Hot weather during summer can increase the risk of electrocution in various industries [19].

The location of the worksite is another important factor that can affect an employee's injury risk in various jobs such as underground mining [20]. More severe injuries might be experienced if an emergency team cannot help an injured person in a timely manner due to limitations of the location or accessibility issues.

The condition of the work environment may also be of interest. Several studies have investigated the effect of surface conditions on occupational accidents [21-23]. Slippery surfaces often result in falls, which cause workplace injuries among construction and mining workers. Thus, the effect of surface condition on occupational accidents should be studied as a potential risk factor, especially in these industries.

Moreover, occupational tasks can influence an employee's risk for workplace injuries. For instance, several studies have found a positive correlation between repetitive motion tasks and occupational injuries such as musculoskeletal disorders [24, 25].

The condition of equipment is another factor that can affect the risk of occupational injuries. One study showed that the risk of hand injuries was increased by using tools and equipment that are not working properly [26].

Therefore, identifying the potential risk factors similar to the ones that are discussed in this section can improve the risk assessment processes. Clearly, many factors can influence the severity of an occupational injury and should therefore be considered in injury surveillance metrics. The new surveillance metric is flexible and can be adjusted for different industries, accounting for unique sets of risk factors.

2.2.2 Severity in Risk Assessment

Once collected, all of this employee and workplace data can be used for health risk assessment purposes and surveillance. Several scoring systems have been developed to assign relative impact scores to healthcare incidents for the purpose of improved surveillance. Patient scores in healthcare, such as the Parsonnet score, Cleveland Clinic score, French score, Euro score, and Ontario Province risk scores, were introduced to monitor medical outcomes (e.g., mortality rate) following surgery [27-30]. Studies have shown that the development of these distinct scoring systems can enhance the quality of patient care [31, 32]. However, there are very few risk assessment scoring systems for occupational injuries. The fishing industry and the Alaska Marine Safety Education Association have generated a risk assessment score sheet to reduce the accidents in fishing vessels [33]. Numerical risk values can be assigned to influential factors based on the assessment criteria for each factor. Darby et.al. [34], used five different fleet driver assessment scores (i.e., exposure to risk score, attitude to safe driving score, behavioral score, knowledge of the rules of the road score, and hazard perception score) to identify, target, and reduce occupational road safety risks. These methods identify high risk factors so that injury prevention strategies can be prioritized.

2.2.3 Aim

Occupational injury risk scores that include employee and workplace risk predictors are necessary for the improvement of current risk assessment tools. Because the current methods for determining accident severity do not include several important factors such as age, weather, and gender, a new severity scoring system is introduced which will be incorporated into a risk assessment tool. This scoring system can be utilized by a variety of industries to quantify injury severity since the method

is generic, and can be modified according to the specific risk factors in the work environment. The proposed severity scores will then be used to create an occupational injury risk scoring system that includes an accident's severity, frequency, and preventability. A three-dimensional risk assessment matrix will be used to analyze these factors.

The remainder of this paper is structured as follows. In Section 2.3, a severity scoring method is developed. Section 2.4 discusses the traditional two-dimensional risk assessment matrix and introduces a new three-dimensional risk assessment matrix. A case study using mining occupational injuries is presented in Section 2.5. Results of the severity scoring method and three-dimensional risk assessment matrix are discussed in this section, followed by practical applications in Section 2.6. Conclusions and future work are discussed in Sections 2.7 and 2.8.

2.3 Accident Severity Grade (ASG)

In this section, a method is developed to quantify a new occupational injury risk score, the Accident Severity Grade (ASG), based on employee and workplace risk factors.

2.3.1 Risk Factors Associated With Accident Severity

There are several factors that will influence the severity of an accident. Table 2.1 represents common risk factors as well as the scoring method that can be used to better quantify the severity. The ASG score sheet in Table 2.1 can be completed for different incidents in order to predict the injury severity.

2.3.2 Approach

It should be noted that the weights of the factors may vary among different occupations and working environments. Thus, it is essential to address any dataset-specific considerations by assigning appropriate weights for each factor. The weights can be determined based on the average number of lost work days from a similar historical dataset. For example, if advanced age is historically correlated with a high number of days away from work, then it should be assigned the highest weight when compared with other factors. Let m_i be the average number of days away from work for factor *i*. Then, the weight of each factor, w_i, is calculated according to Eq. (2.2)

$$w_i = \frac{m_i}{\sum_{i=1}^{n} m_i} \tag{2.2}$$

where $\sum_{i=1}^{n} w_i = 1$ and *n* represents the total number of factors considered.

The level of impact, l_i, for each factor should be assigned as either 0, 0.5, 1, 1.5, or 2, with 0 having the least impact and 2 having the highest impact. Similar scales have been used in other scoring systems such as the Parsonnet score and the fishing risk assessment score sheet [27, 33]. The impact level is assigned empirically based by safety practitioners after observing the incident.

Table 2.1 Accident Severity Grade (ASG) Score Sheet

Factor (*i*)	**Assessment Criteria**	**Level of Impact (l_i)**	**Weight (w_i)**	**Weighted Score ($l_i w_i$)**
Age	(a) <20 years (b) 20-59 years (c) >59 years			
Gender	(a) Female (b) Male			
Weather	(a) Rainy (b) Snowy / Icy			

	(c) Windy (d) Hot	
Equipment	(a) Heavy equipment (b) Old machine (c) Have proper safety equipment? (d) Examined during last year?	
Surface Condition	(a) Slippery (b) Hard/Soft (c) Surface fall	
Repetitive Movement	(a) High frequency (b) Low frequency	
Location	(a) Underground (b) Ground level (c) Above ground (d) Location of equipment	
Training	(a) No training (b) Classrooms and Hands-on training (c) Previous training experience (d) Workplace training	
		ASG

Factors that do not contribute to the accident will have an impact score of $l_i = 0$. In addition, for every incident with an impact score of $l_i = 1$, the impact of factor *i* will be equal to the weight of that factor (i.e., w_i), which is equivalent to the average number of workdays away from work, based on the previously reported accidents. The ASG is calculated as follows:

$$ASG = \sum_{i=1}^{n} w_i l_i \ , \ 0 \leq l_i \leq 2 \tag{2.3}$$

The ASG needs to be categorized for different levels of severity that are shown in Table 2.2 (four categories are used in this example). If the ASG falls between $\left(\frac{c_j}{\sum_{i=1}^{n} m_i}, \frac{c_{j+1}}{\sum_{i=1}^{n} m_i}\right), j = 0, \ldots, 4$, then

it belongs to severity category of $j+1$. We define c_j as the maximum number of acceptable days away from work for category j, which can be determined based on the probability of a case having more lost days than a given magnitude (an example case study is presented in Section 2.5).

It should be noted that individual industries or companies may use different factors to quantify the severity of an incident in their specific work environments. The predictors presented in Table 2.1 are some of the most common factors that affect occupational injuries in the workplace, but if an organization would prefer to consider other factors, the proposed ASG method can be adjusted accordingly. This flexibility enables industries with dissimilar predictors to specialize the ASG scoring method as their individual data collection allows. The case study in Section 2.5 demonstrates more details about modifying the ASG scoring technique.

2.4 Risk Assessment

2.4.1 Two-Dimensional Risk Assessment Matrix

Risk assessment has been traditionally involved in quantifying the risk of an incident based on two or more aspects, such as the likelihood of a risk (frequency) and the impact or consequence of the risk occurring (severity). In the context of risk assessment matrix, the risk of an activity represents the amount of injury that is expected to occur as a result of a potential accident associated with an activity. It can be estimated by multiplying the severity level and the occurrence probability of an accident. Two-dimensional risk assessment matrices have been widely used to define different levels of risk [35]. Table 2.2 shows an example of a commonly used two-dimensional risk assessment matrix which considers an accident's severity and frequency. The color codes in this matrix correspond to various levels of risk (low, medium and high). This matrix has been criticized

for being subjective and qualitative [36]. Assigning numerical values for severity and frequency can provide quantitative risk values for the matrix.

Table 2.2 An example of two-dimensional risk assessment matrix

		Frequency				
		Frequent	Probable	Occasional	Remote	Improbable
Severity	Negligible	Medium	Medium	Low	Low	Low
	Marginal	High	Medium	Medium	Medium	Low
	Critical	High	High	Medium	Medium	Low
	Highly Severe	High	High	High	Medium	Medium

2.4.2 Adding a New Dimension: Preventability

Inherent risk is defined as "the probability of loss arising out of circumstances or existing in an environment, in the absence of any action to control or modify the circumstances" [37]. In other words, inherent risk indicates the gross risk or risk of an accident before any controlling or preventing action, i.e., Inherent Risk = Severity × Frequency, which can be determined using a risk assessment matrix similar to the one in Table 2.2. However, the risk of an injury or death that remains after the elimination of known risks can provide a better estimation of the need for further corrective actions in the workplace. Eq. (2.4) shows the calculation of this residual risk, which is very common in economics and information security risk assessment [38-40]. Therefore, although frequency and severity are the primary characteristics used to monitor risks, the degree of residual risk (also known as the preventability of a risky event) can have a significant impact on risk estimation and further management safety decisions.

$$\textit{Residual Risk} = \textit{Inherent Risk} \times \textit{Preventability} \tag{2.4}$$

Preventability has been discussed frequently in the healthcare literature. Gurwitz et.al. [41] studied the preventability of adverse drug events among older persons in the ambulatory setting. Other

studies concluded that adverse drug events among hospitalized adults in nursing homes are often preventable [42, 43]. Likewise, pre-hospital trauma deaths were found to be preventable [44]. Most occupational injuries are considered to be preventable using injury-prevention strategies [45]. Incorporating preventability as a new risk assessment dimension can enhance the risk estimation, which can in turn lead to more informed safety decision making.

The next section presents a novel risk assessment matrix which has been developed to consider a particular injury's frequency, severity, and preventability in order to quantify the degree of residual risk.

2.4.3 Categorizing Severity, Frequency, and Preventability

Severity levels have been defined by several industries for safety management purposes. For example, NASA's hazard analysis program defines three levels of severity as catastrophic, critical, or marginal depending on the loss/damage to the ground facility or flight vehicle elements [35, 46]. Utilizing a similar approach [17, 23, 24], we use the following definitions for categorizing severity and assign the quantified values of each category in Section 2.4:

a) Highly Severe: an extremely harmful accident that could result in fatality, permanent total disability, or a greater than 30 number of lost workdays,
b) Critical: an accident which could result in serious injury such as permanent partial disability and 10 to 30 number of lost workdays,
c) Marginal: an accident that could cause minor injury with 5 to 10 number of lost workdays,
d) Negligible: an accident or injury that has none to 5 lost workdays or could be treated by applying first-aids.

Quantified values of severity for each category will be calculated based on the new ASG severity scoring method introduced in Section 2.3.

In order to categorize frequency, the probabilities of accident occurrence and quantitative values are defined in five categories (Table 2.3) [47].

Preventability is considered in the proposed risk stratification model by categorizing an accident into three levels of preventability: highly preventable, moderately preventable, and unlikely to be preventable, with their respective numerical values to be 0.5, 1, and 1.5, respectively. These values can be assigned by safety practitioners to incorporate the preventability of an incident. According to Eq. (2.4), a highly preventable accident would decrease the predicted risk of injury by 50% while for the least preventable accident, the predicted risk would increase by 50%. Assessing the preventability of an accident can be based on the possible alterations in work environment. For instance, whether slight changes in work place such as drying the floor can prevent the occurrence of slip and fall, we can assume this accident to be highly preventable. The National Safety Council provides a list of the top 10 preventable workplace incidents and the controlling actions that can be done to prevent these accidents [48]. We use this list (Table 2.4) to recognize and categorize the highly and moderately preventable incidents. Any type of incident that is not classified in Table 2.4 is assumed to be unlikely to be preventable.

To represent the three-dimensional risk assessment matrix, we use three two-dimensional matrices (Tables 2.7, 2.8, and 2.9) for each possible level of preventability (highly, moderately, and unlikely to be preventable). By calculating residual risk values according to Eq. (2.4), the risk of an accident can be assessed using this three-dimensional risk assessment matrix. A case study using this three-dimensional risk assessment matrix with residual risk scores and color codes is presented in the next section.

Table 2.3 Frequency descriptors

Frequency	Chance	Probability
Frequent	very likely to occur in the time period of an event	>95%
Probable	likely to occur in the time period of an event	>65%
Occasional	occurs sometimes in the time period of an event	>35%
Remote	unlikely but still possible to happen in the time period of an event	<35%
Improbable	unlikely and it can be assumed that the accident will not occur in the time period of an event.	<5%

Table 2.4 Top preventable workplace incidents

Preventability	Incident	Action
Highly preventable	**On the job violent acts:** attacks caused by office politics and other personal arguments have led to serious physical injuries.	Provide violence training for employees.
Moderately preventable	**Repetitive motion injuries:** such as typing and excessive use of computer can strain muscles and tendons.	Provide proper ergonomic equipment and training.
Moderately preventable	**Machine entanglement:** loose clothing, shoes, jewelry, fingers, and unbound hair may become caught in machinery.	Provide protective barriers/equipment and train employees.
Moderately preventable	**Vehicle crashes:** employees who drive for business purposes are often injured in auto crashes.	Define safe driving policies and provide safe-driver training.

Highly preventable	**Walking into injuries:** when a person unintentionally runs into static objects such as walls, doors, cabinets, etc.	Maintain a neat workplace, clearly mark potential obstacles/hazards.
Highly preventable	**Falling object injuries:** objects that fall from shelves or are dropped by another person.	Store/stack materials in a safe and secure manner, use signage and protective equipment.
Moderately preventable	**Reaction injuries:** caused by slipping and tripping without falling and can cause muscle injuries, and body trauma.	Address slippery areas around the facility by clearing snow, and placing no slip rugs near entrances/exits.
Moderately preventable	**Falling from heights:** falls that happen from an elevated area such as roofs, ladders, and stairways.	Use of proper personal protection equipment, installation of guard rails, training.
Moderately preventable	**Slipping/Tripping:** slipping on wet floors or tripping over a foreign objector uneven surface.	Use of non-slip rugs in potentially slippery areas, use signage to indicate slippery areas, and training on keeping the floors clean.
Highly preventable	**Overexertion injuries:** pulling, lifting, pushing, and holding activities at work.	Train employees on the proper ways to perform physical activity, use of proper equipment.

2.5 Case Study - Severity of Injuries in the Mining Industry

In this section, calculations for the proposed ASG scoring methodology are presented using a historical dataset of mining accidents. Moreover, the three-dimensional risk assessment matrix will be discussed regarding to ASG values. The publically available mining accident injuries and

illnesses dataset used in this analysis includes data reported by mine operators and contractors from 1983 to 2013 [49].

2.5.1 ASG Score

Table 2.5 shows the weights of important risk factors in the mining industry that were calculated using Eq. (2.2). Around 100 incidents were randomly selected from the dataset. The number of incidents and the total number of lost workdays were used to calculate the average number of lost workdays as a result of each factor, and the contributing factor was determined by reading text descriptions of the accidents. Table 2.6 shows the modified ASG score sheet for this data. Due to the limited description records in the mining dataset, it is assumed that there was only one contributing risk factor for each accident, so that one factor is given an impact score of 2 while impact score for each of the other factors will be 0.

The following accident description is used as an example: "Employee was loading a dust truck. As he began to walk down the catwalk, the employee's foot caught on the non-stick surface. The employee grabbed the handrail to keep from falling and in the process, twisted his right knee [49]." In this case, "Surface condition" was recorded as the major contributing risk factor. Since the employee twisted his knee due to non-stick surface condition, it can be assumed that surface condition highly impacted the accident outcome. Because this analysis is assigning each factor an impact score between 0 and 2, the level of impact for surface condition is recorded as 2, or the equivalent of two times of the factor weight.

The maximum acceptable number of lost workdays for each severity range is calculated based on the probability of an incident resulting in a given number of lost workdays. For simplicity, we assume four categories for the accidents that result in less than one month of lost workdays and

one category for the accidents that contribute to more than thirty lost workdays. Probabilities of 1.00, 0.80, 0.50, and 0.20 are assigned, which correspond to $c_0 = 0$, $c_1 = 5$, $c_2 = 10$, $c_3 = 30$, and $c_4 \geq 30$ lost workdays, in accordance with NIOSH mining injury statistics for 2000-2004 [50]. In other words, the probability of an injury resulting in more than ten lost workdays is 0.50, while the probability of thirty or more lost workdays is only 0.2. The severity ranges for ASG for the case study dataset are calculated as follows.

- Negligible: $[\frac{c_0}{89.58}, \frac{c_1}{89.58}]$ = [0,0.056]
- Marginal: $[\frac{c_1}{89.58}, \frac{c_2}{89.58}]$ = (0.056, 0.112]
- Critical: $[\frac{c_2}{89.58}, \frac{c_3}{89.58}]$ = (0.112, 0.335]
- Highly Severe: $[\frac{c_3}{89.58}, \frac{c_4}{89.58}]$ = (0.335, ∞)

Considering the ASG from the example in Table 2.6, the value 0.74 falls into the Highly Severe range of the ASG scoring system, since it has a value greater than 0.335. Therefore, this incident would be expected to result in more than thirty lost workdays. This accident did, in fact, result in 155 lost workdays. The accurate estimation of an accident's total number of lost workdays immediately following its occurrence will allow real-time surveillance tools to incorporate this measure of severity, in addition to frequency and preventability, without needing to wait until a certain number of days has elapsed for the accident to be categorized.

Table 2.5 Risk Factors in Mining Dataset

Risk Factor	Number of incidents	Total number of lost workdays	Average number of lost workdays (m_i)	Weight (w_i)
Weather	49	379	$^{379}/_{49} = 7.73$	$^{7.73}/_{89.58} = 0.086$
Training	10	88	$^{88}/_{10} = 8.8$	$^{8.8}/_{89.58} = 0.098$
Equipment	12	223	$^{223}/_{12} = 18.58$	$^{18.58}/_{89.58} = 0.207$
Surface	9	298	$^{298}/_{9} = 33.11$	$^{33.11}/_{89.58} = 0.370$
Location	5	14	$^{14}/_{5} = 2.8$	$^{2.8}/_{89.58} = 0.031$
Repetitive Movement	20	371	$^{371}/_{20} = 18.55$	$^{18.55}/_{89.58} = 0.098$
		Total	**89.58**	

2.5.2 Risk Scores and Risk Assessment Matrix

Residual risk scores for this case study are calculated using the ASG ranges in Section 2.4 and Eq. (2.4). Tables 2.6, 2.7, and 2.8 represent the residual risk scores and color codes for mining case study. Light grey cells show the low risk levels, while grey cells correspond to medium levels of risk, and dark grey cells represent the high levels of risk.

