Autonomy vs. artificial intelligence : studies on healthcare work and analytics

Wang, Le

2021

Wang, L. (2021). Autonomy vs. artificial intelligence : studies on healthcare work and analytics. Doctoral thesis, Nanyang Technological University, Singapore.
<https://hdl.handle.net/10356/146910>

<https://hdl.handle.net/10356/146910>

<https://doi.org/10.32657/10356/146910>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0
International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 22:24:09 SGT

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**AUTONOMY VS. ARTIFICIAL
INTELLIGENCE: STUDIES ON
HEALTHCARE WORK AND ANALYTICS**

WANG LE
NANYANG BUSINESS SCHOOL
2021

**AUTONOMY VS. ARTIFICIAL
INTELLIGENCE: STUDIES ON
HEALTHCARE WORK AND ANALYTICS**

WANG LE

Nanyang Business School

A thesis submitted to the Nanyang Technological
University in partial fulfilment of the requirement for the
degree of Doctor of Philosophy

2021

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

28/01/2021

.....

Date

.....

Wang Le

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate with amendments, changes and improvements as suggested by me as the Supervisor. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

02/02/2021

.....

Date

.....

Goh Kim Huat

Authorship Attribution Statement

This thesis contains material from one paper published in the following peer-reviewed journal in which I am listed as an author.

Chapter 3 is published as Goh, K.H., Wang, L., Yeow, A.Y.K. et al. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. Nature Communications 12, 711 (2021). DOI: <https://doi.org/10.1038/s41467-021-20910-4>

The contributions of the co-authors are as follows:

- I and Prof. Kim Huat Goh were involved in the formal analysis and methodology of the paper.
- Prof. Kim Huat Goh, I and Prof. Adrian Yeow were involved in the writing of the paper.
- Hermione Poh, Li Ke, Joannas Yeow and Dr. Gamliel Tan were involved in data identification, extraction, and curation.
- All authors were involved in the conceptualization of the project.

28/01/2021

.....
Date

.....
Wang Le

ACKNOWLEDGEMENTS

First and foremost, I would like to express my most sincere gratitude to my supervisor, Prof. Goh Kim Huat, for his unwavering support, help and guidance for this thesis and my PhD career. My deepest thank also goes to Prof. Goh for his teaching me the truth of being a scholar and a man. It is my great honour to have him as one of the most important people in my life.

I would also like to thank my co-supervisor, Mo Jiahui, and my main co-author, Adrian Yeow, for their valuable time and effort devoted to my research projects and all the help they provided when I encountered any setbacks in my research and life. Without them, there will not be this thesis.

My thanks also go to course teachers and faculty members at NBS, Ang Soon, Boh Wai Fong, Ben Choi, Damien Joseph, Hyeokkoo Eric Kwon, Sadat Reza, Christina Soh, Sia Siew Kien, and Trevor Yu (Alphabetical order by last name), for their imparting of knowledge and the help for my PhD. Among them, Prof Boh, Choi, and Reza are also my thesis advisory committee members. I am deeply grateful for their constructive suggestions and comments in all the stages of this work.

Last but not least, I am also filled with gratitude towards my friends at NBS and my family members, you are the ones who are always supporting me in spirit.

Table of Contents

ACKNOWLEDGEMENTS	iv
SUMMARY	1
CHAPTER 1 INTRODUCTION	3
CHAPTER 2 LITERATURE REVIEW.....	8
Implementing AI in Clinical Judgement and Decision-Making	8
Limitations of Traditional AI Approaches in Healthcare	10
Physicians' Decision-Making and Automated Clinical Alerts	12
Construal Level Theory (CLT)	17
CHAPTER 3 ARTIFICIAL INTELLIGENCE IN SEPSIS EARLY PREDICTION AND DIAGNOSIS USING UNSTRUCTURED DATA IN HEALTHCARE	20
Introduction	20
Data Sample.....	23
Sepsis Early Risk Assessment (SERA) Algorithm	27
Processing of Clinical Notes and Machine Learning	29
The Performance of SERA Algorithm	37
SERA Algorithm vs. Human Prediction.....	39
Benefits of Unstructured Clinical Text in Early Sepsis Prediction	44
Application of SERA Algorithm in Low Sepsis Prevalence Environment.....	46
Proposed Setup of SERA Algorithm Within the Clinical Environment.....	48
Discussion.....	50

CHAPTER 4 EXPERT JUDGMENTS IN OVERRIDING ARTIFICIAL INTELLIGENCE: A HEALTHCARE CONTEXT	54
Introduction	54
Research Context: Field Study.....	58
Theory and Hypothesis Development	61
Construal of AI Algorithm Alerts and Overriding Behavior	61
Environmental Characteristics and Overriding Behavior	64
Alert Exposure and Overriding Behavior	65
Domain Expertise and Overriding Behavior.....	67
Mitigating “Cry Wolf” Overriding Behavior.....	70
Data Collection and Methodology	71
Key Variable Measurement and Description	72
Controlling for Alternative Explanations	74
Model Specification and Estimation	78
Results	79
Main Results	79
Robustness Checks and Further Analysis.....	82
Discussion and Conclusions	86
Theoretical Contributions	87
Contribution to Medical Practice and EMR design.....	90
Limitations and Future Work.....	92