Table 2.6 ASG Score Sheet for Mining Dataset

Factor (i)	**Assessment Criteria**	**Level of Impact (l_i)**	**Weight (w_i)**	**Weighted Score ($l_i w_i$)**
Weather	(a) Rainy (b) Snowy / Icy (c) Windy (d) Hot	0	0.086	0
Equipment	(a) Heavy (b) Old Machine (c) Have proper safety equipment (d) Examined during last year	0	0.207	0
Surface Condition	(a) Slippery (b) Hard/Soft (c) Surface Fall	2	0.370	0.74
Repetitive Movement	(a) High frequency (b) Low frequency	0	0.207	0
Location	(a) Underground (b) Ground level (c) Above ground (d) Location of equipment	0	0.031	0
Training	(a) Classrooms and Hands-on training (b) Previous training experience (c) Workplace training	0	0.098	0
			ASG	**0.74**

The following ranges for residual risk of an injury in this case study are proposed.

- Low risk (light grey cells): [0, 0.036]
- Medium risk (grey cells): (0.036, 0.2]
- High risk (dark grey cells): (0.2, ∞)

These ranges are based on the current case study's residual risk scores and the severity, frequency, and preventability of the incidents in the dataset. For example, it is assumed that only highly severe

and frequent accidents contain high level of residual risk (see Table 2.7). For a high ASG, i.e., (0.335, ∞), if the risk score is greater than 0.2, the risk level is designated as high. However, if the residual risk score is less than 0.2, the risk level is categorized as medium or low. For the highly severe accident (i.e., ASG can be any number greater than 0.335), when the lower bound of risk score is less than 0.2 (i.e., not high risk), the upper bound is assumed to be 0.2.

Table 2.7 Risk Assessment Matrix for Highly Preventable Incidents

		Frequency				
		Frequent	Probable	Occasional	Remote	Improbable
Severity	Negligible	(0, 0.028]	(0, 0.0266]	(0, 0.018]	(0, 0.009]	[0, 0.001]
	Marginal	(0.028, 0.06]	(0.018, 0.053]	(0.009, 0.036]	(0.001, 0.019]	[0, 0.003]
	Critical	(0.06, 0.17]	(0.036, 0.16]	(0.019, 0.109]	(0.003, 0.06]	[0, 0.008]
	Highly Severe	(0.17, ∞)	(0.11, 0.2]	(0.059, 0.2]	(0.008, 0.2]	[0, 0.2]

Table 2.8 Risk Assessment Matrix for Moderately Preventable Incidents

		Frequency				
		Frequent	Probable	Occasional	Remote	Improbable
Severity	Negligible	(0, 0.056]	(0, 0.053]	(0,0.036]	(0, 0.019]	[0, 0.003]
	Marginal	(0.053, 0.112]	(0.036, 0.11]	(0.019, 0.073]	(0.003, 0.039]	(0, 0.006]
	Critical	(0.11, 0.335]	(0.072, 0.318]	(0.039, 0.22]	(0.006, 0.12]	(0, 0.017]
	Highly Severe	(0.335, ∞)	(0.218, ∞)	(0.12, ∞)	(0.017, 0.2]	(0, 0.2]

Table 2.9 Risk Assessment Matrix for Unlikely to be Preventable Incidents

		Frequency				
		Frequent	Probable	Occasional	Remote	Improbable
Severity	Negligible	(0, 0.084]	(0, 0.079]	(0, 0.055]	(0, 0.029]	[0, 0.004]
	Marginal	(0.079, 0.168]	(0.055, 0.16]	(0.029, 0.109]	(0.004, 0.059]	(0, 0.008]
	Critical	(0.160, 0.503]	(0.110, 0.477]	(0.059, 0.327]	(0.10, 0.176]	(0, 0.025]
	Highly Severe	(0.480, ∞)	(0.327, ∞)	(0.176, ∞)	(0.2, ∞)	(0, 0.2]

2.6 Practical Applications

Organizations and industries can implement the proposed method to quantitatively assess the risk of activities within the workplace. An important consideration is how to use the results of the three-

dimensional risk assessment matrix to make decisions that improve workplace safety. In order to use these matrices, different combinations of inherent risk and preventability should be considered.

The residual risk scores can be used to prioritize the implementation of safety improvement practices. For example, if the accident is highly severe, frequent, and highly preventable, it will have a lower residual risk score and be considered less risky than an accident which is also highly severe and frequent, but unlikely to be preventable. It would be logical for organizations to first investigate ways to mitigate the most severe and frequent preventable events, even though the residual risk score might be lower than the score for the less preventable incidents. Next, the highly severe and frequent incidents which are moderately preventable should be investigated. Obviously, the accidents that are categorized as improbable, negligible, and unlikely to be preventable should receive the least investigation. In this way, organizations can optimize their investments in safety improvement processes according to the risk scores while reducing the number of injuries and fatalities.

2.7 Conclusions

This work proposes a new severity metric that incorporates employee and workplace risk factors, as well as a new three-dimensional risk assessment matrix based on residual risk scores, in an effort to include more information in the injury risk assessment process. Using the new Accident Severity Grade (ASG) in the proposed three-dimensional risk assessment matrix, industries can quantify an accident's severity immediately after the incident occurs. This approach allows for real-time monitoring of severity, which will lead to more timely implementation of hazard controls that are specifically targeted toward the most severe accident types and their causes. In order for the proposed system to be as effective as possible, it will be necessary for industries to collect

more detailed data related to worker and workplace factors. The ASG scoring criteria can also be adjusted to accommodate the collection of specialized data that may be relevant in certain industries, locations, or employee demographics. This individualized data will increase the accuracy of the organization's ASG, which will consequently improve the estimation of accident risk. This metric improves upon current severity scores, since severity is now predicted using several risk factors.

2.8 Future Work

There are some limitations of the current study that are worth acknowledging. Due to restrictions associated with publically available occupational injury data, not all potential risk factors could be considered in the case study analysis. For example, information like an employee's lack of training is usually not included in an accident description. Additionally, it is clear that current data collection methods such as the OSHA 300 log were not designed for risk assessment purposes and therefore provide limited data appropriate for this type of surveillance. It was therefore necessary to make some assumptions about the contributing factors to injuries included in the case study, since the accident description rarely states this information explicitly. The collection of more detailed data will allow organizations to avoid these assumptions, and consider the possibility of more than one contributing factor to an incident.

The risk factors discussed in Section 2.3 are common factors derived from literature. However, practitioners can identify relevant risk factors to their work environment, and use this method to predict injury severity. Safety practitioners can determine quantitative severity and risk ranges according to the collected data specific to their industry or industrial setting. The best methods for determining these ranges should be honed in future implementations of this strategy.

There is limited empirical evidence of the validity of the available definitions of preventable workplace accidents. Safety practitioners need to apply the proposed method (i.e., quantifying preventability), and validate the numerical values that are used for different levels of preventability.

Finally, the amount of calculation and data tracking is another limitation of this study. However, the design of a good computer software or a web-based application to calculate the scores and track the metrics will make the surveillance process more convenient for organizations.

References

1. Bureau of Labor Statistics. *Workplace Injury and Illness Summary*. 2011 [cited 2013; Available from: http://www.bls.gov/news.release/osh.nr0.htm.
2. Bureau of Labor Statistics. *Census of Fatal Occupational Injuries Summary*. 2012 [cited 2013; Available from: http://www.bls.gov/news.release/cfoi.nr0.htm.
3. Centers for Diesease Control and Prevention. *NIOSH Strategic Goals*. 2013; Available from: http://www.cdc.gov/niosh/programs/surv/goals.html.
4. Safety, O. and H. Administration, *OSHA technical manual: Section VII, chapter 1*. 1999.
5. OSHA, *FORMULAS for CALCULATING RATES, OSHA Recordable Incident Rate, Lost Time Case Rate, Lost Work Day Rate (LWD), DART Rate, Severity Rate*.
6. Terms, M.-H.D.o.S.T., *Accident Severity Rate*. 2003.
7. Gauchard, G., et al., *Falls and working individuals: role of extrinsic and intrinsic factors.* Ergonomics, 2001. **44**(14): p. 1330-1339.
8. Laflamme, L. and E. Menckel, *Aging and occupational accidents a review of the literature of the last three decades.* Safety Science, 1995. **21**(2): p. 145-161.
9. Kines, P., *Occupational injury risk assessment using injury severity odds ratios: Male falls from heights in the Danish construction industry, 1993-1999.* Human and Ecological Risk Assessment, 2001. **7**(7): p. 1929-1943.
10. Gillen, M., et al., *Injury severity associated with nonfatal construction falls.* American Journal of Industrial Medicine, 1997. **32**(6): p. 647-655.
11. Schuh, A. and J.A. Camelio. *Including Accident Severity in Statistical Monitoring Systems for Occupational Safety*. in *Industrial and Systems Engineering Conference*. 2013.
12. Root, N., *Injuries at work are fewer among older employees.* Monthly Lab. Rev., 1981. **104**: p. 30.
13. Jenkins, E.L. and S.M. Kisner, *Fatal Injuries to Workers in the United States, 1980-1989: A Decade of Surveillance: National and State Profiles*. 1996: DIANE Publishing.
14. Goldberg, R.L., et al., *Fatal occupational injuries in California, 1972–1983.* American Journal of Industrial Medicine, 1989. **15**(2): p. 177-185.
15. Messing, K., et al., *Be the fairest of them all: challenges and recommendations for the treatment of gender in occupational health research.* American Journal of Industrial Medicine, 2003. **43**(6): p. 618-629.
16. DeGroot, D.W., et al., *Epidemiology of US Army Cold Weather Injuries, 19801999.* Aviation, space, and environmental medicine, 2003. **74**(5): p. 564-570.
17. Liao, C.-W. and Y.-H. Perng, *Data mining for occupational injuries in the Taiwan construction industry.* Safety Science, 2008. **46**(7): p. 1091-1102.
18. Leamon, T.B. and P.L. Murphy, *Occupational slips and falls: more than a trivial problem.* Ergonomics, 1995. **38**(3): p. 487-498.
19. Taylor, A.J., et al., *Fatal occupational electrocutions in the United States.* Injury Prevention, 2002. **8**(4): p. 306-312.
20. Maiti, J. and A. Bhattacherjee, *Evaluation of risk of occupational injuries among underground coal mine workers through multinomial logit analysis.* Journal of Safety Research, 1999. **30**(2): p. 93-101.
21. Chang, W.-R., *The effect of surface roughness on the measurement of slip resistance.* International Journal of Industrial Ergonomics, 1999. **24**(3): p. 299-313.

22. STRANDBERG, L., *The effect of conditions underfoot on falling and overexertion accidents.* Ergonomics, 1985. **28**(1): p. 131-147.
23. Simeonov, P. and H. Hsiao, *Height, surface firmness, and visual reference effects on balance control.* Injury Prevention, 2001. **7**(suppl 1): p. i50-i53.
24. Silverstein, B.A., L.J. Fine, and T.J. Armstrong, *Hand wrist cumulative trauma disorders in industry.* British Journal of Industrial Medicine, 1986. **43**(11): p. 779-784.
25. Yassi, A., *Repetitive strain injuries.* The Lancet, 1997. **349**(9056): p. 943-947.
26. Sorock, G., et al., *A case-crossover study of transient risk factors for occupational acute hand injury.* Occupational and environmental medicine, 2004. **61**(4): p. 305-311.
27. Steiner, S.H. and M. Jones, *Risk-adjusted survival time monitoring with an updating exponentially weighted moving average (EWMA) control chart.* Statistics in medicine, 2010. **29**(4): p. 444-454.
28. Wynne-Jones, K., et al., *Limitations of the Parsonnet score for measuring risk stratified mortality in the north west of England.* Heart, 2000. **84**(1): p. 71-78.
29. Nashef, S.A., et al., *European system for cardiac operative risk evaluation (EuroSCORE).* European Journal of Cardio-Thoracic Surgery, 1999. **16**(1): p. 9-13.
30. Lawrence, D., et al., *Parsonnet score is a good predictor of the duration of intensive care unit stay following cardiac surgery.* Heart, 2000. **83**(4): p. 429-432.
31. Geissler, H.J., et al., *Risk stratification in heart surgery: comparison of six score systems.* European Journal of Cardio-Thoracic Surgery, 2000. **17**(4): p. 400-406.
32. Kawachi, Y., et al., *Risk stratification analysis of operative mortality in heart and thoracic aorta surgery: comparison between Parsonnet and EuroSCORE additive model.* European Journal of Cardio-Thoracic Surgery, 2001. **20**(5): p. 961-966.
33. Dzugan, J., *Breaking the Chain, Using risk assessment scores to prevent fishing vessel casualities.* . The Coast gaurd Journal of Safety and Security at Sea, Proceedings of the Marine Safety and Security Council. **67**(4).
34. Darby, P., W. Murray, and R. Raeside, *Applying online fleet driver assessment to help identify, target and reduce occupational road safety risks.* Safety Science, 2009. **47**(3): p. 436-442.
35. U.S. Department of Defense, *Department of Defense Standard Practice for Systems Safety*. 2012.
36. Anthony Tony Cox, L., *What's wrong with risk matrices?* Risk analysis, 2008. **28**(2): p. 497-512.
37. Business Dictionary. *Inherent Risk*. 2013 [cited 2014; Available from: http://www.businessdictionary.com/definition/inherent-risk.html.
38. Amihud, Y. and H. Mendelson, *The Effects of Beta, Bid-Ask Spread, Residual Risk, and Size on Stock Returns.* The Journal of Finance, 1989. **44**(2): p. 479-486.
39. Lehmann, B.N., *Residual risk revisited.* Journal of Econometrics, 1990. **45**(1): p. 71-97.
40. Information Security Handbook. *Inherent and Residual Risk*. [cited 2014; Available from: http://ishandbook.bsewall.com/risk/Assess/Risk/inherent_risk.html.
41. Gurwitz, J.H., et al., *Incidence and preventability of adverse drug events among older persons in the ambulatory setting.* Jama, 2003. **289**(9): p. 1107-1116.
42. Gurwitz, J.H., et al., *Incidence and preventability of adverse drug events in nursing homes.* The American journal of medicine, 2000. **109**(2): p. 87-94.
43. Bates, D.W., L.L. Leape, and S. Petrycki, *Incidence and preventability of adverse drug events in hospitalized adults.* Journal of General Internal Medicine, 1993. **8**(6): p. 289-294.

44. Papadopoulos, I.N., et al., *Preventable prehospital trauma deaths in a Hellenic urban health region: an audit of prehospital trauma care.* Journal of Trauma-Injury, Infection, and Critical Care, 1996. **41**(5): p. 864-869.
45. Rivara, F.P. and D.C. Thompson, *Systematic reviews of injury-prevention strategies for occupational injuries: an overview.* American journal of preventive medicine, 2000. **18**(4): p. 1-3.
46. Watson, C., *Risk Assessment Using the Three Dimensions of Probability (Likelihood), Severity, and Level of Control*. NASA Technical Reports Server.
47. Talbot, J., *What's right with risk matrices?* Jakeman Business Solutions (JBS).
48. National Safety Council, *Top 10 Preventable Workplace Incidents*. 2013.
49. Data.gov. 2013; Available from: https://explore.data.gov/Energy-and-Utilities/Accident-Injuries-Data-Set/qbck-cccw.
50. Coleman, P.J. and J.C. Kerkering, *Measuring mining safety with injury statistics: Lost workdays as indicators of risk.* Journal of Safety Research, 2007. **38**(5): p. 523-533.

3. Can objective early warning scores and subjective risk assessment predict a patient's hospital length of stay and mortality?

3.1 Abstract

During admission, can a patient's in-hospital length of stay (LOS) and mortality be predicted by early risk factors? This paper presents a systems model of patient health outcomes and LOS based on initial health risk and physician assessment of risk. The model elaborates on the interdependent effects of hospital service and a physician's subjective risk assessment on LOS and mortality. The model is used to offer hypotheses about the predictive power of early warnings that are empirically tested by analyzing a detailed dataset of 1,031 patients admitted to a large hospital in the southeastern United States. We find that early warnings are not good predictors of LOS and mortality; rather, LOS is associated more with physicians' subjective risk assessments. Notably, physicians' early assessments of mortality risk are negatively associated with the actual mortality rate, potentially depicting a feedback mechanism that compensates for high values of early risk assessments. Furthermore, controlling for patient-related health risks, we find that different physicians assign different risk values for medically similar patients.

Keywords: Early warning scores; MEWS; Risk management; Length of stay (LOS); Mortality

3.2 Introduction

Risk evaluation and control have been important components of healthcare operations. Ideally, providers would like to predict health risks early during hospital admission and take controlling actions. Different methods and techniques have been developed for this purpose, one of which is the early warning system.

An early warning system is a measurement tool to assess patients' health risk objectively and quickly determine the degree of illness. They aim at reducing the risk of sudden, life-threatening events for patients with the help of hospital rapid response teams [1]. These systems help nurses or patient family members call for a designated group of healthcare professionals to a patient's bedside to react immediately to a deteriorating condition.

There are several major early warning systems; they slightly differ in the parameters they use for risk assessment. The Modified Early Warning Score (MEWS) [2] is a commonly used triage tool to determine quickly the severity of a hospitalized patient's illness [2-5]. Its score is an aggregate number calculated from five major vital signs of the patient: temperature, systolic blood pressure (SBP), heart rate, respiratory rate, and patient's level of consciousness (or AVPU). (We describe the measure in more detail later in this paper.) Other early warning systems follow a similar logic. The Standardized Early Warning System (SEWS) [6] adds oxygen saturation level (SpO2%) to the five MEWS vital signs to detect a patient's deterioration. The Decision-Tree Early Warning Score (DTEWS) [7] is a decision-tree analysis based on a database of vital signs. The National Early Warning Score (NEWS) [8-11] uses seven variables to identify deteriorating patients: respiratory rate, oxygen saturation, any supplemental oxygen, temperature, SBP, heart rate, and level of consciousness.

The main use of these measures is to provide early warnings to health providers to spur quick preventive reactions. Some have argued that these measures have much more to offer, such as helping to predict patients' length of stay (LOS) in hospitals or health outcomes such as the chance of in-hospital mortality [4, 6, 12]. Given the importance of LOS and mortality in assessing healthcare quality, resource allocations, and costs [13], the hope is that early warning scores can make it possible to predict and improve hospital utilization and outcomes.

The purpose of this study is to assess whether a patient's admission MEWS and vital signs are informative in predicting LOS and in-hospital mortality. We build on other studies that have investigated a similar problem (e.g., [13-17], but with a very rich dataset that includes a wide range of variables about patients and physicians. We develop hypotheses derived from a conceptual systemic model of patients' health risk.