REFERENCE94

SUMMARY

With the advance and prevalence of artificial intelligence (AI), many believe that the healthcare industry is ripe for AI disruption, and a wide variety of AI technologies have been piloted within healthcare. While AI, computer scientists and medical informatics researchers have extensively researched the roles of AI technologies in assisting clinical judgement and decision making, most initiatives are still in research and development stage and others are facing numerous implementation challenges. This dissertation aims at providing some insights on how to improve the application of AI technologies in healthcare. In the first essay, I developed an artificial intelligence algorithm, SERA algorithm, which uses both structured data and unstructured clinical notes to predict and diagnose sepsis. I test this algorithm with independent, clinical notes and achieve high predictive accuracy 12 hours before the onset of sepsis (AUC 0.94, sensitivity 0.87 and specificity 0.87). I compare the SERA algorithm against physician predictions and show the algorithm's potential to increase the early detection of sepsis by up to 32% and reduce false positives by up to 17%. Mining unstructured clinical notes is shown to improve the algorithm's accuracy compared to using only clinical measures for early warning 12 to 48 hours before the onset of sepsis. In addition, I demonstrated the role human experts play in an increasing algorithmic world of artificial intelligence. Except for improving AI algorithms and artifacts themselves, researchers are also trying to improve the actual adoption and usage of these AI tools to improve the

effectiveness and efficiency of healthcare AI. I touch this aspect by examining physicians' behaviour towards AI-enabled clinical alerts, which is one of the most common AI application in current clinical settings. One of the benefits of EMR systems is their ability to leverage on AI to provide automated clinical alerts, and this leads to the ubiquitous use of automated clinical alerts in clinical settings. The excessive use of automated clinical alerts, however, leads to the excessive dismissal of such alerts by physicians in a phenomenon described as alert fatigue. In the second essay, I tracked the actions of 1,152 different physicians when they encountered automated clinical alerts in a hospital over a period of 22 months. I collected a total of 66,320 instances of automated clinical alerts and examined the physicians' behaviour towards these alerts. This paper posits that the physicians' dismissal of the alerts is due to more than just alert fatigue, and I argue that the psychological distance of the alert encounters impacts the physicians' construal of the alerts, i.e., the way in which (or the process of) people perceive, comprehend, and interpret the alerts, and a high-level construal will result in their corresponding excessive dismissal of the alerts. My findings suggest that the context in which the AI-enabled alerts appear influences physicians' adherence to these alerts, and I examine the boundary conditions that mitigate the biases that cause the excessive dismissal of these alerts.

CHAPTER 1 INTRODUCTION

Clinical decision making is a knowledge intensive work process that focuses on decision making on diagnosis and prognosis of symptoms and signs and the appropriate treatment (Norman 2000; Sox et al. 2013) and is often shaped by a certain level of uncertainty, complexity, and equivocality, albeit at varying levels (Croskerry 2005; Norman 2000). Uncertainty is characterized as a lack of information about either about the signs or the alternative treatments or their consequences, which makes interpreting a situation and making a decision more difficult (Lipshitz and Strauss 1997; Platt and Huettel 2008). Complex situations are characterized by an abundance of elements or variables (Payne 1976). And lastly, equivocality refers to the presence of several simultaneous but divergent interpretations of a decision domain (Weick and Roberts 1993). The increasing availability of healthcare data and the fast development of big data analytics techniques has made the application of AI in healthcare happen. AI can assist clinical decision making by surfacing useful information hidden in the huge amount of clinical data (Murdoch and Detsky 2013), which is helpful to alleviate the problem of uncertainty and complexity using analytical methods. Humans, however, excel in dealing with equivocal problems compared to machines. This is because equivocality often occurs due to the conflicting interests of stakeholders – clinicians, nursing staff, allied health, and radiologists/pharmacists. This transforms decision making from an impartial, objective process (as assumed in an analytical, rational approach) into

an inherently subjective and political process that attempts to fulfil the conflicting needs and objectives of multiple parties.

Many believe that the healthcare industry is ripe for AI disruption. A wide variety of AI technologies have been piloted within healthcare, including: technologies that enable machines to sense, comprehend, act and learn so they can perform administrative and clinical healthcare functions, others are programmed to take over administrative workflow assistance tasks (Poon et al. 2008), automated image diagnosis (Liu et al. 2017), and clinical decision support and early detection (Jones et al. 2016; Murray et al. 2017). With an increasing prevalence of EMR systems in healthcare industry, more clinical judgement and decision making tools are built on EMR data analytics (Agarwal et al. 2010; Romanow et al. 2012). While AI, computer scientists and medical informatics researchers have extensively researched the roles of AI technologies in assisting clinical judgement and decision making, most initiatives are still in research and development stage (Bardhan et al. 2015; Helm et al. 2016; Henry et al. 2015a) and others are facing numerous implementation challenges. Researchers are trying to solve the problem from two perspectives: one is to improve the algorithm or artifacts themselves and the other is to improve the actual adoption and usage of these AI tools.

From the perspective of algorithm or artifacts, several factors may account for their failing to reach people's expectation. First, many medical problems are often equivocal and ill-structured (Meyer et al. 2014). Intuitive approaches possessed by human experts instead of common AI technologies are

deemed to be more appropriate to deal with such problems and early studies have suggested that human beings are reasonably more capable of predicting patient outcomes than AI algorithms (Rocker et al. 2004; Sinuff et al. 2006). Second, unlike typical business applications, the balance between Type 1 (false positive) and Type 2 (false negative) errors for prediction in clinical contexts is complex and have more significant consequences. Usually there is a lower tolerance for wrong identifications for either positive or negative outcomes. For example, while it is acceptable for an algorithm that predicts potential users for targeted online advertisements to err on the “liberal” side such that there are more false positive targets, the same is not true for AI-based clinical systems used to identify patient ailments. Third, because clinical decisions carry a level of risk and have significant consequences, healthcare AI applications, especially those that are operated autonomously, have ethical and legal ramifications for healthcare personnel and institutions (Cohen et al. 2014). Finally, the medical profession often has a relatively robust set of evidence before they are ready to adopt the use of AI in their daily work, especially if the AI technologies are part of their patient work (Timmermans and Mauck 2005).

Considering these challenges, in the first study I adopt a symbiotic view of AI development, where designs of AI should incorporate the complementary strengths of computers and algorithms and human intuition in problem solving (Jacucci et al. 2015; Jain et al. 2018; Kamar et al. 2012; Qian et al. 2011). In this study, I propose a Collective Experts-AI (“CE-AI”) design for clinical judgement and decision-making tools. I argue that for healthcare AI to be

applicable, systems should be designed to simultaneously leverage on both collective experts' strengths and computing capabilities such that computing agents and human clinicians collaboratively access wider knowledge stores and patterns to better diagnose highly uncertain, complex, and equivocal problems in healthcare. Specifically, my study looks at designing a clinical predictive algorithm that leverages on collective experts' insights and judgement.

From the perspective of actual adoption and usage of AI-enabled healthcare tools, I focus on automated clinical alerts in this dissertation. As one type of typical IT-based Clinical decision support systems (CDSS), automated clinical alerts that assist with diagnostic decisions have been more ubiquitous in EMR systems and have been playing important roles in recent years (Ancker et al. 2017; Daniel et al. 2019). While some AI-enabled CDSS and automated clinical alerts have been promising (Kucher et al. 2005; Schedlbauer et al. 2009; Strom et al. 2010b), many of these systems continue to face significant challenges in actual implementation and use (Garg et al. 2005; Jaspers et al. 2011; Strom et al. 2010b; Wilson et al. 2015). Whereas some recent work has pointed to the need to assist or explain the AI algorithms to help physicians trust such systems (Ahmad et al. 2018; Kelly et al. 2019), the more pressing issue is the common problem of physicians' overriding such alerts (Van Der Sijs et al. 2006).

Extant research on automated clinical alerts and why physicians override alerts suggest that alert fatigue is one significant factor, and that has led to many recent studies that looked at how to better design alert systems to

mitigate alert fatigue (Ancker et al. 2017; Ash et al. 2004; Carspecken et al. 2013; Kesselheim et al. 2011; Nanji et al. 2014; Oppenheim et al. 2002; Phansalkar et al. 2013; Weingart et al. 2003). However, growing research suggests that alert fatigue is not the only main factor. Instead, a physician's decision to override alerts is also related to individual physician factors (e.g., physicians perceive alerts as less knowledgeable) and other contextual factors such as workflow or time constraints (Van Der Sijs et al. 2006). Unfortunately, while these studies have given made these factors salient, there is no holistic framework to explain how they influence the physicians' perception and interpretation of alerts and eventually their decisions towards these alerts. In the second essay, I provide a novel perspective to why physicians override automated clinical alerts.

CHAPTER 2 LITERATURE REVIEW

Implementing AI in Clinical Judgement and Decision-Making

Scholars working in healthcare IS research have noted the importance of AI technologies such as predictive analytics and machine learning that can provide advanced decision support function for medical teams to improve clinical decision making (Agarwal and Dhar 2014; Chen et al. 2012; Fichman et al. 2011). Clinical decision making is a knowledge intensive work process that focuses on decision making on diagnosis and prognosis of symptoms and signs and prescribing appropriate treatment (Norman 2000; Sox et al. 2013). At a fundamental level, the clinicians' decision making involves gaining information of symptoms (subjective patients' presentation of complaints), signs (medical facts based on physical response or characteristic detected by a clinician, nurse, or medical device during the examination of a patient) and disease. As discussed in the literature, clinical decision making is often shaped by a certain level of uncertainty, complexity, and equivocality, albeit at varying levels (Croskerry 2005; Norman 2000).