We analyze the predictive power of physicians' subjective risk assessments early after admission. Put simply, is there a correlation between physicians' assessment of patients' health risks and what happens to those patients? To that end, we employ in this study two common subjective measures assessed by doctors at the time of admission to emergency departments: severity level and mortality risk. We also analyze factors that influence physicians' subjective risk assessment of patients' illness and their assessment of mortality likelihood.

Table 3.1 Modified Early Warning Score (MEWS) [4]

	3	**2**	**1**	**0**	**1**	**2**	**3**
SBP (mm Hg)	<70	71-80	81-100	101-199		≥200	
Pulse Rate (bpm)		<40	41-50	51-100	101-110	111-129	≥130
Respiratory Rate (bpm)		<9		9-14	15-20	21-29	≥30
Temperature (℃)		<35		35-38.4		≥38.5	
AVPU score				Alert	Reacting to voice	Reacting to pain	Unresponsive

AVPU (A: alert, V: responding to voice, P: responding to painful stimuli, U: unresponsive); SBP (systolic blood pressure)

Obtaining a MEWS involves assigning a number between 0 and 3 to each of the five vital signs mentioned above, as depicted in Table 3.1 [12]. For example, as Table 3.1 shows, the nurse assigns a score of 2 to a patient's SBP that is between 71 and 80 or more than 200. Value 0 is assigned when a vital sign is normal. The sum of the five vital signs yields the patient's total MEWS

between 0 and 15. The total score of 4 or greater than 4 prompts the nurse to call the patient's physician or the hospital's rapid response team.

Mathematically, MEWS for each patient is calculated as an additive function of vital signs, as follows:

$$M_j = \sum_i P_{ij}, \quad \forall i = 1, 2, \ldots, 5$$

(1)

P_{ij} $(0 \leq P_{ij} \leq 3)$ represents the score of vital sign i for patient j;

M_j $(0 \leq M_j \leq 15)$ represents the value of MEWS for patient j.

While a warning system can be an efficient and rapid way of reducing or preventing life-threatening events, there is mixed evidence regarding the predictability of outcome and LOS using different types of early warning scores. On the one hand, MEWS can predict increased risk of death or admission to an intensive care unit (ICU) or high dependency unit (HDU) [2]. Moreover, MEWS can be used to identify patients needing hospital admission as well as those at higher risk of in-hospital death [12]. The proportion of patients admitted to the hospital and who died in the hospital was significantly higher for higher values of MEWS [12]. Furthermore, Standardized Early Warning Score (SEWS) is correlated with a patient's LOS in a hospital [6]. On the other hand, though, some studies suggest further work is needed to derive and validate early warning scores, and that scores that utilize inappropriate parameters and cut-off points should be replaced with ones having higher diagnostic accuracy [18, 19]. However, a potential limitation to these studies is a lack of detailed administrative-level data on patients' characteristics. This can introduce methodological limitations such as the omitted variable bias, and reduce the predictive

power of statistical methods even where significant values are found. Our study differs by taking a systems approach, looking at the entire process, and employing a detailed rich dataset.

The study proceeds as following. In Section 3.3, we develop the hypotheses based on a systems model of risk assessment. We describe the data and our empirical approach in Section 3.4, and report the empirical results in Section 3.5. Our discussions and conclusions are in Sections 3.6 and 3.7.

3.3 Conceptual Model

A systems approach to assessing how predictable early risk indicators are requires considering different reactions within a system to the indicators throughout the provision of healthcare service. In this section, we develop a coherent theory about the predictability of outcomes using early warning scores in healthcare systems.

A linear approach might assume the healthcare system as a black box system, simply turning inputs (i.e., early indicators) to outputs (i.e., mortality and LOS). By contrast, in a systems approach we consider the possibility that the system reacts "endogenously" to the inputs and adapts itself to the observed level of risk [20]. In a systems approach, given risk preventability [21], people's perception, communication, and reaction to risk indicators are central in final outcome [22, 23]. On the one hand, perception is a complex construct based on various environmental and demographic queues [24, 25], and on the other hand, human sensitivity to warnings is not fixed but can change over time due to various reasons such as experiencing too many false alarms [26, 27]. In our specific problem, following a systems approach, physicians and nurses are not simply passive components of the system; rather, by evaluating the risk indicators, their level of attention

to patients and their reactions may change patients' health status. Here, we offer a conceptual model that can still demonstrate the system response to the risk indicators, including the human element of the system.

Figure 3.1 depicts a model of how early risk indicators may influence health outcome (here, mortality) through different paths.

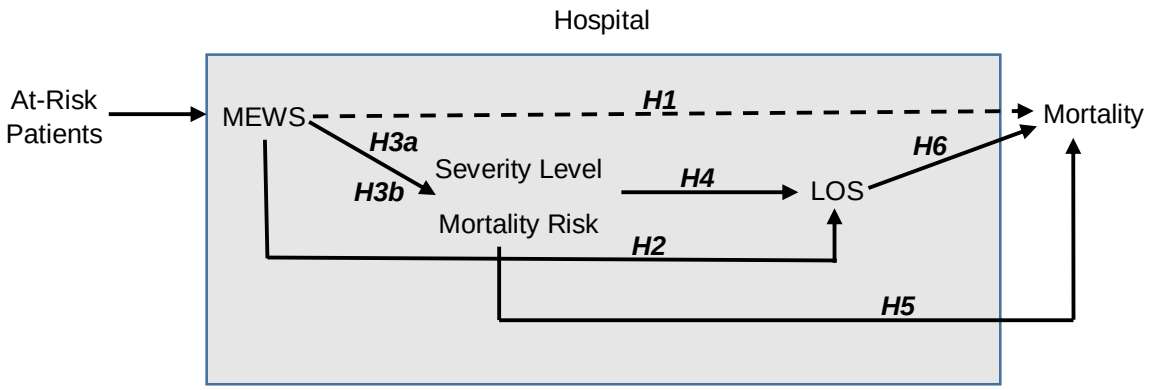

Figure 3.1 Risk assessment steps in hospital -proposed hypotheses graph

The figure shows three major steps inside the hospital and four potential paths through which the final outcome (mortality) might be influenced. The first step is hazard identification for at-risk patients by measuring vital signs and calculating an early warning score; it happens very early, as soon as the patient enters the system. Such early warning scores can be associated with outcome (e.g., a less-sick patient is likely to have a better health outcome) and can also trigger a wide range of reactions from providers that potentially influence the final outcome. What we call the "direct" path (the dashed line in Figure 3.1) is what is usually investigated in simple correlational analysis of MEWS versus mortality. Overall, it is expected that patients who enter the hospital with more severe conditions are more likely to have severe outcomes. The same logic dictates that there should be an association between MEWS and LOS, that is, for higher MEWS values we should have extended LOS.

Therefore, we investigate the following hypotheses:

H1: There is a positive association between MEWS and in-hospital mortality.

H2: There is a positive association between MEWS and hospital LOS.

Healthcare processes involve much more complex actions such as continuous assessment of patients, controlling actions, and treatment procedures that follow physicians' subjective assessments. One major characteristic of such processes is that the high level of "human" (physicians, nurses, etc.) involvement – which includes continuous observation, potential reactions, and changes in decisions in response to sudden changes in patient health status – renders the system more complex and uncertain.

Considering that healthcare is not a passive system, we argue that it can potentially react to early warnings and adapt itself to patients' conditions by offering more intensive care to those who need it. As described in the social judgment theory, the way that human (here, physicians) evaluate various information cues and form a subjective perception of health risks become important [24]. Physicians' risk perception can influence both process-level measures (such as LOS) and outcome measures (such as mortality). In some cases, early risk factors that trigger physicians' subjective assessments and, potentially, the level of service a patient receives in a medical setting can even weaken the first "direct" path.

We hypothesize that physicians are careful observers of patients' initial conditions and so their early subjective assessments should be correlated with objective measures. MEWS in particular may affect physicians' subjective assessments of patients, and those assessments along with several other patient-related variables and physicians' characteristics influence processes. Past studies focused on this part of the system have suggested that clinician judgment alone has a low

sensitivity for detecting critical illness in the pre-hospital environment, and that adding MEWS improves detection at the cost of reduced specificity [28]. Thus, we are particularly interested to see how those subjective measurements are constructed, and whether MEWS affects a physician's subjective assessment. Thus we offer the following hypothesis:

H3: There is a positive association between MEWS and the subjective assessment of physicians.

As there are two measures of subjective assessment (severity level and mortality risk), we investigate the association between MEWS and each of those separately, referring to them as Hypothesis 3a and 3b, respectively.

Following the path, physicians' risk assessment can influence process-level variables. This is the human component of the system, depicting that the healthcare system is not a pure technical (medical) system but in fact a socio-technical system in which physicians' perceptions of risk may be as important as or even more important than objective measures about patients. LOS represents the process intensity in our model. While we agree there may be other variables that can better represent the level of attention to patients, LOS is more accessible and comparable across patients. Thus, simply put, we assume that longer care represents more attention to patients and offer the following hypothesis:

H4: There is a positive association between a physician's subjective assessment of a patient and LOS.

Such subjective assessments and process intensity can influence outcomes. They may even compensate for the initial condition by making it that patients with high-risk initial indicators receive more care. If such an endogenous reaction exists in the system, we expect higher assessment of mortality risks to result in lower mortality incidents. If physicians are incapable of

reacting properly to the conditions despite risk assessments, we may observe a positive relation between the assessments and mortality. Whether the association is negative or positive is an empirical question we investigate in this hypothesis:

H5: There is an association between a physician's subjective assessment of a patient and in-hospital mortality.

Finally, we expect a relation between the severity of health conditions and LOS and mortality. Patients with worse health situations may stay in the hospital longer and may have a higher likelihood of death. Thus we offer the following hypothesis:

H6: There is a positive association between LOS and in-hospital mortality.

Investigating these six hypotheses helps reveal the dominant path in the healthcare system. If patients' initial conditions and the direct path are dominant, we expect H1 and H2 to be significant. If internal human dynamics are dominant, we expect H4, and H5 to be significant and H3 will help uncover how those internal human dynamics are triggered. H6 may be significant in both cases. We also supported our hypothesis with a dynamic, stochastic simulation model which can be found in Appendix A of this dissertation.

3.4 Methodology

3.4.1 Data

Our hypotheses are tested empirically using patient-level data. We collected detailed data for 1,031 randomly selected patients admitted between January and September 2014 to a 500-plus-bed medical center in the state of Virginia, United States. Patient records were accessed through the electronic patient management systems (MEDITECH and Crimson), which contain full patient

demographic data along with individual records and medical documents including vital signs (i.e., blood pressure, pulse rate, temperature, respiratory rate, and AVPU) and MEWS. We also had access to clinician judgment measures – specifically, severity level and mortality risk – assigned to each patient by the attending physician, as well as patient outcomes such as hospital LOS and disposition conditions through the MEDITECH system. Physicians' identifiers were included in the data. Newborns and patients admitted to the ICU were excluded from our study. Of the initial sample of 1,031 randomly selected patients, we dropped the data of 10 patients since their MEWS values were not reported.

Table 3.2 shows the descriptive statistics for some of the main variables in our analysis. In addition, we controlled for gender, weight, height, BMI, physician, and additional physiological measurements including oxygen saturation level and diastolic blood pressure (DBP). The variable *physician* is a nominal variable assigned to each attending physician who makes the subjective measures. AVPU is a categorical and nominal variable with four values (A, P, V, and U).

Table 3.2 Descriptive Statistics of Some Variables

Variable	Sample N	Mean	Std. Dev.	Min	Max
Severity Level (1 to 4)	1021	2.12	0.87	1	4
Mortality Risk (1 to 4)	1021	1.91	0.94	1	4
Mortality (0 or 1)	1021	0.04	0.18	0	1
Length of Stays (Days)	1021	4.72	4.29	1	31
MEWS	1021	1.60	1.17	0	9
Age	1021	66.55	16.94	18	101
Temperature	1021	97.69	4.48	33.7	103.2
Pulse Rate	1021	82.63	19.09	20	160
Respiratory Rate	1021	18.43	3.07	8	44
SBP	1021	135.97	24.72	79	243

We also investigated the correlations between any pairs of variables. The results are in the Appendix B, Table B.1. We found no serious correlation between the variables other than that between BMI and weight ($r = 0.89$) and between two subjective assessments of physicians, that is, mortality risk and severity level ($r = 0.73$) – which were expected.

3.4.2 Statistical Model and Data Analysis

We ran four different sets of regression analyses with different dependent variables to explore our hypotheses. For the first and second analyses, the dependent variables were the physician's judgments of severity level and mortality risk. These models explored hypotheses 3a and 3b and are ordinary least square models. LOS was the dependent variable for the third analysis, in which we explored hypotheses H2 and H4. The models were also ordinary least square models. In the fourth analysis, the dependent variable was mortality. Mortality is a binary variable and refers to the patient who died before discharge. We used logistic regression models and explored hypotheses H1, H5, and H6.

For each analysis, we ran several models with different predictors. At the first stage, every outcome variable was regressed over MEWS, and then we added other predictors such as vital signs and patient features in every step. We compared the models and monitored the significance of each predictor.

3.5 Results

3.5.1 Assessment of Severity Level

Table 3.3 shows the regression analysis for the severity level for six different models. In model 1, we controlled severity over MEWS, and added vital signs and additional physiological measures in models 2 and 3. Patient features were added in model 4, and *physician* was added as a fixed effect in model 5. Finally, in model 5' we controlled for the patient's disease and other health issues. In our analysis, we used attending physicians who are ultimately responsible for all patient care, and assumed primary care for each patient. As soon as they arrive at the hospital, patients are assigned to an attending physician randomly based on their availability. A patient's disease group was specified according to the attending physician's specialty. Our data included 105 physicians and 25 disease groups.

The results show MEWS to be a significant variable in three of the analyses, meaning that higher severity values are assigned to patients with higher MEWS values. All vital signs are significant predictors of severity in model 2. However, after adding patient demographics (model 4) and controlling for *physician* and disease group (models 5 and 5') as model predictors, MEWS loses its predictive power. In the final model, which has a considerably higher R^2, MEWS is not significant, while age, three of the vital signs, and a physiological measure are significant predictors of physicians' subjective assessment.

Notably in model 5, when we run a fixed effect model controlling for *physician* as dummy variables, the predictive power of the model almost doubles, reaching to $R^2 = 0.38$. In this model, *physician* is significant. Simply put, controlling for patients' characteristics, physicians show variation in their assessment, meaning that different physicians potentially assign different severity

levels to patients with similar physiological conditions. We observe similar results in model 5', where the type of disease is controlled along with vital signs. However, *physician* is a better predictor of severity level than the type of disease, since the R^2 value decreases to 0.31 in model 5'.

Moreover, patient characteristics such as age and gender are significant predictors of severity level in model 5. This implies that older patients are assigned higher values of severity than younger patients, and females are assigned lower values of severity than males.

In summary, the analysis shows that MEWS is not a good predictor of severity level (no support for hypothesis *H3a*) once we control for demographic characteristics of patients, especially age and gender, and for *physician*. Controlling for *physician* doubles the predictive power of our model, depicting significant variation across physicians.

Table 3.3 Regression analysis for the severity level

Source (x_i)	**M1**	**M2**	**M3**	**M4**	**M5**	**M5'**
Patient Features						
Age				0.01*** (0.001)	0.01*** (0.002)	0.01*** (0.002)
Gender				-0.09** (0.03)	-0.06* (0.03)	-0.07** (0.03)
Weight				0.004 (0.01)	0.003 (0.01)	0.002 (0.01)
Height				-0.003 (0.01)	-0.001 (0.01)	-0.002 (0.01)
BMI				-0.01 (0.02)	-0.003 (0.02)	-0.001 (0.02)
MEWS	0.22*** (0.02)	0.07** (0.04)	0.07** (0.04)	0.06 (0.04)	0.03 (0.03)	0.04 (0.03)
Vital Signs						
Temperature		0.1** (0.01)	0.01** (0.01)	0.01* (0.01)	0.01 (0.01)	0.01 (0.01)
Pulse Rate		0.01*** (0.002)	0.01*** (0.002)	0.01*** (0.001)	0.01*** (0.002)	0.01*** (0.001)
Respiratory		0.02** (0.01)	0.02* (0.01)	0.02* (0.01)	0.02* (0.01)	0.01 (0.01)
SBP		-0.004***	-0.001	-0.002**	-0.004***	-0.003**

		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
AVPU		Significant	Significant	Significant	Significant	Significant
Additional Physiological Measures						
DBP			-0.01***	-0.01***	-0.01**	-0.01***
			(0.002)	(0.002)	(0.002)	(0.002)
SpO2%			-0.01	-0.002	-0.003	-0.003
			(0.01)	(0.01)	(0.01)	(0.01)
Physician					Significant	
Disease Group						Significant
Intercept	1.85***	0.80	1.95	1.48	1.74	1.90
	(0.04)	(0.62)	(1.07)	(1.92)	(1.92)	(1.83)
R^2	0.09	0.15	0.16	0.20	0.38	0.31
R^2 adjusted	0.09	0.14	0.16	0.19	0.30	0.28
Observations	1021	1021	1010	1010	1010	1010

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Note: Standard errors are in parentheses.

3.5.2 Assessment of Mortality Risk

Mortality risk is another subjective measure of physician. Table 3.4 represents the regression analysis for the mortality risk for six different models. We followed steps similar to those discussed in Section 3.5.1 and controlled mortality risk over MEWS, vital signs, patient features, *physician*, and disease group in every step.

The results show that MEWS is a significant variable in four of the analyses, meaning that higher mortality risk values are assigned to patients with higher MEWS values. Other than respiratory rate, four of the vital signs are significant predictors of mortality risk in model 2. Age is a significant predictor of mortality risk in models 4, 5, and 5'. Adding *physician* as a fixed effect to model 5 eliminates the significance of MEWS but improves R^2 from 0.27 to 0.42. Every physician, however, has different criteria for assigning mortality risk values. In other words, different physicians have different perceptions of mortality risk for similar patients. Controlling for patient's disease in model 5' reduces the R^2 value.

In summary, after controlling for *physician*, MEWS loses its significance (no support for *H3b*). Controlling for *physician* increases the predictive power of our model, depicting significant variation across physicians on how risk of mortality is assessed even though we control for a wide range of vital signs.