Despite positive views of using AI technologies to support clinical judgment and decision making, many studies have found that AI use has serious limitations and is associated with mixed outcomes (Bardhan et al. 2015; Helm et al. 2016; Henry et al. 2015a). IS researchers acknowledge that the successful adoption and implementation of AI technologies needs to consider the unique characteristics of medical work and clinical decision-making process (Agarwal

and Dhar 2014; Kohli and Tan 2016). They argue that one of the key reasons why AI technologies may not be as effective may be because AI design and implementation may not be aligned with how actual clinical work and decision-making process are conducted.

To support clinical decision making that is often faced with a significant degree of complexity and significant consequences, IS and medical informatics researchers tend to focus on the information issues (uncertainty and complexity aspects) of the decision making process and use an analytical approach to support the clinicians (Amarasingham et al. 2014; Croskerry 2005; Norman 2000). Uncertainty is characterized as a lack of information about either the signs, the alternative treatments or their consequences, which makes interpreting a situation and making a decision more difficult (Lipshitz and Strauss 1997; Platt and Huettel 2008). Complex situations are characterized by an abundance of alternatives or dimensions (Payne 1976). To address the uncertainty (lack of information) and complexity (multiple information sources and relationships), IS and medical informatics designers of AI technologies tend to adopt an analytical approach in how they model the decision-making process. Analytical approach assumes that clinicians engage in methodical, laborious information gathering and analysis, where they carefully weigh all the best evidence and develop alternative solutions in an attentive, rational, and objective fashion. Thus an analytical approach involves analysing knowledge through conscious reasoning and logical deliberation (Croskerry 2005; Dane et al. 2012).

AI technologies' problem-solving and big data information processing capabilities are thus designed to support this type of analytical decision-making process by reducing uncertainty and complexity. For example, AI technologies such as machine learning, and predictive analytics enable well-deliberated calculations that integrate otherwise unmanageable amounts of data; these technologies are built with formalized analytical models to help evaluate alternative decision options. By assuming that all potential information can be collected and processed, AI technologies are designed to help inform clinicians of various alternatives (treatment strategies) with their respective consequences (Norman 2000). At the same time, AI technologies are capable of processing massive amount of information at a speed beyond the cognitive capabilities of even the smartest human decision makers and thus help reduce the complexity of decision making.

Limitations of Traditional AI Approaches in Healthcare

Current predictive analytics and other AI technologies whilst effective in solving probability-based decision contexts are unfortunately ill-equipped to tackle such equivocal and ill-structured problems (Guszcza et al. 2017). First, AI typically require numerous examples to be trained and are less capable in environments that are outside of a predefined domain of knowledge or where the problem space is not well defined and known (Brynjolfsson and McAfee 2012). Organizations rarely have the number of examples needed to understand relationships between everything in the world that can affect an organization, i.e., the ratio of “examples of past similar decisions” to “stuff that might be

important for those decisions” can be abysmally low (Ransbotham 2016).

Second, because clinicians’ decision-making process at times can be more intuitive than analytical, current AI designs that focuses mainly on speed, size, and range of information processing may not be as effective for intuitive approaches, which often requires more focus on breadth of the issues as well as the ability to consider the problem from a more holistic and abstract view. For example, AI currently may not be able to simulate clinician’s tacit experiences that lead to relatively inexplicable perceptions to ambiguous problems and emergent issues (Meyer et al. 2014; Sadler-Smith and Shefy 2004). Third, even as IS and medical informatics researchers attempt to address some of the obvious shortcomings of predictive analytics in healthcare (e.g., Meyer et al. (2014) study in managing ill-structured problems by refining control strategies or Lin et al. (2017) study in modelling adverse health events using multiple events and outcomes simultaneously), these refinements still assume a rational, deterministic analytic approach in their designs (e.g., Meyer et al. (2014) implicitly assumes that the initial set of control strategies used for diabetic care are comprehensive and applicable to all patients). In sum, given that they are designed mainly from an analytical approach, current AI are not aligned to the needs and processes of clinical judgement and decision making.

One approach to this issue is to acknowledge that analytical and intuitive approaches are not mutually exclusive and can be employed as an amalgamated solution where analytical and intuitive approaches work in parallel (Hung 2003). For example, better decision making can be achieved by having

algorithms “sift through vast amounts of data to highlight the most interesting things while human managers can drill down, using human intelligence, to reach conclusions and take actions” (Hoffman 2016). Here I propose the CE-AI design i.e., where designs of AI technologies attempt to incorporate and complement the advantages of computers and algorithms with collective human intuition in problem solving.

Physicians’ Decision-Making and Automated Clinical Alerts

Extant research suggests that CDSS and automated clinical alerts can potentially have positive impacts on physician decision-making. For example, one study has shown that an electronic prescribing system with decision support can prevent about half of harmful prescription errors (Bobb et al. 2004). In addition, other small scale studies have shown that automated clinical alerts were relatively effective in reducing drug interactions (Strom et al. 2010a; Strom et al. 2010b), increasing the rate of venous thromboembolism prophylaxis (Kucher et al. 2005), reducing hospital stay (Durieux et al. 2008), and improving the rates of various other preventive measures to positively influence outcomes and improve patient safety (Dexter et al. 2001; Schedlbauer et al. 2009).

However, many implemented automated clinical alerts have had varying impact on prescribing quality (Schedlbauer et al. 2009) and an even less clear effect on some clinical outcomes (Garg et al. 2005; Strom et al. 2010b). A synthesis of the results of several studies examining automated clinical alert systems noted that 57% had effect on physician decision-making and only 30%

have shown that their decisions had a positive effect on patient outcomes (Jaspers et al. 2011). For example, a study by Wilson et al. (2015) showed that the clinical alert for acute kidney injury did not improve clinical outcomes among patients.

Factors that explain the above mixed findings include poorly designed clinical alerts or failed implementations. However, extant research has shown that user non-adherence by overriding of alerts (i.e., failure to perform the recommended action) is one of the most significant factors for the suboptimal efficacy of the alerts (Durieux et al. 2008; Van Der Sijs et al. 2006; Weingart et al. 2003). Studies show that physicians override up to 90% of presented alerts, which makes the impact of these alerts substantially limited (Van Der Sijs et al. 2006; Weingart et al. 2003). This is problematic, as these alerts are designed to assist medical experts in their decision-making and may have significant impact on the patient's health.

Why do physicians override alerts? As there is limited research on why experts override AI algorithm alerts, I adopted a broader approach to examine why individuals in general override AI algorithm prescriptions or other data-driven alerting systems. A review of the literature suggests three broad reasons for why overriding behavior occurs.

First, the *characteristic of the individual* exposed to the AI algorithm influences how she reacts to its alert. Kahneman (2011) argued that when individuals feel that they have more information or are more knowledgeable than the AI algorithms, they tend to override these alerts. For example,

physicians may choose to override AI algorithm alerts because they trust more in their clinical knowledge or their intimate knowledge of the patient (Magnus et al. 2002; Taylor and Tamblyn 2004) and therefore downplay the “advice” or prescriptions provided by the alerts (Timmermans and Mauck 2005). This view is echoed in Lee et al.’s (2015) study where they observed that Uber drivers typically trust their knowledge about the conditions of the street more than suggestions provided by AI algorithms. Further, they found that Uber drivers used their knowledge to work around how the algorithm assigns rides to drivers.

Other studies also found that the individual’s experience and familiarity with the algorithm (i.e., the number of prior interactions the individual had with the algorithm) determined if the individual will override the algorithm over time (Xiao and Benbasat 2007). In other words, how individuals *construe* the AI algorithm alert is another factor of overriding behavior related to an individual characteristic. The users’ perception of the algorithm’s sophistication is negatively associated with overriding behavior, i.e., the more sophisticated a user perceives the AI algorithm, the less likely the user will override the alert (Muir and Moray 1996; Wang and Benbasat 2007). Likewise, in recommendation algorithms, the more users experience the algorithm’s degree of personalization, the more the user would adhere to the AI algorithm’s recommendations (Komiak and Benbasat 2006), and the user’s characteristics such as motivation, ability, and personality traits influence their level of information processing and eventual acceptance of the recommendation (Tam

and Ho 2005). Additionally, prior research in IT security warning systems has also shown that individuals' susceptibility to fear influences users' overriding of IT security warnings (Johnston and Warkentin 2010).

Second, the *environmental* and *contextual factors* in which the AI algorithm alert is triggered and how the alert relates to the context impact how an individual reacts to the alert. Individuals are primed by the environment in which they make decisions, and the environment provides cues that influence how they construe AI algorithm alerts (Kahneman 2011). For example, physicians may be inclined to override AI algorithm alerts when they occur during busy time periods (Overhage et al. 1997) or when they interrupt the physicians' workflow (Ash et al. 2004). Conversely, AI algorithm alerts that are perceived to be aligned with the individuals' decision-making process are more likely to be accepted by the users (Tam and Ho 2005). This is reflected in a study by Krall and Sittig (2002), who found that alerts were more likely to be helpful if they appear when the physician is deciding on a related clinical problem. However, in certain contexts, it may be a norm that AI algorithm alerts are ignored. In a study of emergency security alerts (Han et al. 2015), individuals exposed to the alerts will act based on how others in their social circle react to the alert, i.e., social norms influence whether alerts are ignored. On the other hand, how the AI algorithm alert is framed also impacts the user's perception of the alert. Algorithms framed as more transparent and interpretable are less likely to be overridden by users (Pasquale 2015; Rader et al. 2018; Rudin 2019).