Table 3.4 Regression analysis for the mortality risk

Source (x_i)	**M1**	**M2**	**M3**	**M4**	**M5**	**M5'**
Patient Features						
Age				0.02*** (0.002)	0.02*** (0.002)	0.02*** (0.002)
Gender				-0.09** (0.04)	-0.06 (0.04)	-0.07* (0.04)
Weight				-0.004 (0.01)	-0.01 (0.01)	-0.01 (0.01)
Height				0.007 (0.01)	0.01 (0.01)	0.01 (0.01)
BMI				0.01 (0.03)	0.02 (0.03)	0.03 (0.02)
MEWS	0.24*** (0.02)	0.11** (0.04)	0.11*** (0.04)	0.08** (0.04)	0.06 (0.04)	0.05 (0.04)
Vital Signs						
Temperature		0.02*** (0.01)	0.01** (0.01)	0.01** (0.01)	0.01 (0.01)	0.01 (0.01)
Pulse Rate		0.01** (0.002)	0.01*** (0.002)	0.01*** (0.001)	0.01*** (0.002)	0.01*** (0.002)
Respiratory		0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
SBP		-0.002*** (0.001)	0.00 (0.001)	-0.004*** (0.001)	-0.004*** (0.001)	-0.004*** (0.001)
AVPU		Significant	Significant	Significant	Significant	Significant
Additional Physiological Measures						
DBP			-0.01*** (0.002)	-0.002 (0.002)	0.00 (0.01)	0.00 (0.002)
SpO2%			-0.01 (0.01)	-0.001 (0.01)	-0.01 (0.01)	-0.005 (0.01)
Physician					Significant	
Disease Group						Significant
Intercept	1.53*** (0.05)	0.22 (0.68)	1.90 (1.19)	-1.75 (1.99)	1.74 (1.92)	-1.63 (1.92)
R^2	0.09	0.12	0.13	0.27	0.42	0.36

R^2 adjusted	0.09	0.11	0.13	0.26	0.34	0.33
Observations	1021	1021	1010	1010	1010	1010

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Note: Standard errors are in parentheses.

3.5.3 Length of Stay

Table 3.5 summarizes the main results of our hospital LOS analysis. We controlled for similar variables as described earlier, in addition to physicians' two subjective assessments (severity level and mortality risk).

The first column in Table 3.5 shows that MEWS is a significant predictor of LOS. Controlling for vital signs, though, renders MEWS no longer significant. Temperature, pulse rate, and SBP are the significant control variables in model 2, which means that patients with higher temperatures and pulse rates and lower SBPs stay longer in hospital. Oxygen saturation level (SpO2) and DBP are added as predictors in model 3, in which SBP is no longer significant and DBP turns to be a significant variable. This implies that patients with lower DBPs will stay longer in hospital. DBP loses its significance after patient demographics were added in model 4. Age, gender, height and BMI are significant patient characteristics of LOS. Therefore, older, female, and taller patients, as well as patients with large BMI values, have higher hospital LOS compared to others.

Notably, Model 5 shows that while MEWS – an objective assessment tool – is not a significant predictor of length of stay, severity level and mortality risk – both subjective measures – are significant predictors of LOS (support for *H4*). This suggests that physicians' early assessments of risk are associated with LOS.

In model 6, controlling for *physician* fixed-effect, R^2 more than doubles and *physician* becomes a significant predictor of LOS. Simply put, *physician* is as good a predictor as all other variables in our model in predicting LOS, which implies that patients of specific physicians stay longer than

those of other physicians. Controlling for disease group in model 6' shows that disease is a significant predictor of LOS, meaning that patients with special health issues stay longer in the hospital, even though all vital signs are controlled. However, overall, *physician* seems to be a better predictor of LOS, since the R^2 value in model 6' is reduced in compare to model 6.

In summary, our analysis shows that LOS can be predicted at the time of admission according to some predictors – gender, weight, height, BMI, severity level, and mortality risk – but not MEWS. So, while MEWS itself is not a good predictor of LOS (no support for *H2*), some of its components (vital signs) can be used to predict hospital LOS. Conversely, our analysis provides evidence that a clinician's judgment is a better predictor of LOS than is MEWS.

To examine the robustness of our findings, we use the log transformation on the dependent variables and replicate the models. Table B.2 in the Appendix B summarizes the results. The model R^2 improved slightly and similar results were obtained: physicians' subjective assessments and *physician* are still significant predictors of LOS. Moreover, DBP and pulse rate contribute significantly to LOS prediction. Patients with lower values of DBP and higher pulse rates stay longer in the hospital.

Table 3.5 Regression analysis for length of stay

Source (x_i)	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M6'**
Patient Features							
Age				0.02**	-0.003	0.004	-0.0004
				(0.01)	(0.009)	(0.01)	(0.01)
Gender				0.32*	0.51***	0.32*	0.43**
				(0.19)	(0.17)	(0.18)	(0.18)
Weight				-0.07	-0.07*	-0.08*	-0.07*
				(0.05)	(0.04)	(0.04)	(0.04)
Height				0.09*	0.09**	0.09*	0.09*
				(0.05)	(0.05)	(0.05)	(0.05)
BMI				0.22*	0.23*	0.24*	0.24*
				(0.13)	(0.22)	(0.13)	(0.12)

MEWS	0.57*** (0.11)	0.27 (0.19)	0.28 (0.19)	0.30 (0.19)	0.16 (0.18)	0.26 (0.18)	0.17 (0.18)
Vital Signs							
Temperature		0.1*** (0.03)	0.09*** (0.03)	0.09*** (0.03)	0.07** (0.03)	0.06 (0.03)	0.06** (0.03)
Pulse Rate		0.02** (0.01)	0.03*** (0.01)	0.03*** (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)
Respiratory		0.01 (0.05)	0.01 (0.05)	-0.001 (0.05)	-0.04 (0.05)	-0.05 (0.05)	-0.04 (0.05)
SBP		-0.01* (0.01)	0.00 (0.01)	-0.004 (0.01)	0.002 (0.01)	0.01 (0.01)	0.004 (0.01)
AVPU		Not Significant	Not Significant	Not Significant	Not Significant	Not Significant	Not Significant
Additional Physiological Measures							
DBP			-0.03*** (0.11)	-0.02 (0.01)	-0.01 (0.01)	-0.01 (0.1)	-0.01 (0.01)
SpO2%			0.03 (0.05)	0.05 (0.05)	0.05 (0.04)	0.05 (0.04)	0.05 (0.04)
Subjective Assessments							
Severity Level					1.76*** (0.22)	1.77*** (0.23)	1.72*** (0.22)
Mortality Risk					0.47** (0.21)	0.53** (0.22)	0.56*** (0.21)
Physician						Significant	
Disease Group							Significant
Intercept	3.80*** (0.22)	-4.47 (3.23)	-6.55 (0.25)	-25.21** (10.34)	-27.02*** (9.52)	-26.22*** (9.92)	-26.72*** (9.56)
R^2	0.02	0.05	0.05	0.06	0.21	0.33	0.24
R^2 adjusted	0.02	0.04	0.04	0.05	0.20	0.24	0.21
Observations	1021	1021	1010	1010	1010	1010	1010

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Note: Standard errors are in parentheses.

3.5.4 Mortality

Table 3.6 shows the results of our analysis of mortality. We controlled for the same variables as those described in Section 3.5.3. The results indicate that MEWS is significant only in model 1.

Adding vital signs and additional physiological measures to the model reveals respiratory rate, level of consciousness, and oxygen saturation level as significant predictors of in-hospital mortality. Weight, height, BMI, mortality risk, and LOS are significant in three of the models (models 5, 6, and 6').

The significance of LOS implies a correlation between LOS and mortality (support for *H6*). *Physician* in model 6 is not a significant predictor of mortality, meaning that despite all variations among physicians regarding how patients are assessed and how long they stay in the hospital, we see no significant variation across physicians when it comes to mortality.

It is interesting to note that there is a negative association between physicians' subjective assessments and mortality. Physicians' subjective assessment of severity level in model 5 is slightly significant with a negative sign. Moreover, physicians' subjective assessment of mortality risk is negatively correlated with actual mortality. This leads to the conclusion that patients with higher mortality risk values at the time of admission are less likely to die. This evidence points to a possible feedback mechanism that compensates for high values of early risk assessment, since patients assessed as high risk at admission are likely to receive more attention. In other words, higher risk assessment at admission leads to lower risk of mortality at exit.

Finally, controlling for disease group in model 6' shows that a patient's disease is not a significant predictor of mortality, which implies that patients with special health issues do not die more than others.

Table 3.6 Regression analysis for mortality

Source (x_i)	M1	M2	M3	M4	M5	M6	M6'
Patient Features							
Age				-0.05*** (0.02)	-0.02 (0.02)	-0.02 (0.02)	-0.02 (0.02)
Gender				0.62** (0.29)	0.47 (0.32)	0.39 (0.36)	0.53 (0.33)
Weight				-0.14** (0.07)	-0.15** (0.07)	-0.25** (0.10)	-0.18** (0.08)
Height				0.13* (0.07)	0.14* (0.07)	0.22** (0.10)	0.16** (0.08)
BMI				0.45** (0.20)	0.48** (0.22)	0.77*** (0.30)	0.56** (0.25)
MEWS	-0.52*** (0.09)	-0.13 (0.22)	0.04 (0.23)	0.14 (0.24)	0.22 (0.27)	0.29 (0.39)	0.35 (0.29)
Vital Signs							
Temperature		0.02 (0.02)	0.03 (0.02)	0.04* (0.02)	0.06* (0.03)	0.29 (0.25)	0.18 (0.18)
Pulse Rate		-0.005 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.08)	-0.01 (0.02)
Respiratory		-0.09* (0.05)	-0.11* (0.05)	-0.14** (0.06)	-0.11* (0.06)	-0.11 (0.08)	-0.13** (0.06)
SBP		0.01 (0.01)	0.004 (0.01)	0.01 (0.01)	0.003 (0.01)	0.001 (0.01)	0.002 (0.01)
AVPU		Significant	Significant	Not Significant	Not Significant	Not Significant	Not Significant
Additional Physiological Measures							
DBP			0.01 (0.02)	-0.003 (0.02)	-0.01 (0.02)	0.00 (0.02)	-0.01 (0.02)
SpO2%			0.12*** (0.05)	0.12** (0.05)	0.10* (0.05)	0.09 (0.07)	0.011* (0.06)
Subjective Assessment							
Severity Level					-0.79* (0.43)	-0.82 (0.51)	-0.78* (0.43)
Mortality Risk					-1.36*** (0.39)	-1.50*** (0.52)	-1.43*** (0.40)
LOS					0.13** (0.05)	0.19*** (0.07)	0.15*** (0.05)
Physician						Not Significant	

Disease Group							Not Significant
Intercept	4.40*** (0.29)	2.35 (2.18)	-10.87** (5.24)	-31.22** (13.21)	-28.39** (14.09)	-50.16 (16318.33)	-29.09 (27034.22)
R^2	0.09	0.12	0.15	0.23	0.42	0.56	0.46
Log Likelihood	-141.71	-137.4	-132.60	-118.95	-90.21	-69.07	-84.66
Observations	1021	1021	1010	1010	1010	1010	1010

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Note: Standard errors are in parentheses.

3.6 Discussion

Figure 3.2 summarizes our findings. Briefly, the analysis shows that mortality can be predicted ahead of time by some predictors such as weight, height, physicians' assessment of mortality risk (though, in the opposite direction!), and LOS. These provide support for *H5* and *H6*. In other words, after controlling for LOS and physician's subjective assessment on mortality (i.e., mortality risk), MEWS and its components are not good predictors of mortality (no support for *H1*).

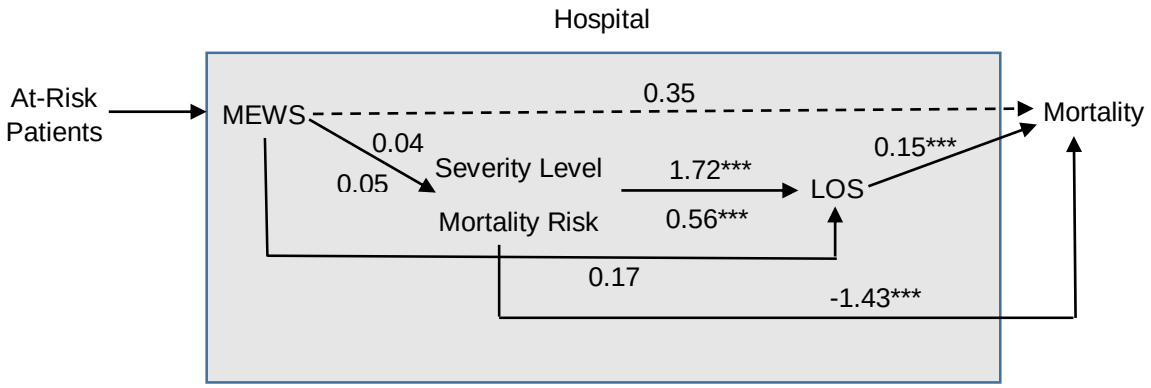

Figure 3.2 A summary of the results of analysis ($***p < 0.01$, $**p < 0.05$, $*p < 0.1$)

Our results contradict findings from previous studies [4, 12]. We believe this is because we control for a much greater number of variables in our analysis than in other studies, and we have a better control of in-hospital processes such as physicians' subjective assessments, which trigger the intensity of healthcare services. In particular, in our models the significance of MEWS disappears

as we add patients' characteristics and physicians' subjective assessments. Our results imply that although LOS and risk of mortality may increase for the higher values of MEWS where only patient's physiological data is used, MEWS is not as an effective predictor on its own when the role of humans in the system is added. Our study differs theoretically from previous studies because we analyze the entire process, taking a systems approach and including how physicians react to initial vital signs and how those reactions may prevent catastrophic events such as mortality. Our analysis shows that such hospital-level processes are effective in preventing undesired outcomes. In sum, these differences from previous studies comprise our investigation's main contributions to risk analysis in healthcare.

This study has several implications for designing risk assessment systems. While research on the effectiveness of early warning scores in predicting healthcare outcome and LOS rely heavily on the consideration of patient's physiological factors, it typically fails to consider the contributions of healthcare providers (physicians and nurses) in the hospital. Overall, our study shows that early warnings are not good predictor of final outcome. In a sense, this finding can be encouraging, since the primary goal of developing early warning scores is early reaction to and prevention of catastrophic outcomes. For the purpose of outcome prediction, other metrics are needed as part of the final outcome prediction tool, including physicians' reactions and the healthcare system's performance.

This study has several limitations that suggest future avenues of research. While we used a rich data source, we cannot rule out all potential variables that might affect the outcome and patients' LOS. While we tried to use most of the available variables, including patient's demographic data and physician's subjective assessments, it could be useful to study the effect of organizational variables as well as insurance payers on LOS. Future studies could analyze multiple hospitals with

different patient demographics and over a longer period. Keeping track of what happens to patients after discharge could also enrich the analysis. We also suggest an analysis of risk indicators over the length of a patient's hospital stay, that is, how patient health risks change over the period in the hospital – including the potential side effects of long hospital stays.

3.7 Conclusions

We analyzed the predictability of hospital LOS and in-hospital mortality by using MEWS, vital signs, demographic data, and physicians' subjective assessments. Our main hypothesis was that LOS is extended for the higher values of MEWS and the probability of mortality increases. In addition, we analyzed the effect of patients' MEWS on physicians' subjective assessments and the effects of those assessments on LOS and mortality rate. We show that MEWS is not a good predictor of hospital LOS and mortality. Physicians' subjective assessments of patients (i.e., severity level and mortality risk) are better predictors of both outcome (mortality) and LOS. Furthermore, unobservable physicians' characteristics (represented as fixed-effect controls) are very strong predictors of LOS, which means that patients of specific physicians may stay longer in the hospital. Moreover, how humans (physician) react and make judgments are better predictors of both LOS and outcome than patient physiological measures (MEWS). Overall, subjective assessment in our context were better predictors of the process (represented in LOS), and were inversely predictors of death, while MEWS as an objective measure was not associated with LOS or death.

References

1. Institute for Healthcare Improvement. *Early Warning System: Scorecards That Save Lives*. Improvement Stories 2014 [cited 2014; Available from: http://www.ihi.org/resources/Pages/ImprovementStories/EarlyWarningSystemsScorecardsThatSaveLives.aspx.
2. Subbe, C., et al., *Validation of a modified Early Warning Score in medical admissions.* Qjm, 2001. **94**(10): p. 521-526.
3. Gardner-Thorpe, J., et al., *The value of Modified Early Warning Score (MEWS) in surgical in-patients: a prospective observational study.* Annals of the Royal College of Surgeons of England, 2006. **88**(6): p. 571.
4. Cei, M., C. Bartolomei, and N. Mumoli, *In-hospital mortality and morbidity of elderly medical patients can be predicted at admission by the Modified Early Warning Score: a prospective study.* International journal of clinical practice, 2009. **63**(4): p. 591-595.
5. Subbe, C., et al., *Effect of introducing the Modified Early Warning score on clinical outcomes, cardio-pulmonary arrests and intensive care utilisation in acute medical admissions.* Anaesthesia, 2003. **58**(8): p. 797-802.
6. Paterson, R., et al., *Prediction of in-hospital mortality and length of stay using an early warning scoring system: clinical audit.* Clinical Medicine, 2006. **6**(3): p. 281-284.
7. Badriyah, T., et al., *Decision-tree early warning score (DTEWS) validates the design of the National Early Warning Score (NEWS).* Resuscitation, 2014. **85**(3): p. 418-423.
8. Prytherch, D.R., et al., *ViEWS—towards a national early warning score for detecting adult inpatient deterioration.* Resuscitation, 2010. **81**(8): p. 932-937.
9. Smith, G.B., et al., *The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death.* Resuscitation, 2013. **84**(4): p. 465-470.
10. McGinley, A. and R.M. Pearse, *A national early warning score for acutely ill patients.* Bmj, 2012. **345**.
11. Hill, K., *National Early Warning Score.* Nursing in Critical Care, 2012. **17**(6): p. 318-318.
12. Burch, V., G. Tarr, and C. Morroni, *Modified early warning score predicts the need for hospital admission and inhospital mortality.* Emergency Medicine Journal, 2008. **25**(10): p. 674-678.
13. Brownell, M. and N. Roos, *Variation in length of stay as a measure of efficiency in Manitoba hospitals.* CMAJ: Canadian Medical Association Journal, 1995. **152**(5): p. 675.
14. Blais, M.A., et al., *Predicting length of stay on an acute care medical psychiatric inpatient service.* Administration and Policy in Mental Health and Mental Health Services Research, 2003. **31**(1): p. 15-29.
15. Toumpoulis, I.K., et al., *Does EuroSCORE predict length of stay and specific postoperative complications after cardiac surgery?* European journal of cardio-thoracic surgery, 2005. **27**(1): p. 128-133.
16. Lavoie, A., et al., *The Injury Severity Score or the New Injury Severity Score for predicting intensive care unit admission and hospital length of stay?* Injury, 2005. **36**(4): p. 477-483.
17. Wee, J.Y., H. Wong, and A. Palepu, *Validation of the Berg Balance Scale as a predictor of length of stay and discharge destination in stroke rehabilitation.* Archives of physical medicine and rehabilitation, 2003. **84**(5): p. 731-735.
18. Cuthbertson, B.H. and G. Smith, *A warning on early-warning scores!* British journal of anaesthesia, 2007. **98**(6): p. 704-706.
19. Barlow, G., D. Nathwani, and P. Davey, *The CURB65 pneumonia severity score outperforms generic sepsis and early warning scores in predicting mortality in community-acquired pneumonia*. Thorax, 2007. **62**(3): p. 253-259.

20. Ghaffarzadegan, N., *How a System Backfires: Dynamics of Redundancy Problems in Security.* Risk analysis, 2008. **28**(6): p. 1669-1687.
21. Azadeh-Fard, N., et al., *Risk assessment of occupational injuries using Accident Severity Grade.* Safety science, 2015. **76**: p. 160-167.
22. Fischhoff, B., *Risk perception and communication unplugged: Twenty years of process1.* Risk analysis, 1995. **15**(2): p. 137-145.
23. Slovic, P., et al., *Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality.* Risk analysis, 2004. **24**(2): p. 311-322.
24. Hammond, K.R. and T.R. Stewart, *The essential brunswik: Beginnings, explications, applications*. 2001: Oxford University Press.
25. Mumpower, J.L., et al., *Psychometric and demographic predictors of the perceived risk of terrorist threats and the willingness to pay for terrorism risk management programs.* Risk analysis, 2013. **33**(10): p. 1802-1811.
26. Paté-Cornell, M.E., *Warning systems in risk management.* Risk analysis, 1986. **6**(2): p. 223-234.
27. Ghaffarzadegan, N. and D.F. Andersen, *Modeling behavioral complexities of warning issuance for domestic security: A simulation approach to develop public management theories.* International Public Management Journal, 2012. **15**(3): p. 337-363.
28. Fullerton, J.N., et al., *Is the Modified Early Warning Score (MEWS) superior to clinician judgement in detecting critical illness in the pre-hospital environment?* Resuscitation, 2012. **83**(5): p. 557-562.

4. An Analysis of Mixed Evidence for the Predictive Power of Early Warning Systems in Healthcare

4.1 Abstract

Early warning systems have been widely used in healthcare to predict adverse outcome. The objective of this study is to systematically review the available literature to assess the predictive power of early warning systems and prognostic risk indicators in predicting different outcomes in health such as mortality, disease diagnosis, adverse outcomes, care intensity, and survival. Web of Science database between January 2000 and March 2015 was searched for relevant papers. Inclusion criteria were original empirical studies that assessed prediction tests by reporting *sensitivity* and *specificity*, or area under the receiver operating characteristic curve (*ROC AUC*). Exclusion criteria were previous reviews, conference proceedings, and non-English studies. A total of 44 papers met our criteria. We analyzed their methods of study and results. Our analysis shows that the reviewed papers provide mixed evidence indicating that there is no robust trend in predictive power of these early indicators. Average ROC AUC and sensitivity of these tests were relatively low. Based on the analysis, we propose a dynamic model of healthcare risk assessment which hypothesizes a potential source of mixed evidence.

Keywords: Early warning system, clinical outcome, mortality, adverse outcome, health risk management

4.2 Introduction

Early prediction of adverse outcomes are at the heart of risk management, and help decision makers take appropriate actions to mitigate the potential risks. For example, physicians would like to

prevent any adverse outcomes such as cardiac arrest, or mortality for their patients, and like to use all pieces of information that help accurate early diagnosis and interventions. A systematic way to be proactive and be prepared for risky events is to develop and use an early warning system which encompass a wide range of indicators and offer an accurate prediction [1-3].

Early warnings in healthcare are aimed at providing timely predictions of high consequent and catastrophic events [4, 5]. They are often used to prevent catastrophic events such as intensive care unit (ICU) admission and in-hospital mortality [2]. In simple words, early warning systems consist of judgment and decision making models which utilize various pieces of information as input and predict chances of specific critical events as outputs. Strength of such systems depend on the accuracy of predictions as well as preventability of outcome. All decision makers would like to receive early and trustable warnings, however, the question is that with the growing complexities and uncertainties in healthcare domain, how useful these early warning systems are? The concern is whether early warning systems and prognostic indicators are sensitive and specific enough to predict health outcome.

A wide range of studies have been conducted to develop early warning indicators or assess the predictive power of common warning indicators in predicting healthcare outcome. The results are mixed [6]. In our recent empirical investigation of predictive power of early warnings, we found that the Modified Early Warning Score (MEWS) [1], a common early warning system, cannot predict hospital length of stay and mortality when one controls for subjective assessments of physicians. This brings up the question that how useful early warning measures are and how they can be improved. Understanding the state of knowledge and offering an overall analysis of the results can help further develop better warning systems. This is the main goal of this study. We

will offer a comprehensive analysis of the existing empirical evidence and evaluate conditions where early warning measures have had a better predictive power.

In addition, this study connects to the previous papers of this dissertation [6, 7], by taking a step back and providing a theoretical framework on why risk indicators can or cannot predict healthcare outcomes, and how better predictors can be designed.

In following we layout our research method and results of the study, and then discuss how early warning systems could be improved to help risk management in healthcare.

4.3 Research Method

4.3.1 Search Strategy

We conducted a systematic review to assess the predictive power of early warning scores in predicting adverse outcomes. We did not limit our study to ones that specifically mention the term "early warning" but we also included studies that use any predictor or prognostic variables to assess the predictability of a clinical outcome. We looked for relevant papers listed in Web of Science. Search terms included: (*early warning score* or *prognostic indicator* or *leading indicator*) and *prediction*.

4.3.2 Inclusion/Exclusion Criteria

Peer reviewed full-text papers written in English published between January 2000 and February 2015 were included. The inclusion and exclusion criteria were fully specified in the review protocol before conducting the review. Papers had to conduct an empirical analysis on

predictability of outcome variables using early warning scores or prognostic indicators. Papers were excluded if they did qualitative studies, conceptual design, or did not report sensitivity, specificity, or the area under receiver operating characteristic curve (ROC AUC) in their analysis.

4.3.3 Review Process and Search Results

The review process is represented in Figure 4.1. Using our keywords we identified 976 papers from Web of Science. After removing 148 duplicate papers, and conference proceedings, titles and abstracts of 828 papers were screened and reviewed. After excluding non-empirical studies, review papers, and studies not implemented in healthcare domain, 269 papers remained. We then focused on study design and results; 99 papers were remained that evaluated the predictive power of risk indicators in predicting outcome variable. In the next step, by reviewing the full text of these 99 papers, 54 studies were excluded because they did not report sensitivity and specificity, or ROC AUC measures in their analysis. Finally, 44 papers satisfied our inclusion criteria. A brief information on these measures are offered in Section 4.2.5.

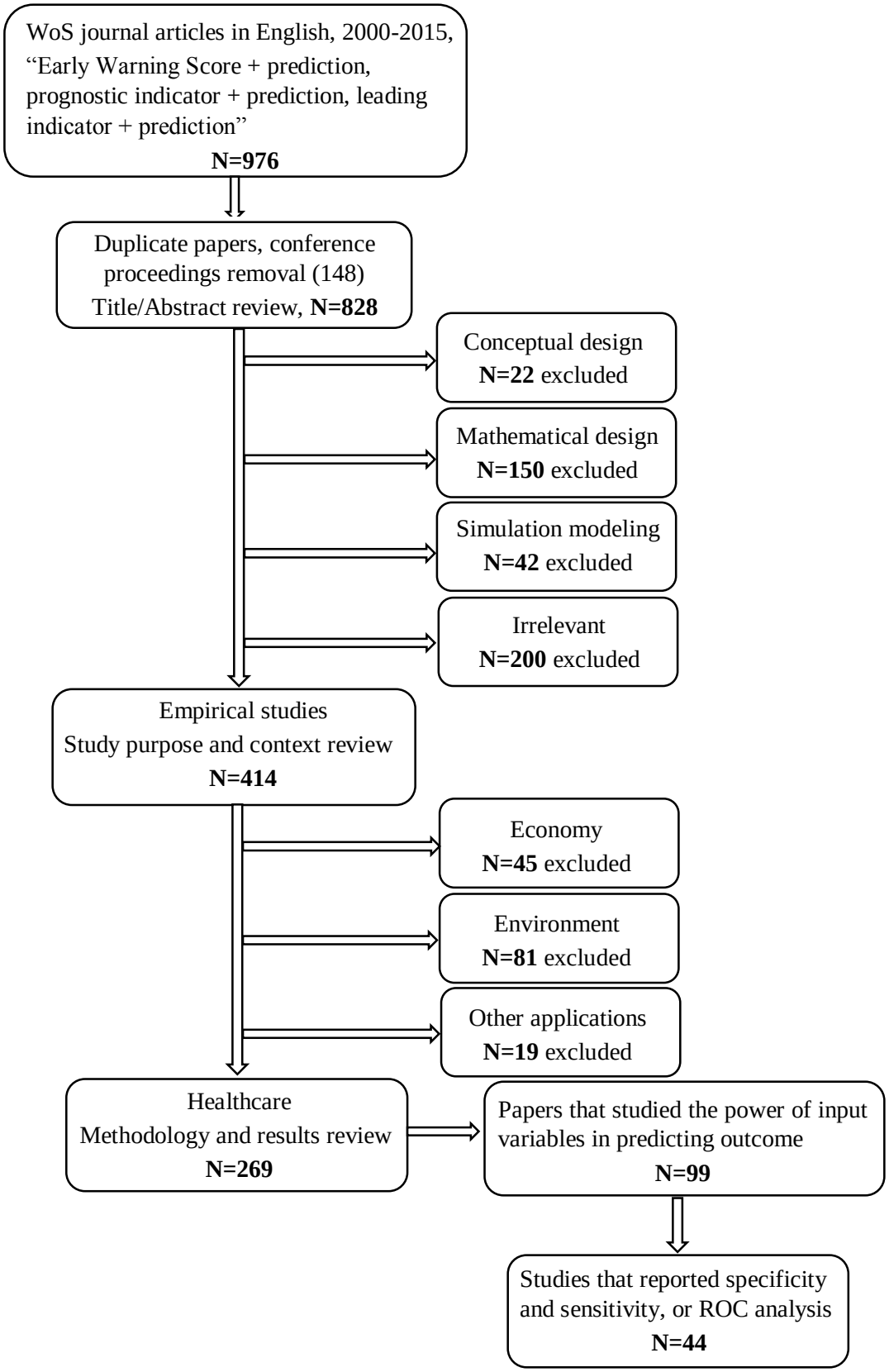

Figure 4.1 Flow diagram of Study selection process

The list of the final selected papers and their characteristics are presented in Table 4.1. In addition to authors and publication date, the table provides information about their sample of study (sample size and patient group). The table reports risk indicators and the predicted outcome variable for each paper. We categorize risk indicators into three different groups that include *MEWS*, *other type of early warning score*, and *physiological measure*.

Out of 44 papers, 7 papers studied the predictive ability of MEWS in predicting health outcome, 9 studies focused on other early warning scores, and 37 papers studied other physiological measures as outcome predictors. Furthermore, eight papers studied MEWS or other early warning score in addition to a physiological measure. A complete list of the excluded studies is available from the authors upon request.

4.3.4 Data Collection and Data Analysis

The data extracted from each study includes the source and full reference, main topic area, sample size, patient group, risk indicators, outcome measures, sensitivity and specificity measure, or ROC AUC, and overall conclusion on predictability of risk indicators.

Table 4.1 Characteristics of 44 healthcare papers studying the power of risk indicator in predicting outcome variable

Author(s), Date	Sample Characteristics		Design Characteristics	
	Sample Size	Patient Group	Risk Indicators	Outcome Variables
Almaraz *et al.* 2009 [8]	227	Comatose post cardiac arrest	P	Neurological outcome
Alvarez *et al.* 2013 [9]	7,466	General medical and surgical	P, M	Out of ICU cardiopulmonary arrest; acute respiratory compromise; unexpected death

Author(s), Date	Sample Characteristics		Design Characteristics	
	Sample Size	Patient Group	Risk Indicators	Outcome Variables
Andrijevic *et al.* 2014 [10]	101	Community acquired pneumonia (CAP)	P	30-day mortality
Asadollahi *et al.* 2011 [11]	550	General medical and surgical	P	Mortality
Badreldin *et al.* 2013 [12]	5,207	Postoperative cardiac surgical	P	Clinical deterioration
Bhorat *et al.* 2014 [13]	29	Pregnant women with severe gestational diabetes on insulin therapy in the third trimester	E	Adverse prenatal outcome, e.g., death; change in cardiac function in fetus
Bradman *et al.* 2014 [14]	946	ED	P	Need for hospital admission
Cattermole *et al.* 2009 [15]	330	ED	P	Early ICU Admission; death
Churpek *et al.* 2014 [16]	269,999	On the wards	P	Cardiac arrest; ICU transfer; death
Cildir *et al.* 2013 [17]	230	ED with sepsis	E	Mortality in sepsis
Clements *et al.* 2007 [18]	43,611	13 common surgical procedures	E	Surgical site infection outcomes
Cuthbertson *et al.* 2006 [19]	140	General surgical high dependent	P	ICU admission
Fullerton *et al.* 2012 [20]	3,504	ED	M, P	Occurrence of an adverse event with 24hr of admission
Gaebel *et al.* 2012 [21]	339	Schizophrenia outpatients	P	Occurrence of prodromal symptoms
Geier *et al.* 2013 [22]	151	ED suspected sepsis	P, M, E	In-hospital mortality
Ghanem-Zoubi *et al.* 2011 [23]	1,072	Sepsis	M, E	In-hospital mortality
Goodell *et al.* 2006 [24]	104	Women at the time of their initial definitive surgery for ovarian cancer	P	Ovarian cancer survival
Hauzman *et al.* 2004 [25]	150	IVF pregnancy	P	Outcome of IVF pregnancies
Hayashida *et al.* 2014 [26]	495	Post cardiac arrest registry	P	Neurological outcome
Hendriks *et al.* 2004 [27]	109		P	Ovarian hyperresponse
Hu *et al.* 2012 [28]	From 2005-2010	Chickenpox infectious	P	Infectious disease
Ito *et al.* 2005 [29]	197	Gastric cancer	P	Peritoneal relapse

Author(s), Date	Sample Characteristics		Design Characteristics	
	Sample Size	Patient Group	Risk Indicators	Outcome Variables
Jalali *et al.* 2014 [30]	104	Brain injury	E	Mortality
Jalkanen *et al.* 2013 [31]		Operative and non-operative ventilated	P	Mortality; acute lung injury; sepsis; renal replacement
Kim *et al.* 2010 [32]	103	Non-muscle invasive bladder cancer	P	Disease progression
Lam *et al.* 2009 [33]	122	Laryngosocopically confirmed	M, P	Airways intervention; ICU admission
Lee *et al.* 2014 [34]	1,049	ICU	P	Organ failure status in ICU patients; mortality
Liu *et al.* 2014 [35]	546	Chest pain in ED	M, P	Acute cardiac complications
Loekito *et al.* 2013 [36]	42,701	Admitted for more than 24 hours	P	Imminent death
Merle *et al.* 2009 [37]	25	Acute liver failure	P	Need of transplantation
Mishra *et al.* 2007 [38]	76	Cirrhosis (chronic liver disease)	E, P	Death
Mizuguchi *et al.* 2011 [39]	105	Hepatocellular carcinoma underwent hepatectomy	P	Severity of liver disease; survival after hepatectomy
Oku *et al.* 2002 [40]	40	Rectal cancer	P	Rectal cancer
Ong *et al.* 2012 [41]	925	Critically ill in ED	M, P	Cardiac arrest
Purkayastha *et al.* 2007 [42]	529	Rectal cancer	P	Preoperative evaluation of circumferential margin involvement
Rozen *et al.* 2013 [43]	28	Implantable cardiac defibrillators (ICD)	P	Imminent ventricular arrhythmias
Smith *et al.* 2012 [44]	572	General and trauma surgery wards	E	Major adverse clinical events
Suppiah *et al.* 2013 [45]	146		P	Acute pancreatitis
Ugajin *et al.* 2014 [46]	445	Aged>64 and hospitalized because of nursing home acquired pneumonia	P	Nursing home acquired pneumonia; community acquired pneumonia
Umscheid *et al.* 2015 [47]	4,575	Adult non-ICU	E	Composite of ICU transfer, RRT call, or death

Author(s), Date	Sample Characteristics		Design Characteristics	
	Sample Size	Patient Group	Risk Indicators	Outcome Variables
Vis *et al.* 2002 [48]	81	Men with prostate cancer	P	Tumor features
Vladimirov *et al.* 2004 [49]	17	Undergoing IVF treatment	P	Poor ovarian response
Yildirim *et al.* 2007 [50]	704	With positive axillary lymph nodes (PN)	P	Distant recurrence; locoregional recurrence; disease-free survival
Yu *et al.* 2014 [51]	328	Non-ICU wards	P	Clinical deterioration (ICU admission, or death)

Note: M: MEWS, E: Other Early Warning Scores, P: Physiological measures

4.3.5 Measures of Predictive Power of Early Warnings

There are different measures of evaluating the predictive power of early warning systems such as R^2 value [52], and likelihood ratios [53]. Here, our main variables of interest are binary variables, and we focus on measures that are commonly used to evaluate the predictive power of screening tests, i.e., the sensitivity, specificity, and ROC AUC of the tests.

Sensitivity and specificity measure the predictive power of a test that classifies a set into two groups (e.g., sick vs. healthy) [54]. In clinical epidemiology and medical testing, they are used to determine whether a patient has a certain disease. Sensitivity or true positive rate refers to the percentage of patients who are correctly identified as having the disease. Specificity or true negative rate is the percentage of healthy group who are correctly identified as not having the disease. So if in our sample of study, there are 300 people, and 100 of them are sick, and our test can find 80 sick people, the test's sensitivity is 80/100 = 0.8. If the same test misdiagnose 50 healthy people as being sick, it means that it has correctly diagnosed (200-50) people as healthy, so the specificity of the test is (200-50)/200 = 0.75. In general terms, we can write:

$$Sensitivity = \frac{number\ of\ true\ positives}{(number\ of\ true\ positives + number\ of\ false\ negatives)}$$

$$Specificity = \frac{number\ of\ true\ negatives}{(number\ of\ true\ negatives + number\ of\ false\ positives)}$$

For any clinical test, there is a trade-off between sensitivity and specificity. This means that in a given accuracy, if we increases sensitivity, specificity decreases, and vice versa. For example, in a hospital setting in which one is testing for a potential cancer risk factors in a patient, physicians may consider low-risk signs like the variation of body temperature (low specificity), to reduce the risk of missing signs that do pose a threat to a person's health status (high sensitivity). Although temperature could be an important sign in identifying other types of diseases but in order to diagnose a cancer physicians may conduct more specific tests. In the simple example that we provided above, if the hospital becomes more conservative about risk indicators, they can potentially find more than 80 people as true-sick people, but in the expense of misdiagnosing more than 50 healthy people. If each of these numbers increases by 20%, sensitivity will change from 0.8 to 96/100=0.96, and specificity will change from 0.75 to (200-60)/200=0.7. In this simple numerical example, sensitivity increases in the expense of losing specificity.