Third, the *equivocality* of the *underlying task* the AI algorithm seeks to solve also influences the likelihood of an alert being overridden. In highly equivocal tasks, experts tend to exhibit a higher level of confidence and hubris when comparing themselves to algorithms, as such there is a higher occurrence of alerts being overridden by experts (Kahneman 2011). This finding is also observed in product purchase settings, where consumers assess their preferences for products. In settings where the purchasing task is unambiguous, and consumers can state their preferences, they are less likely to override AI algorithm purchase recommendations (Lee and Benbasat 2011). On the other hand, Xiao and Benbasat's (2007) review of AI recommendation systems concluded that the complexity of products moderates user-decision quality and perception of AI algorithms. When consumers construe the product choice task as complex, they are more likely to consider more product attributes in the decision-making process, and thereby increasing the probability of overriding the AI recommendations. Finally, AI algorithm alerts that provide actionable follow-up tasks are less likely to be overridden (Han et al. 2015).

Put together, my literature review shows that the reasons for overriding AI algorithm prescriptions and alerts are linked primarily to how an individual *perceives, comprehends, and interprets* the AI algorithms given the individual's current work context, current knowledge, and their assessment of the situation. Thus, the individual's *construal* of the AI algorithm is a consistent conceptual theme across those studies examining overriding behavior. As such, I propose that the individual expert's characteristics and the context of the alert

influence how experts *construe* the AI algorithm alert in the healthcare context, and in turn, this construal influences the overriding actions that follow. I thus apply the Construal Level Theory to understand how physicians construe AI algorithm alerts and how that impacts their decision to override alerts.

Construal Level Theory (CLT)

CLT is a theory that accounts for the connection between *psychological distance* and *construal level* (i.e. levels of abstraction) in people's daily cognition, such as the perception of a long term plan or evaluation of a new-launched product (Trope et al. 2007). CLT posits that the greater a person's psychological distance from a referent target (e.g., person, event, object), the more likely the target will invoke a high-level construal. On the contrary, the smaller a person's psychological distance is to a target, the target will more likely invoke a low-level construal.

Psychological distance is a "subjective experience that something is close to or far away from the self, here and now" (Trope and Liberman 2010, p. 440). CLT literature has identified four dimensions of psychological distance: spatial, social distance, hypotheticality, and temporal (see Trope and Liberman (2010) for a comprehensive review). For instance, attending an academic conference in two years – a temporally distant event – is associated with high psychological distance. Similarly, a car accident that occurs two blocks away is a spatially distant event compared to an accident that occurs next to oneself. Numerous research has found that as the psychological distance from the target

increases, people also increasingly tend to invoke a high-level construal about that target (Trope et al. 2007).

High-level construals are viewed as “[more] abstract, [less] coherent, and superordinate mental representations” compared with low-level construals (Trope and Liberman 2010). High-level construals also have fewer details; for example, the word “illness” invokes a higher construal level compared to “skin cancer”, which invokes a low-level of construal with more concrete details.

The construal level in turn affects an individual’s prediction of outcomes and evaluation of tasks (Trope et al. 2007). Construal levels in one’s mind determine whether people will focus more on full picture or general aspect of the object (“the forest”) or the detailed aspect (“the trees”) (Smith and Trope 2006) and whether the object will be categorized into rough and bigger groups, or specific and smaller groups (Liberman et al. 2002). In situations where high-level construals are invoked, people tend to interpret the situation by asking more *why* instead of *how* (Fujita et al. 2006; Liberman and Trope 1998) and tend to rely more on dispositional inferences rather than situational inferences (Henderson et al. 2006; Nussbaum et al. 2003).

In summary, CLT suggests that one’s psychological distance towards the target can be represented by different forms of distance between a person and a target ; that psychological distance leads to different construal levels about the target, which impacts how one perceives, evaluates, and behaves towards the target (Cubitt et al. 2018). CLT has been applied in many IS studies to explain phenomena such as user contribution motivations in online feedback

forums (Phang et al. 2015), consumers' acceptance of online recommendations (Köhler et al. 2011), IT project risk management activities (Lee et al. 2019), information search and decision-making (Shani et al. 2009), effectiveness of mobile targeting (Luo et al. 2014), and IT adoption tendencies (Shen and Chiou 2010).

CHAPTER 3 ARTIFICIAL INTELLIGENCE IN SEPSIS EARLY PREDICTION AND DIAGNOSIS USING UNSTRUCTURED DATA IN HEALTHCARE

Introduction

Sepsis is a leading cause of death in United States hospitals (Komorowski et al. 2018), accounting for half of all hospital deaths (Sweeney et al. 2018). Early prediction of sepsis is crucial in preventing mortality, given that sepsis management is highly time sensitive (van der Poll et al. 2017). Based on international medical guidelines (Rhodes et al. 2017), early fluid resuscitation is recommended to commence within the first three hours to stabilize sepsis-induced tissue hypoperfusion (Rhodes et al. 2017), and administration of intravenous antimicrobials is recommended to commence at the earliest possible time, specifically within one hour of sepsis (Ferrer et al. 2014; Kumar et al. 2006; Rhodes et al. 2017). Further, sepsis mortality increases significantly with each hour of delay in antimicrobials administration (Ferrer et al. 2014; Kumar et al. 2006). As sepsis management is often based on a standardized management approach (Rhodes et al. 2017), early sepsis identification may be practically challenging and operational constraints in healthcare delivery can lead to unacceptably high mortality rates. For example, delay in communication among clinicians, nurses, and pharmacist exacerbates delay in sepsis management. Hence, the early prediction of sepsis before its

onset in a patient gives clinicians additional lead time to plan and execute treatment plans.