The trade-off between sensitivity and specificity can be illustrated graphically by the ROC curve. The curve is a fundamental tool for evaluation of diagnostic tests and represents the "accuracy" of a test type in which the x-axis is [1-specificity] (false positive rate) and the y-axis is sensitivity [55]. ROC AUC is the area under the ROC curve and is a measure that shows how well the test can discriminate between sick patients versus healthy group. Mathematically, it is the cumulative distribution function of the probability of true positive rate (or detection) in the y-axis over the cumulative distribution function of the false positive rate in x-axis. Figure 4.2 represents the ROC

curve of three sample predictors for different sensitivity and specificity thresholds. The best possible condition is the point in coordinate (0,100), representing 100% sensitivity (no false negatives) and specificity (no false positives), but obviously almost all curves don't pass that point.

The diagonal line ($y = x$), also known as no-discrimination line, in the graph represents a test that is equivalent to a random guess. Specifically, the point that corresponds to (0.5, 0.5) on this curve (line) is equivalent to flipping a coin. The accuracy of the tests increases as we move to the curves on the left, shown by the arrow in Figure 4.2.

After clarifying which curve belongs to a specific test, decision makers have to pick a threshold to discriminate between positive and negative cases. In simple words, based on Figure 4.2, on any specific curve, we can pick a point that corresponds to a threshold. For example, if the yellow curve represents the accuracy of our test, point *A* is a potential threshold which corresponds to sensitivity of 38% and specificity of 85%. If we move *A* upward, on its current ROC curve, sensitivity will increase but specificity will decrease. Decision analysis, first figure out the ROC curve, and then pick the threshold (e.g., point *A*) that maximizes their payoff function.

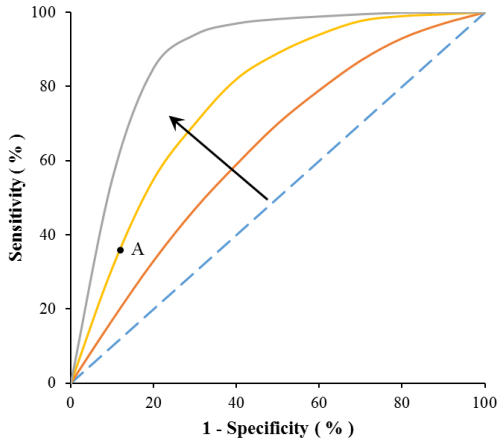

Figure 4.2 ROC curve of three sample predictors

4.4 Results

We categorize the outcome variables into four different groups of *adverse outcome*, *care intensity*, *disease prediction*, and *survival*. Table 4.2 depicts the number of papers in each category as well as the classification criteria. Out of 44 paper, 20 studied adverse outcomes such as mortality, cardiac arrest, and surgical site infection; 5 papers looked at care intensity (e.g., ICU admission, rapid response team call, or need for hospital admission) that a patient may need during hospitalization; 16 papers studied the predictability of clinical conditions (disease prediction) such as neurological outcome or organ failure status based on different risk indicators; 3 papers studied survival predictions (for example for cancer cases) based on early risk indicators. Furthermore, based on the format of their reports, we can group the studies into the ones that report sensitivity and specificity of their tests, and ones that report the ROC AUC.

Figures 4.3, 4.4, and 4.5 depict the distributions of sensitivity, specificity, and ROC AUC in the reviewed papers. Using the results presented in distribution analysis, the average sensitivity is 68.4%, the average specificity is 78.4%, and the average ROC AUC is 73.1%. The results show that ROC AUC data are less varied ($\sigma = 8.6\%$) than sensitivity ($\sigma = 20.6\%$), and specificity ($\sigma = 16.7\%$). Figures 4.3 and 4.4 show that 50% of papers reported a sensitivity of at least 73% and specificity of 82%. The minimum of sensitivity and specificity were 11% and 30.4%, respectively.

Table 4.2 Categorizing health outcome variables: number of papers and definitions based on findings from 44 included studies

Outcome category	No. of studies	Classification criteria
Adverse outcome	20	In-hospital mortality
		30-day mortality
		Cardiac arrest
		Surgical site infection
		Disease progression
Care intensity	5	ICU admission
		Rapid response team call
		Need for hospital admission for ED patients
Disease prediction	16	Neurological outcome
		Outcome of IVF pregnancies
		Organ failure status
		Acute cardiac complications
		Severity of disease
		Cancer
		Poor ovarian response
		Need for transplantation
		Pneumonia
		Sepsis
Survival	3	Cancer survival
		Survival after hepatectomy
		Disease free survival

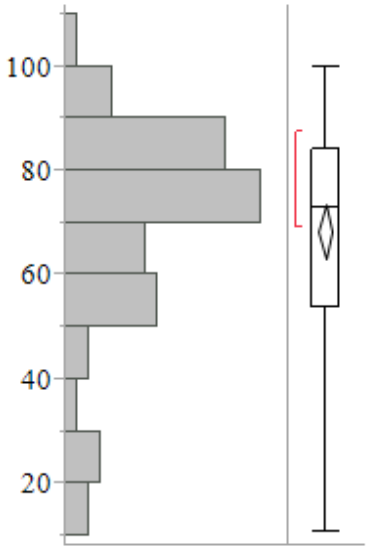

Figure 4.3 Distribution of sensitivity

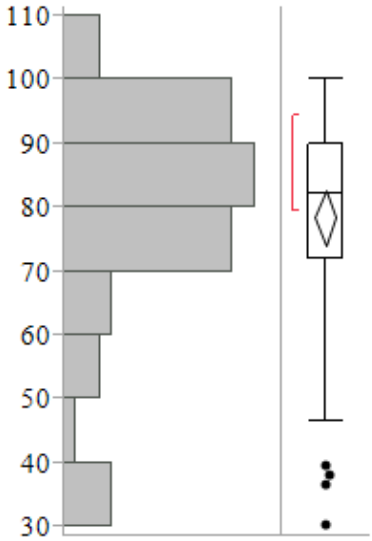

Figure 4.4 Distribution of specificity

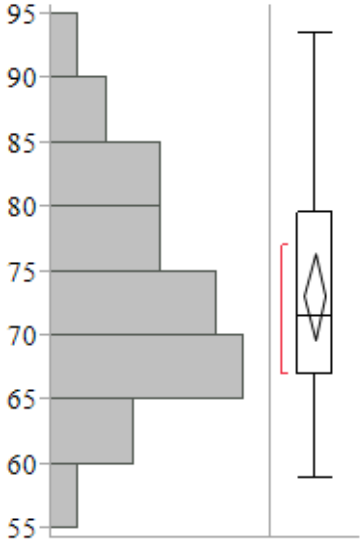

Figure 4.5 Distribution of ROC AUC

Figures 4.6 and 4.7 demonstrate the distributional characteristics of sensitivity and specificity, respectively for outcome categories. Figure 4.6 shows that overall, papers reported relatively higher sensitivity values for predicting survival where 50% of the papers found a sensitivity value greater than 78%. On the other hand, the graph suggests that papers report quite different sensitivity values for predicting adverse outcome and diseases. The top 50% of papers reported at least 74.4% sensitivity for predicting adverse outcome, and 72% for disease prediction. For the care intensity prediction, papers found very similar sensitivity values for the most positive quartile group, and sensitivities were varied among the least positive quartile group. The median of sensitivity for predicting care intensity is 72%.

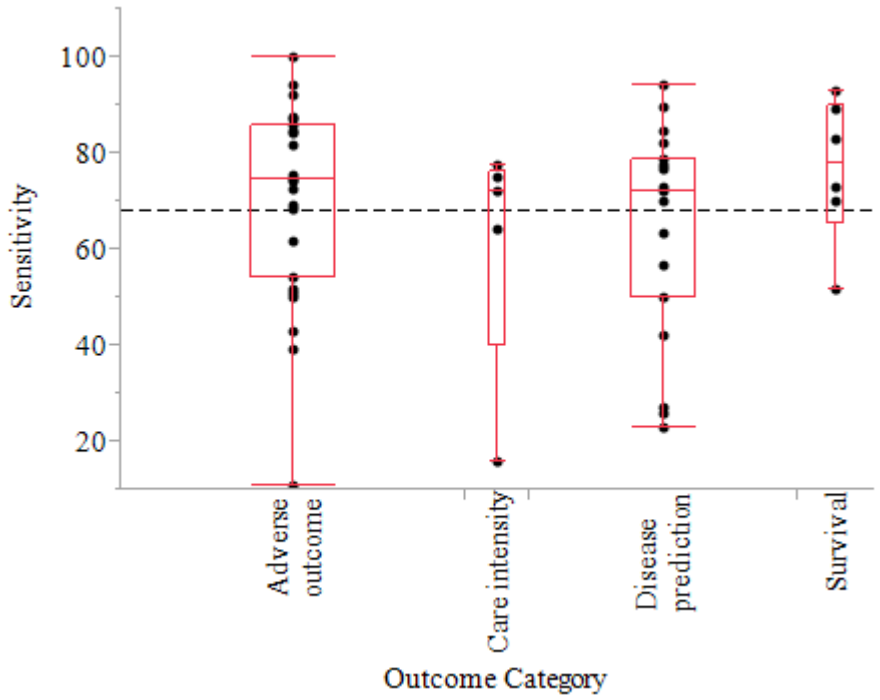

Figure 4.6 Reported sensitivity (in percentage) by health outcome category

Figure 4.7 shows that the overall average of reported specificities (i.e., 79%) is higher than the overall average of sensitivities (i.e., 64%). Overall, papers reported relatively higher specificity values for predicting care intensity where 50% of the papers found a specificity value greater than 89%, and all papers reported a specificity of at least 72%. Moreover, this graph shows that papers report quite different specificity values for predicting adverse outcome. The top 50% of papers reported at least 80.4% specificity for adverse outcome prediction. It is notable that papers found very similar specificity values for the most positive quartile group, but specificities were varied among the least positive quartile group. For disease prediction, 50% of studies reported at least 82% specificity, however, the variation is high among the lower 10%.

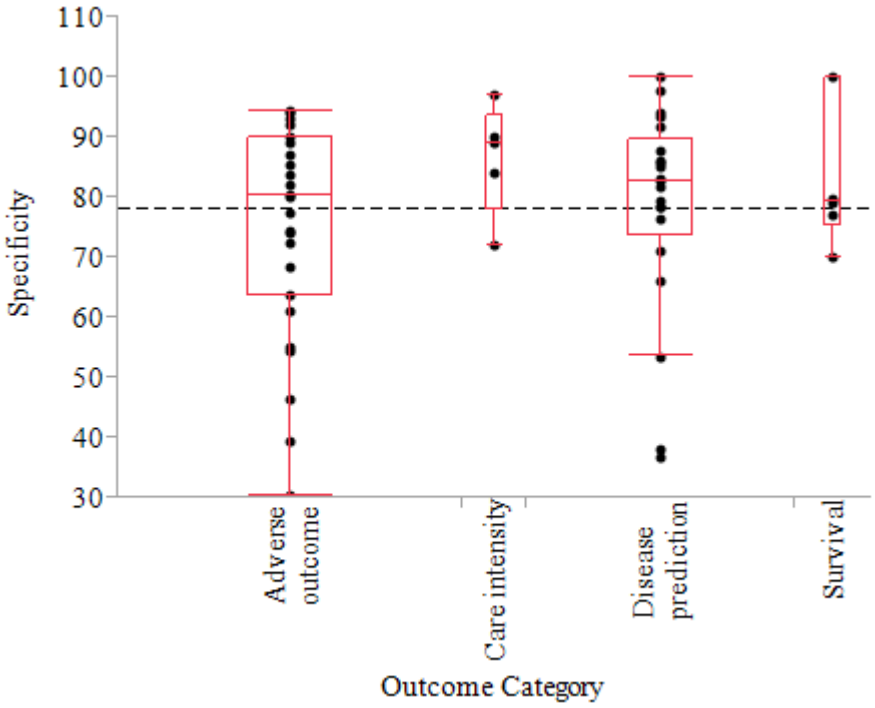

Figure 4.7 Analysis of specificity (in percentage) by health outcome category

Figure 4.8 represents the distribution of the results for the ROC AUC by outcome category for papers that reported ROC AUC metric in their results. The results show that the ROC AUC values for predicting survival is lower than overall mean (73%). 50% of papers found an ROC AUC greater than 76.7% for predicting adverse outcome, however, the variation between values is high. For disease prediction, papers report a relatively low ROC AUC values where the median is 69.5%. Finally, there was only one study that used ROC AUC measure for care intensity prediction with the value of 71%.

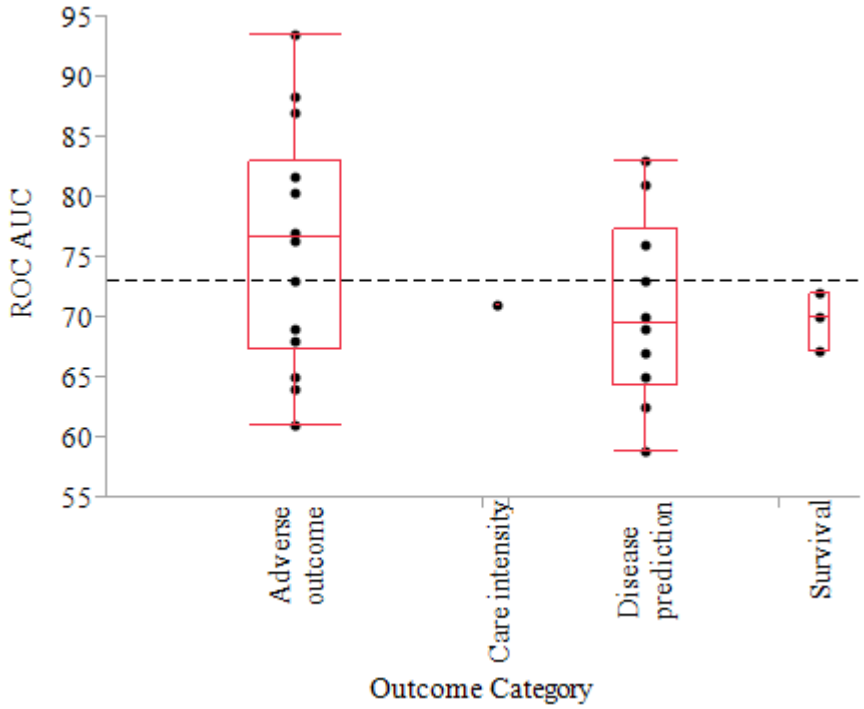

Figure 4.8 Analysis of ROC AUC (in percentage) by health outcome category

We also categorized the outcome variables to two setting classes, i.e., hospital and non-hospital. If outcome prediction is related to hospital processes, actions, or triage tools then it is classified as a hospital setting. The ROC AUC values are calculated for all papers based on their sensitivity and specificity measures. We assume the ROC curve function to be $y = x^{\alpha}, \alpha \leq 1$, where y is sensitivity and x is (1-specificity). Then ROC AUC can be calculated as follows:

$$ROC\ AUC = \int_0^1 x^{\alpha} dx = \frac{1}{\alpha + 1}, \quad \alpha \leq 1$$

$$ROC\ AUC = \frac{1}{1 + \frac{\mathrm{Ln}(y)}{\mathrm{Ln}(x)}} = \frac{1}{1 + \frac{\mathrm{Ln}(sensitivity)}{\mathrm{Ln}(1 - specificity)}}$$

Table 4.3 contains information that compare the mean and standard deviation of ROC AUC values in different outcome categories between the two settings. We also categorized adverse outcome to mortality and non-mortality outcomes. The results show that the accuracy of tests that predict non-

mortality adverse outcomes in a non-hospital setting is high. Also, accuracy of the tests that predict care intensity such as ICU admission in a hospital setting is higher than adverse outcome prediction.

Table 4.3 Average of ROC AUC (%) in each outcome category and setting type (standard deviations are in parentheses)

	Healthcare Services	Adverse Outcome			
Setting Type	Care Intensity	Non-mortality		Mortality	
Hospital	79.38 (9.20)	74.80 (12.84)		77.79 (9.50)	
Non-hospital	–	Non-mortality	Disease Prediction	Survival	Mortality
		98.30 (1.50)	77.19 (11.20)	87.28 (12.22)	79.5 (2.80)

While we don't aim at claiming a statistically significant effect as the sample size is small, but looking at the differences in average ROC AUC provides first-order insights, a base for future investigation. For example, it seems that these are better in predicting the care intensity in hospitals that is in-hospital outcomes. These resonates with our findings from the previous paper. Furthermore, the risk measures provide higher average scores from non-hospital outcomes. This might be due to the fact that when a high risk situation is perceived in a hospital, preventive interventions take place.

Figures 4.9, 4.10, and 4.11 represent the results of sensitivity, specificity, and ROC AUC analysis for hospital and non-hospital settings. The median of sensitivity, specificity, and ROC AUC for non-hospital group is higher than hospital group. There is more variation in sensitivity values reported by non-hospital setting papers, however, these studies had more agreement on specificities. Therefore, models of papers that predicted diseases, survival, and adverse outcome for non-hospital setting category are more specific than other group.

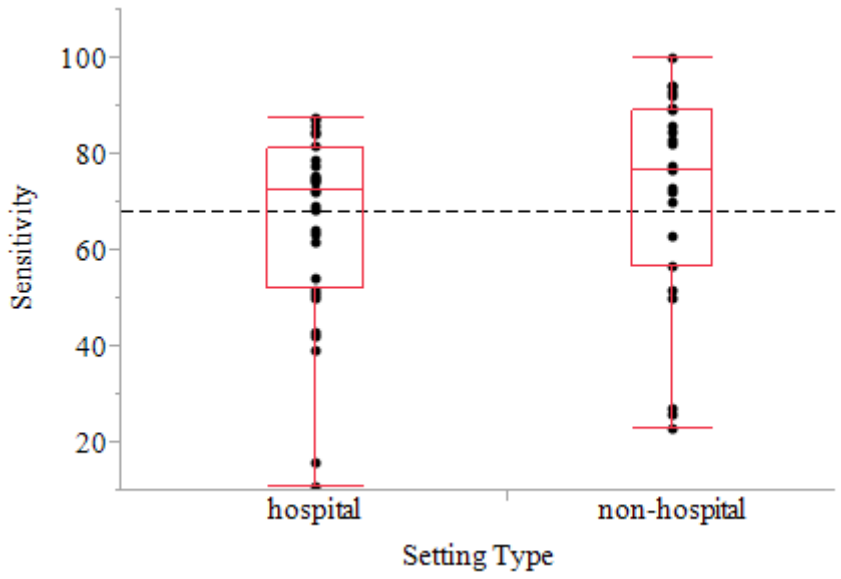

Figure 4.9 Analysis of sensitivity (in percentage) by setting type

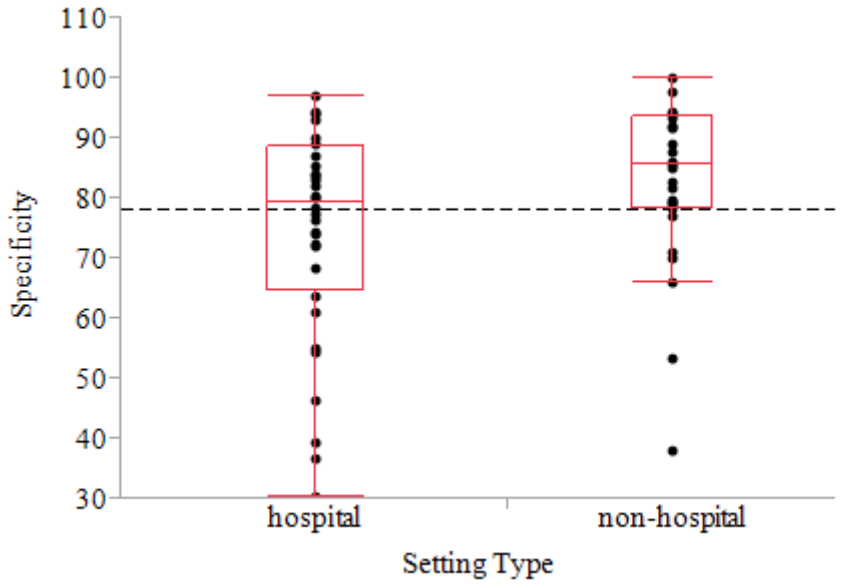

Figure 4.10 Analysis of specificity (in percentage) by setting type

ROC AUC values for hospital setting papers varied between 61% and 93.5%, and for non-hospital setting was between 60% and 81.7%. Figure 4.11 shows that although the maximum ROC AUC value reported for non-hospital papers is lower than hospital papers, but the upper 50% of papers

in non-hospital category reported similar ROC AUC values than hospital category. Overall, there is a large variation in ROC AUC values for hospital setting papers.

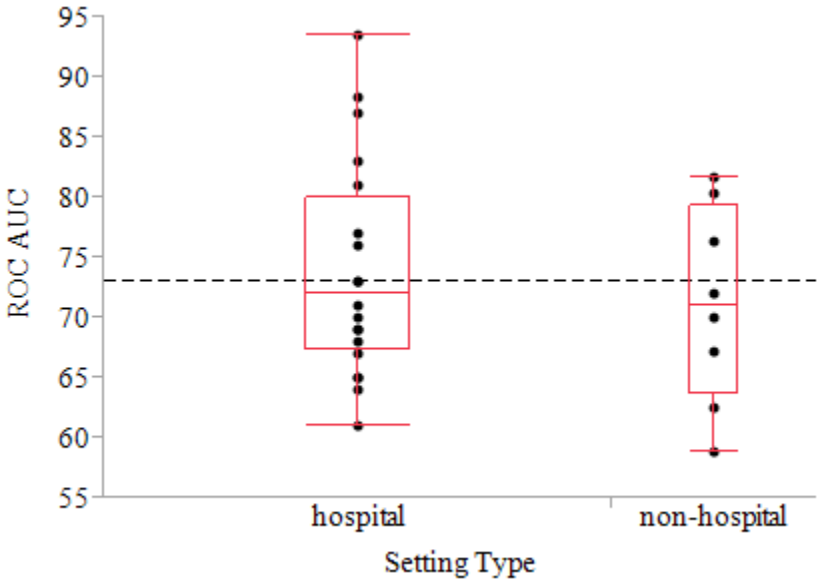

Figure 4.11 Analysis of ROC AUC (in percentage) by setting type

4.5 Discussion

The review has shown that there is a variety of published papers that studied the predictability of health outcome according to risk indicators with little evidence of their agreement on prediction. Although this review only considered published early warning systems that used sensitivity, specificity, and ROC AUC measures, there are many more early risk indicators being used in a variety of healthcare systems, which have not been published or used another measures to evaluate their prediction capabilities.

Our results show that sensitivities were relatively low. This may be due in part to rapidly changing health status, especially in the context of acute diseases, and infrequent and non-standardized risk

measures. Sensitivities could potentially be improved at the cost of decreased specificities, by reducing the early warning trigger threshold [56].

Our results show that early warnings for survival had high specificity and acceptable sensitivities, however this contradicts the fact that the ROC AUC was relatively low. A former risk identification model provides a benchmark at ROC AUC = 77% [57]. Our ROC AUC analysis of survival, from the evaluation of available two papers, indicates that the probability of true survival rate against the false survival rate is not high (ranged from 67.2% to 72%). One paper was concerned about survival prediction for patients with positive axillary lymph nodes[1], and the other paper studied the survival prediction after hepatectomy[2]. Both of these procedures involve a surgery for a type of cancer patients, and it may be that the current early warning systems of these outcome variables developed is not the ideal. Other early warning systems that predict adverse outcome, care intensity, and diseases had more varied sensitivities than specificities and would require further validation in different patient groups and settings.

Considering the mixed evidence in predictive power of warning measures, we offer a dynamic model to help make sense of the variation in results. One potential source of variation in results is in the purpose of using early warning systems. Early warning systems are generally used to identify deteriorating patients based on their health risk status.

Decision makers desire to have a system that would predict health outcome accurately. On the other hand, preventive actions and their effects on decreasing the probability of adverse outcome play an important role in preventability of catastrophic event.

1: Positive axillary lymph node: a lymph node in the area of armpit that cancer has spread and determined by surgical procedure [58].

2: Hepatectomy: liver surgery on patients with liver cancer [59].

Figure 4.12 demonstrates the predictability and preventability of adverse outcome according to risk measure accuracy, attention to risk, human intervention, and risk mitigation. We expect to see more adverse outcomes for high risk patients (in the figure, the bold arrow line from *Risk Status* to *Adverse Outcome*). However, the risk status will result in series of reactions in the healthcare setting which can compensate for effect of initial risks by providing better care for patients. Notably, in Figure 4.12, risks are measured (dependent on how accurate the risk measures are), are perceived by healthcare providers (if they pay enough attention to the risk measures), result in preventive actions (dependent on resource availability), and mitigate the chance of adverse outcome (dependent on how proper the interventions were).

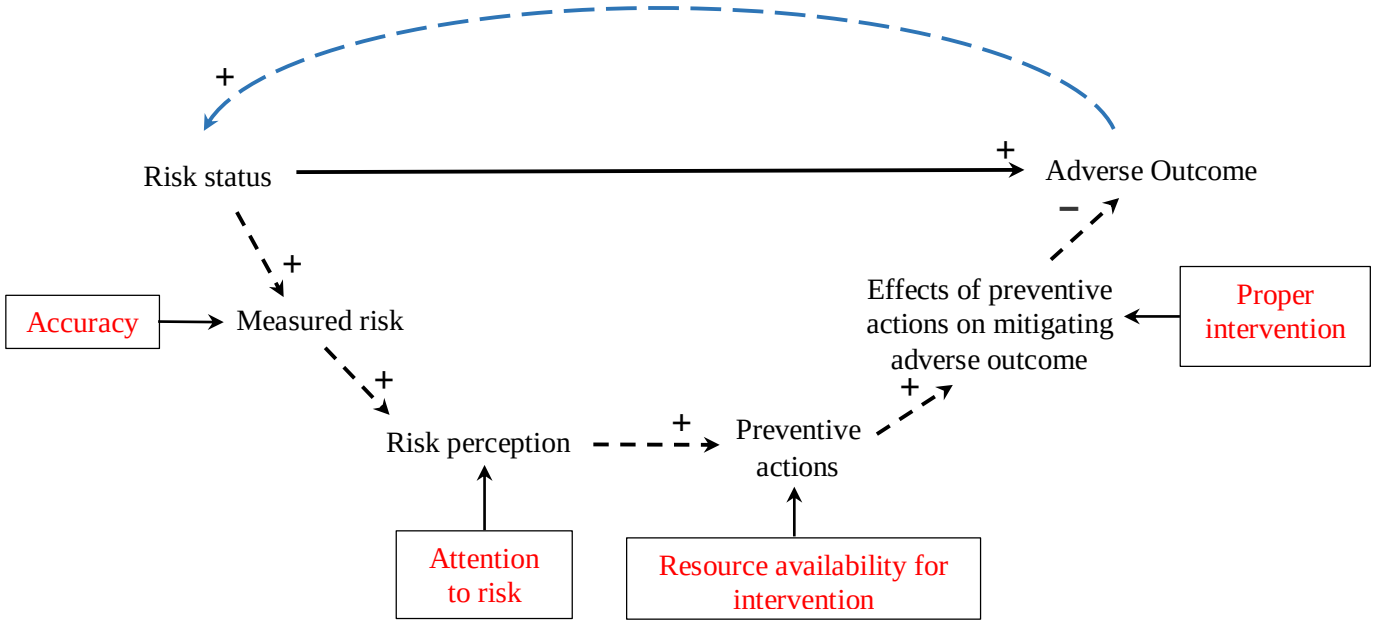

Figure 4.12 Conceptual representation of predictability and preventability of health outcomes in hospital

The overall result is mitigating adverse outcome, compensating and working in the opposite direction of the bold line. As stated, the effect of this line depends on 1) how good the warning system is (in Figure 4.12, Accuracy), 2) if healthcare providers such as physicians use and pay

attention to these measures (in Figure 4.12, Attention to risk), 3) if the resources are available for intervention (in Figure 4.12, Resource availability for intervention), and 4) if the interventions are correctly selected and implemented (in Figure 4.12, Proper intervention).

Since these two effects (paths) can work simultaneously, it may result in confusion on the predictive power of early warning score. At the end, the overall objective is to increase the extent to which early warning systems in healthcare achieve their objectives, i.e., predicting and preventing adverse outcome. Figure 4.13, demonstrate the potential source of conflict.

In the case where outcome is predicted and is happened we conclude a successful prediction but poor prevention. In contrary, we have poor prediction for the system where outcome is not predicted but is happened.

The main source of confusion is the cell with outcome predicted of "Yes", and outcome happened of "No". There are two different reasons to be in this cell. Either we observed a poor prediction, or the system could successfully predict and prevent the outcome should be investigated to achieve an accurate conclusion. Notably, an ideal powerful early warning system not only would predict the adverse outcome successfully but would prevent it too. Finally, a system could successfully predict the outcome where outcome is not predicted and is not happened.

		Outcome Predicted	
		Yes	No
Outcome Happened	Yes	Successful Prediction, Poor Prevention	Poor Prediction
	No	Poor prediction OR Successful prediction, Successful prevention	Successful Prediction

Figure 4.13 Evaluation of early warning systems

4.6 Conclusions

In conclusion, this review finds mixed evidence on the predictive power of early warning or diagnostic scores. We hypothesize that the variation may relate to the purpose of use of these measures, i.e., prediction vs. prevention. We found better prediction for healthcare intensity than outcome, which resonates with our finding in the previous paper. We also hypothesize that poor evidence on predictive power of these measures may relate to *accuracy*, *attention to measures*, and *proper interventions*. We suggest that future research focus on improving accuracy as well as attention and intervention, looking at early warning systems as a part of a "whole system" which influence healthcare response to measured risks.

References

1. Subbe, C., et al., *Validation of a modified Early Warning Score in medical admissions.* Qjm, 2001. **94**(10): p. 521-526.
2. McGaughey, J., et al., *Outreach and Early Warning Systems (EWS) for the prevention of intensive care admission and death of critically ill adult patients on general hospital wards.* Cochrane Database Syst Rev, 2007. **3**.
3. Cuthbertson, B.H., et al., *Can physiological variables and early warning scoring systems allow early recognition of the deteriorating surgical patient?**. Critical care medicine, 2007. **35**(2): p. 402-409.
4. Paterson, R., et al., *Prediction of in-hospital mortality and length of stay using an early warning scoring system: clinical audit.* Clinical Medicine, 2006. **6**(3): p. 281-284.
5. Smith, G.B., et al., *The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death.* Resuscitation, 2013. **84**(4): p. 465-470.
6. Azadeh-Fard, N., N. Ghaffarzadegan, and J.A. Camelio, *Can early risk assessments predict a patient's hospital length of stay and mortality?* 2015.
7. Azadeh-Fard, N., et al., *Risk assessment of occupational injuries using Accident Severity Grade.* Safety science, 2015. **76**: p. 160-167.
8. Almaraz, A., et al., *Serum Neuron Specific Enolase to Predict Neurological Outcome After Cardiopulmonary Resuscitation A Critically Appraised Topic.* Neurologist, 2009. **15**(1): p. 44-48.
9. Alvarez, C.A., et al., *Predicting out of intensive care unit cardiopulmonary arrest or death using electronic medical record data.* Bmc Medical Informatics and Decision Making, 2013. **13**.
10. Andrijevic, I., et al., *Interleukin-6 and procalcitonin as biomarkers in mortality prediction of hospitalized patients with community acquired pneumonia.* Annals of Thoracic Medicine, 2014. **9**(3): p. 162-167.
11. Asadollahi, K., et al., *Prediction of hospital mortality from admission laboratory data and patient age: A simple model.* Emergency Medicine Australasia, 2011. **23**(3): p. 354-363.
12. Badreldin, A.M.A., et al., *Rapid clinical evaluation: an early warning cardiac surgical scoring system for hand-held digital devices.* European Journal of Cardio-Thoracic Surgery, 2013. **44**(6): p. 992-998.
13. Bhorat, I.E., et al., *Use of the myocardial performance index as a prognostic indicator of adverse fetal outcome in poorly controlled gestational diabetic pregnancies.* Prenatal Diagnosis, 2014. **34**(13): p. 1301-1306.
14. Bradman, K., M. Borland, and E. Pascoe, *Predicting patient disposition in a paediatric emergency department.* Journal of Paediatrics and Child Health, 2014. **50**(10): p. E39-E44.
15. Cattermole, G.N., et al., *Derivation of a prognostic score for identifying critically ill patients in an emergency department resuscitation room.* Resuscitation, 2009. **80**(9): p. 1000-1005.
16. Churpek, M.M., et al., *Multicenter Development and Validation of a Risk Stratification Tool for Ward Patients.* American Journal of Respiratory and Critical Care Medicine, 2014. **190**(6): p. 649-655.

17. Cildir, E., et al., *Evaluation of the modified MEDS, MEWS score and Charlson comorbidity index in patients with community acquired sepsis in the emergency department.* Internal and Emergency Medicine, 2013. **8**(3): p. 255-260.
18. Clements, A.C.A., et al., *Risk stratification for surgical site infections in Australia: evaluation of the US National Nosocomial Infection Surveillance risk index.* Journal of Hospital Infection, 2007. **66**(2): p. 148-155.
19. Cuthbertson, B.H., et al., *Can physiological variables and early warning scoring systems allow early recognition of the deteriorating surgical patient?* Critical Care Medicine, 2006. **35**(2): p. 402-409.
20. Fullerton, J.N., et al., *Is the Modified Early Warning Score (MEWS) superior to clinician judgement in detecting critical illness in the pre-hospital environment?* Resuscitation, 2012. **83**(5): p. 557-562.
21. Gaebel, W. and M. Riesbeck, *Revisiting the relapse predictive validity of prodromal symptoms in schizophrenia.* Schizophrenia Research, 2007. **95**(1-3): p. 19-29.
22. Geier, F., et al., *Severity illness scoring systems for early identification and prediction of in-hospital mortality in patients with suspected sepsis presenting to the emergency department.* Wiener Klinische Wochenschrift, 2013. **125**(17-18): p. 508-515.
23. Ghanem-Zoubi, N.O., et al., *Assessment of disease-severity scoring systems for patients with sepsis in general internal medicine departments.* Critical Care, 2011. **15**(2).
24. Goodell, V., et al., *Antibody immunity to the p53 oncogenic protein is a prognostic indicator in ovarian cancer.* Journal of Clinical Oncology, 2006. **24**(5): p. 762-768.
25. Hauzman, E., et al., *Use of serum inhibin A and human chorionic gonadotropin measurements to predict the outcome of in vitro fertilization pregnancies.* Fertility and Sterility, 2004. **81**(1): p. 66-72.
26. Hayashida, K., et al., *Estimated cerebral oxyhemoglobin as a useful indicator of neuroprotection in patients with post-cardiac arrest syndrome: a prospective, multicenter observational study.* Critical Care, 2014. **18**(4): p. 500-500.
27. Hendriks, D.J., et al., *Use of stimulated serum estradiol measurements for the prediction of hyperresponse to ovarian stimulation in in vitro fertilization (IVF).* Journal of Assisted Reproduction and Genetics, 2004. **21**(3): p. 65-72.
28. Hu, Q., et al., *Early warning CUSUM plans for surveillance of infectious diseases in Wuhan, China.* African Journal of Microbiology Research, 2012. **6**(23): p. 4989-4992.
29. Ito, S., et al., *Prospective validation of quantitative CEA mRNA detection in peritoneal washes in gastric carcinoma patients.* British Journal of Cancer, 2005. **93**(9): p. 986-992.
30. Jalali, R. and M. Rezaei, *A comparison of the glasgow coma scale score with full outline of unresponsiveness scale to predict patients' traumatic brain injury outcomes in intensive care units.* Critical care research and practice, 2014. **2014**: p. 289803-289803.
31. Jalkanen, V., et al., *SuPAR and PAI-1 in critically ill, mechanically ventilated patients.* Intensive Care Medicine, 2013. **39**(3): p. 489-496.
32. Kim, S.-K., et al., *Identification of S100A8-correlated genes for prediction of disease progression in non-muscle invasive bladder cancer.* Bmc Cancer, 2010. **10**.
33. Lam, P.K., et al., *Adult acute epiglottitis: predictors for airway intervention and intensive care unit admission.* Hong Kong Journal of Emergency Medicine, 2009. **16**(4): p. 198-207.
34. Lee, K.S., et al., *Consideration of additional factors in Sequential Organ Failure Assessment score.* Journal of Critical Care, 2014. **29**(1).