Most of the existing methods for sepsis diagnosis and early prediction only take advantage of structured data stored in the electronic medical records (EMR) system (Desautels et al. 2016; Islam et al. 2019; Mani et al. 2014). However, research has shown that about 80% of the clinical data in EMR systems consist of unstructured data – i.e., data stored without a pre-determined or standardized format (Jensen et al. 2017). Common examples of such unstructured data are free-form text (e.g., clinical notes) or images (e.g., radiological images). These unstructured clinical data contain rich information, i.e., additional clinical details not captured in the EMR structured data fields. Physicians use such unstructured clinical data fields to record “free-form” clinical notes as structured data is designed to store only pre-determined discrete data (e.g., patient vital signs). Physicians also depend on the unstructured data to review judgments and critical clinical information entered by other clinicians to gain a better understanding of a patient’s condition or effects of their treatment. Consequently, unstructured EMR data is a potentially rich data source to develop better artificial intelligence (AI) tools, especially for medical conditions such as sepsis, where early symptoms are ambiguous and difficult to recognize. To this end, more recent work has incorporated text-mining of clinical notes to improve the accuracy of early sepsis prediction (Horng et al. 2017; Liu et al. 2019).

In the area of clinical notes mining, researchers have used natural language processing (NLP) mainly to identify and extract medical events, medication information, and clinical workflow from unstructured text data stored in EMR systems (Hsu et al. 2015; Popejoy et al. 2014; Xu et al. 2011; Yang et al. 2009). While most NLP applications on clinical notes mining have mainly focused on identifying and extracting concepts, my study extends prior text-mining sepsis predicting algorithms (Horng et al. 2017; Liu et al. 2019) to diagnose and predict sepsis with higher accuracy and with a longer early warning lead time of up to 48 hours.

It is a well-known fact that the diagnosis of sepsis is often equivocal due to the varied nature of infection sources and wide-ranging patients' responses (Desautels et al. 2016). Prior research that used text mining to augment sepsis prediction advocates the use of extracting common words (Horng et al. 2017; Liu et al. 2019) that are found in clinical notes to improve model predictive accuracy. In my study, I extend this stream of research by showing that modeling for common topics can lead to higher levels of predictive accuracy and the ability to predict up to 48 hours prior to the onset of sepsis. I argue that the use of topics - as opposed to words - is preferred, as lexicographical topics are more stable compared to individual words (Blei 2012; Wallach 2006). Further, this method is more generalizable because it can mitigate challenges such as idiosyncratic words used by different physicians due to differences in writing styles.

Here, I developed a topic-based, NLP-enabled AI algorithm that combines the NLP analysis of physicians' clinical notes with structured EMR data to improve our ability to predict the risk of sepsis. Specifically, my algorithm extracts, analyzes, and summarizes physicians' clinical notes and combines these sets of summarized clinical information with structured clinical variables to classify if a patient has sepsis at the time of analysis. For patients who are not classified as having sepsis at the time of analysis, the algorithm will then predict the risk of those patients having sepsis in the following 4, 6, 12, 24, and 48 hours. Unlike prior NLP sepsis predictive models, my model operates in situations where the prevalence of sepsis is as low as six percent – equivalent to the prevalence of sepsis in hospitals observed in historical studies (Rhee et al. 2017). Additionally, I found that mining clinical notes provide significant improvements to the predictive accuracy over structural variables for predictions 12 to 48 hours ahead of the onset of sepsis; this is also beyond the accuracy of physician predictions. As discussed above, under standard sepsis management protocol, every hour delay in completion of the administration of the 3-hour bundle is found to be associated with a 4% increase in mortality (Seymour et al. 2017); thus the ability of my algorithm to speed up diagnosis and early detection would potentially reduce overall mortality in hospitals in a significant way.

Data Sample

I obtained approval from a Singapore government hospital's medical board to extract de-identified patients' data from the EMR system to build the

predictive algorithm. MySQL 8.0 was used to extract the data from the Epic™ electronic medical record system. The sample consists of 5,317 patients admitted from 2nd April 2015 to 31st Dec 2017 with 114,602 clinical note entries. I separated this sample into a training and validation sample and an independent, hold-out, test sample. I used a 10-fold cross-validation methodology to train and validate the model on the training and validation sample. In order to test the stability of the text topics extracted from my text-mining algorithm, I deliberately designed the hold-out test sample to include patients admitted from a later period (10th May 2017 to 31st December 2017). The training and validation sample consists of 3,722 patients (80,162 clinical note entries) while the independent, hold-out test sample consists of 1,595 patients with 34,440 clinical note entries. I use each patient consultation instance (clinical note) as the unit of analysis (see Figure 1 in for more details about the data sample).