35. Liu, N., et al., *Risk Scoring for Prediction of Acute Cardiac Complications from Imbalanced Clinical Data.* Ieee Journal of Biomedical and Health Informatics, 2014. **18**(6): p. 1894-1902.
36. Loekito, E., et al., *Common laboratory tests predict imminent death in ward patients.* Resuscitation, 2013. **84**(3): p. 280-285.
37. Merle, U., et al., *Sensitivity and specificity of plasma disappearance rate of indocyanine green as a prognostic indicator in acute liver failure.* Bmc Gastroenterology, 2009. **9**.
38. Mishra, P., et al., *Applicability of MELD as a short-term prognostic indicator in patients with chronic liver disease: An Indian experience.* Journal of Gastroenterology and Hepatology, 2007. **22**(8): p. 1232-1235.
39. Mizuguchi, T., et al., *Prognostic Impact of Preoperative the Branched-Chain Amino Acid to the Tyrosine Ratio in Hepatocellular Carcinoma Patients after Initial Hepatectomy.* Journal of Gastrointestinal Surgery, 2011. **15**(8): p. 1433-1439.
40. Oku, S., et al., *FDG-PET after radiotherapy is a good prognostic indicator of rectal cancer.* Annals of Nuclear Medicine, 2002. **16**(6): p. 409-416.
41. Ong, M.E.H., et al., *Prediction of cardiac arrest in critically ill patients presenting to the emergency department using a machine learning score incorporating heart rate variability compared with the modified early warning score.* Critical Care, 2012. **16**(3).
42. Purkayastha, S., et al., *Diagnostic precision of magnetic resonance imaging for preoperative prediction of the circumferential margin involvement in patients with rectal cancer.* Colorectal Disease, 2007. **9**(5): p. 402-411.
43. Rozen, G., et al., *Multipole Analysis of Heart Rate Variability as a Predictor of Imminent Ventricular Arrhythmias in ICD Patients.* Pace-Pacing and Clinical Electrophysiology, 2013. **36**(11): p. 1342-1347.
44. Smith, T., et al., *Accuracy of an expanded early warning score for patients in general and trauma surgery wards.* British Journal of Surgery, 2012. **99**(2): p. 192-197.
45. Suppiah, A., et al., *The Prognostic Value of the Neutrophil-Lymphocyte Ratio (NLR) in Acute Pancreatitis: Identification of an Optimal NLR.* Journal of Gastrointestinal Surgery, 2013. **17**(4): p. 675-681.
46. Ugajin, M., et al., *Prognostic value of severity indicators of nursing-home-acquired pneumonia versus community-acquired pneumonia in elderly patients.* Clinical Interventions in Aging, 2014. **9**: p. 267-274.
47. Umscheid, C.A., et al., *Development, Implementation, and Impact of an Automated Early Warning and Response System for Sepsis.* Journal of Hospital Medicine, 2015. **10**(1): p. 26-31.
48. Vis, A.N., et al., *Value of tissue markers p27(kip1), MIB-1, and CD44s for the pre-operative prediction of tumour features in screen-detected prostate cancer.* Journal of Pathology, 2002. **197**(2): p. 148-154.
49. Vladimirov, I.K., D.M. Tacheva, and K.B. Kalinov, *Mean ovarian diameter (MOD) as a predictor of poor ovarian response.* Journal of Assisted Reproduction and Genetics, 2004. **21**(3): p. 73-77.
50. Yildirim, E. and U. Berberoglu, *Lymph node ratio is more valuable than level III involvement for prediction of outcome in node-positive breast carcinoma patients.* World Journal of Surgery, 2007. **31**(2): p. 276-289.
51. Yu, S., et al., *Comparison of risk prediction scoring systems for ward patients: a retrospective nested case-control study.* Critical Care, 2014. **18**(3).

52. Chau, P.Y. and P.J.-H. Hu, *Investigating healthcare professionals' decisions to accept telemedicine technology: an empirical test of competing theories.* Information & management, 2002. **39**(4): p. 297-311.
53. Powell, J., et al., *Suicide in psychiatric hospital in-patients Risk factors and their predictive power.* The British Journal of Psychiatry, 2000. **176**(3): p. 266-272.
54. Sackett, D.L., R.B. Haynes, and P. Tugwell, *Clinical epidemiology: a basic science for clinical medicine*. 1985: Little, Brown and Company.
55. Fawcett, T., *An introduction to ROC analysis.* Pattern recognition letters, 2006. **27**(8): p. 861-874.
56. Gao, H., et al., *Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward.* Intensive care medicine, 2007. **33**(4): p. 667-679.
57. Ryan, P.B., et al., *Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership.* Statistics in medicine, 2012. **31**(30): p. 4401-4415.
58. Veronesi, U., et al., *Sentinel lymph node biopsy and axillary dissection in breast cancer: results in a large series.* Journal of the National Cancer Institute, 1999. **91**(4): p. 368-373.
59. Fan, S.-T., et al., *Hepatectomy for hepatocellular carcinoma: toward zero hospital deaths.* Annals of surgery, 1999. **229**(3): p. 322.

5. Conclusions

Risk assessment is a systematic procedure for describing and quantifying the risks associated with hazards or catastrophic events. Risk assessments are very important as they help create awareness of hazards and risks, and determine if existing control measures are adequate or if more should be done. Risk indicators are metrics that are widely used in risk management to indicate how risky an activity is. Among different types of risk indicators, early warning systems are designed to help decision makers predict and be prepared for catastrophic events. Especially, in complex systems where outcomes are often difficult to predict, early warnings can help decision makers manage possible risks and take a proactive approach.

5.1 Risk Assessment of Occupational Injuries

The consideration of additional severity factors improves risk assessment and the estimation of injury severity of occupational injuries. A three dimensional risk assessment matrix allows for the analysis of an incident's degree of preventability, frequency, and severity all at once. A new severity metric that incorporates employee and workplace risk factors, as well as a new three-dimensional risk assessment matrix based on residual risk scores was proposed, in an effort to include more information in the injury risk assessment process. Using the new Accident Severity Grade (ASG) in the proposed three-dimensional risk assessment matrix, industries can quantify an accident's severity immediately after the incident occurs. This approach allows for real-time monitoring of severity, which will lead to more timely implementation of hazard controls that are specifically targeted toward the most severe accident types and their causes. In order for the

proposed system to be as effective as possible, it will be necessary for industries to collect more detailed data related to worker and workplace factors.

5.2 Predictability of Mortality and Hospital LOS

Risk evaluation and control have been important components of healthcare operations. Ideally, providers would like to predict health risks early during hospital admission and take controlling actions. Different methods and techniques have been developed for this purpose, one of which is the early warning system.

We analyzed the predictability of hospital LOS and in-hospital mortality by using MEWS, vital signs, demographic data, and physicians' subjective assessments. In addition, we analyzed the effect of patients' MEWS on physicians' subjective assessments and the effects of those assessments on LOS and mortality rate. We showed that MEWS is not a good predictor of hospital LOS and mortality. Physicians' subjective assessments of patients (i.e., severity level and mortality risk) are better predictors of both outcome (mortality) and LOS. Furthermore, unobservable physicians' characteristics (represented as fixed-effect controls) are very strong predictors of LOS, which means that patients of specific physicians may stay longer in the hospital. Moreover, how humans (physician) react and make judgments are better predictors of both LOS and outcome than patient physiological measures (MEWS). Overall, subjective assessment in our context were better predictors of the process (represented in LOS), and were inversely predictors of death, while MEWS as an objective measure was not associated with LOS or death.

5.3 Analyzing the Predictive Power of Healthcare Early Warning Systems

Early warning systems have been widely used in healthcare to predict adverse outcome. The available literature was systematically reviewed to assess the predictive power of early warning systems and prognostic risk indicators in predicting different outcomes in health. Inclusion criteria were original empirical studies that assessed prediction tests by reporting *sensitivity* and *specificity*, or area under the receiver operating characteristic curve (*ROC AUC*). Mixed evidence was found on the predictive power of early warning or diagnostic scores. We hypothesized that the variation may relate to the purpose and use of these measures, i.e., prediction vs. prevention. It is also hypothesized that poor evidence on predictive power of these measures may relate to accuracy, attention to measures, and proper interventions. We suggested that future research should focus on improving all these measures, looking at early warning systems as a part of a "whole system" which influence healthcare response to measured risks.

5.4 Summary

In summary, a new objective risk indicator for evaluating the risk of occupational injuries was introduced. Moreover, a risk indicator in healthcare systems, i.e., MEWS, was studied in compare to physician's subjective risk measures to predict the health outcome in hospital. Finally, the predictive power of specific risk indicators, i.e., early warning systems, in predicting risky events was examined in healthcare.

Appendix A: A Simulation Model of Health Risk Assessment

We developed a dynamic, stochastic, simulation model of patients admitted to hospital to examine lasting effects of their initial risk factors depicted by early warning scores. A patient level model is designed to simulate hospital LOS and mortality based on patient's initial health risk and physician's subjective assessment of risk. The goal of this was to develop a coherent theory about predictability of outcome by using early warning scores in healthcare systems.

In this simulation model, patients with initial health risks are admitted to hospital. The initial health risk values of patients at time 0 are assumed to be a type of early warning score which are characterized by normally distributed random numbers between 0 and 1 ($\mu = 0.5, \sigma = 0.25$).

At each time period, physicians visit their patients, assign subjective risk scores (also, between 0 and 1), and offer service accordingly. Physician's subjective risk score is a function of true risk score and an error. As in reality, we assume that receiving hospital service usually improves medical state, however, patient's medical state may get worse despite receiving healthcare service. At the end of each time period, if physicians' assessment of risk values for their patients is lower than a specific threshold (discharging threshold), the patient will get discharged from the hospital.

We run this model for 10,000 patients who are randomly assigned to 100 physicians over a 30 day time period. The goal of the model is to examine how long patients stay in the hospital and what proportion die as a function of their initial health risk and physician's initial assessments. In following we describe the model formulation in more details.

Model Formulation

We want to define a function that describes the actual health risk of patient over time, from the time of hospital admission to discharge. We index a particular patient by i and let the health risk of patient i at time t be $r_{i,t}$. We can write patient's health risk as

$$r_{i,t} = r_{i,t-1} + \Delta r_{i,t}, \qquad 0 \leq r_{i,t} \leq 1, \ \forall i = 1..N, \forall t = 1..T, \tag{A.1}$$

where $\Delta r_{i,t}$ is the rate of change in medical condition at time t. Thus, $\Delta r_{i,t} > 0$ represents a deteriorating condition for patient i at time t. Furthermore, we define a function that characterizes physician's subjective assessment of patient's health risk over time. That is,

$$\bar{r}_{i,t} = r_{i,t} + \varepsilon + b_d, \qquad 0 \leq \bar{r}_{i,t} \leq 1, \ \forall i = 1..N, \forall t = 1..T, \forall d = 1..D, \tag{A.2}$$

where $\bar{r}_{i,t}$ is the value of physician's subjective assessment of health risk for patient i at time t. In this equation, ε is a normally distributed error term ($\varepsilon \sim N(0, \sigma)$) that describes the deviation of subjective risk value from its true value of health risk, and b_d refers to the systematic bias of physician d in her patients' examination. In this equation b_d represents difference in physicians' style and characteristics assumed to be constant during our simulation period.

Hospital service and interventions can change the value of the risk defined in Eq. (A.1). Mathematically the following difference equation, Eq. (A.3), represents change in $r_{i,t}$:

$$\Delta r_{i,t} = K.f\left(r_{i,t-1}, S_{i,t}\right), \qquad \forall i = 1..N, \forall t = 1..T, \tag{A.3}$$

where $S_{i,t}$ refers to the service that patient i receives at time t. It is notable that, $\Delta r_{i,t}$ will increase as $r_{i,t-1}$ increases, representing cascading patterns of health in critical conditions, but $\Delta r_{i,t}$ will decrease as $S_{i,t}$ increases, i.e., $\frac{\partial}{\partial r} f > 0, \frac{\partial}{\partial S} f < 0$. In Eq.(A.3), K is a constant parameter

representing all other factors and assumed to be the same for all patients. Let r_d represent the discharging threshold, so the risk value below r_d indicates that patients are safe and can be discharged from hospital. For the purpose of parsimony, we assume the following simple function for $f(r_{i,t-1}, S_{i,t})$:

$$f(r_{i,t-1}, S_{i,t}) = (r_{i,t-1} - r_d) - S_{i,t} \ . \tag{A.4}$$

The equation has two terms. The first term represents deteriorating conditions as a function of one's current health risk, i.e., with no health service, sick patients get sicker. The second term represents the service that one receives in hospital. In Eq. (A.4), whether the first term (deteriorating effect) is stronger or the second term (health service) determines if one heals or not. Such services is assumed to be proportional to the level of attention that physicians pay to a patient, $(\bar{r}_{i,t-1} - r_d)$. We formulate health service as:

$$S_{i,t} = K'(\bar{r}_{i,t-1} - r_d), \qquad S_{i,t} \geq 0. \tag{A.5}$$

In this equation, K' is a coefficient that represents proportional power of the second term, and helps us to simulate different healthcare performance functions. As it is notable, higher values of K' depict better health service quality. While as shown in Eq. (A.4), f is assumed to be a simple linear model, our analysis for a range of non-linear conditions for each of the terms provides qualitatively similar results.

By substituting Eq. (A.5) in Eq. (A.4), we determine the change in medical state for each patient over time.

We can calculate LOS for each patient in our model by adding up the time periods that patient have been hospitalized as presented in Eq. (A.6):

$$L_{i,t} = \sum_{t=1}^{T} \Delta L_{i,t}\,,\ \forall i = 1..N, \tag{A.6}$$

where $\Delta L_{i,t}$ is a binary variable that represents whether patient i stayed in the hospital at time period t ($\Delta L_{i,t} = 1$) or not ($\Delta L_{i,t} = 0$). It is worth to note that, the patient stays in the hospital if physician's subjective risk value is greater than the discharge threshold, and if patient is already in the hospital in previous time period (we assume no return to hospital after discharge). This is represented in Eq. (A.7):

$$\Delta L_{i,t} = \begin{cases} 1, & if\ \bar{r}_{i,t} > r_d\ and\ \Delta L_{i,t-1} = 1 \\ 0, & if\ \bar{r}_{i,t} \leq r_d\ \ or\ \Delta L_{i,t-1} = 0 \end{cases}, \qquad \forall i = 1..N, \forall t = 1..T. \tag{A.7}$$

Appendix B: Additional Analysis

Here, we present more information about our data and variables and report additional analyses of LOS prediction. Table B.1 shows the matrix for correlation coefficients between control variables. Table B.2 depicts the results of regression analysis in which we did a ln transformation on dependent variable, as follows:

$$\ln(y) = \beta_0 + \beta_i x_i + \varepsilon_i \tag{B.1}$$

The results in Table B.2 are consistent with Table 3.4, which shows that the findings are robust.

Table B.1 Correlations between variables

	Age	**Weight**	**Height**	**BMI**	**MEWS**	**Temperature**	**Pulse Rate**	**Respiratory**	**SBP**	**DBP**	**SpO2%**	**Severity Level**	**Mortality Risk**	**LOS**	**Mortality**
Age	1.00														
Weight	-0.29	1.00													
Height	-0.15	0.45	1.00												
BMI	-0.26	0.89	0.00	1.00											
MEWS	-0.02	0.01	0.10	-0.04	1.00										
Temperature	0.03	0.02	-0.00	0.02	-0.26	1.00									
Pulse Rate	-0.11	0.02	0.04	0.00	0.48	0.05	1.00								
Respiratory	0.02	0.11	0.08	0.08	0.51	-0.00	0.27	1.00							
SBP	0.13	0.05	-0.03	0.08	-0.13	0.01	-0.14	-0.01	1.00						
DBP	-0.18	0.12	0.13	0.08	-0.04	-0.02	0.10	-0.01	0.55	1.00					
SpO2%	-0.12	-0.06	0.03	-0.08	-0.15	-0.05	-0.14	-0.11	0.03	0.06	1.00				
Severity Level	0.15	-0.01	0.06	-0.05	0.30	-0.00	0.26	0.18	-0.17	-0.18	-0.09	1.00			
Mortality Risk	0.35	-0.10	0.05	-0.14	0.30	-0.00	0.22	0.17	-0.12	-0.14	-0.10	0.73	1.00		
LOS	0.06	-0.01	-0.01	0.00	0.16	0.05	0.15	0.08	-0.09	-0.10	-0.02	0.43	0.36	1.00	
Mortality	0.10	-0.04	0.04	-0.07	0.21	-0.15	0.06	0.12	-0.06	-0.05	-0.13	0.26	0.30	0.01	1.00

Table B.2 Regression analysis for length of stay with logarithm transformation

Source (x_i)	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**
Patient Features						
Age				0.01***		0.004**
				(0.002)		(0.002)
Gender				0.07**		0.07**
				(0.03)		(0.03)
Weight				-0.01		-0.01*
				(0.01)		(0.01)
Height				0.01*		0.02*
				(0.01)		(0.01)
BMI				0.04		0.04**
				(0.02)		(0.02)
MEWS	0.08***	-0.10	0.03	0.03	0.01	0.02
	(0.02)	(0.58)	(0.03)	(0.03)	(0.03)	(0.03)
Vital Signs						
Temperature		0.01**	0.01**	0.01**	0.01	0.003
		(0.01)	(0.01)	(0.01)	(0.01)	(0.006)
Pulse Rate		0.003**	0.01***	0.01***	0.002	0.003**
		(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
Respiratory		0.01	0.01	0.004	-0.002	-0.01
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
SBP		-0.002*	0.001	-0.001	0.00	0.001
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
AVPU		Not Significant	Not Significant	Not Significant	Not Significant	Not Significant
Additional Physiological Measures						
DBP			-0.01***	-0.01**	-0.002	-0.003*
			(0.002)	(0.002)	(0.002)	(0.002)
SpO2%			0.003	0.01	0.01	0.01
			(0.01)	(0.01)	(0.01)	(0.01)

Subjective Assessments						
Severity Level					0.33***	0.31***
					(0.04)	(0.04)
Mortality Risk					0.07**	0.10***
					(0.04)	(0.04)
Physician						Significant
Intercept	1.12***		-0.18	-3.79**	-4.16**	-4.06**
	(0.04)		(1.01)	(1.82)	(1.66)	(1.69)
R^2	0.02	0.03	0.05	0.08	0.23	0.38
R^2 adjusted	0.02	0.03	0.04	0.06	0.22	0.30
Observations	1021	1021	1010	1010	1010	1010

$***p < 0.01$, $**p < 0.05$, $*p < 0.1$

Note: Standard errors are in parentheses.