*** ICD-10 Code:**

Sepsis: 'A40.0','A40.1','A40.8','A40.9','A41.2','A41.0','A41.0Z16','A41.1','A40.3',
'A41.4','A41.50','A41.3','A41.51','A41.52','A41.53','A41.59','A41.81','A41.89',
'A41.9'

Severe Sepsis: 'R65.20','R65.21','R65.10','R65.11'

Figure 1. Data Sample for Training, Validation, and Testing of SERA Algorithm

I construct the dependent variable for sepsis using the ICD-10 code for sepsis, severe sepsis, or sepsis shock, and the presence of ICU admission. The hospital practice is such that patients diagnosed as having sepsis are transferred to the ICU ward; hence, patients with at least one of these ICD-10 codes and admitted to ICU ward are allocated to the sepsis case cohort. All other patients not meeting these criteria are allocated to the non-sepsis control cohort. There are 240 sepsis patients in the training and validation sample and 87 sepsis patients in the test sample. I define the sepsis onset time as the ICU ward admission time as per the hospital's practice.

At the patient encounter level, the prevalence of sepsis in my case-control sample is 6.15%, which is equivalent to the estimated sepsis's prevalence of around 6% in hospitals (Rhee et al. 2017). Given that the data is imbalanced, I applied SMOTE (Synthetic Minority Over-sampling Technique) to achieve 1:1 balanced data for sepsis cases and non-sepsis controls (at the clinical note level). Prior literature argued that oversampling (instead of undersampling) will result in more accurate models (Batista et al. 2004; Carnielli et al. 2018; Chawla et al. 2002) and SMOTE has been used in earlier studies that develop machine learning classifiers for other clinical conditions such as oral cancer detection (Carnielli et al. 2018) and cell identification/classification (Rennie et al. 2018; Xia et al. 2020) where the prevalence of the positive cases are low. For comparative purposes, I also develop, test, and report the models without any oversampling to present the possibility of

operating this algorithm in a normal clinical environment where the prevalence of sepsis is relatively low.

Sepsis Early Risk Assessment (SERA) Algorithm

Figure 2 outlines the steps used to develop the SERA algorithm. The unit of analysis for the SERA algorithm is each patient consultation instance given that the algorithm utilizes clinical notes and structured data to make a sepsis risk assessment. I adopt this unit of analysis to ensure that the algorithm can operate in typical clinical context where clinicians consult, assess, and diagnose patients. SERA algorithm consists of two inter-linked algorithms—a diagnosis algorithm and an early prediction algorithm. The diagnosis algorithm determines if the patient has sepsis at the time of consultation and if not, the early prediction algorithm will determine the patient’s risk of having sepsis in the next 4, 6, 12, 24, and 48 hours.

For the diagnosis algorithm, I combined EMR system clinical notes for each consultation with the most recent structured variables available in the EMR system, as listed in Table 1. These data are used to classify if the patient has sepsis at the time of consultation. For the early prediction algorithm, the data structure is similar to that of the diagnosis algorithm except that the early prediction algorithm will not consider any patient consultations from the sample when the patient has been confirmed to have sepsis (i.e., transferred to an ICU). This omission is necessary to prevent positively biasing the predictive power of the algorithm because these positive sepsis consultations would have traits that are strongly associated with sepsis conditions.

Figure 2. Setup of SERA Algorithm. The flow diagram shows the steps used to develop the SERA Algorithm. Both structured data (vitals, investigations, treatment) and unstructured data (clinical notes) are used in the process of diagnosing and predicting sepsis.

Table 1. List of Structural Variables Used in Predictive Algorithm

Category	Predictors
Patient Information	Age, gender
Vital Signs	Blood pressure, heart rate, temperature, oxygen saturation, respiratory rate
Investigations	Total white cell, culture results ^a , lactate, high-sensitivity C-reactive protein, procalcitonin, arterial blood gas
Treatment	Use of vasopressor, use of antibiotics

Note: ^aThe cultures include: culture of urine, blood culture (anaerobic/aerobic), general culture, wound culture, stool culture, molecular, fluid culture, respiratory culture, cerebrospinal fluid culture, fungus smear, tissue culture, fluid culture (bactec bottle), fluid culture (aerobic), ear culture, genital culture, sterility testing, tip culture, wound culture (aerobic), appearance, stain results, or any real-time polymerase chain reaction. Cultures are typically used to confirm test presence of carbapenem-resistant enterobacteriaceae, methicillin-resistant staphylococcus aureus, clostridium difficile antigen, clostridium difficile toxin, clostridium difficile influenza, influenza B (real-time polymerase

chain reaction), influenzae virus, strep pneumoniae antigen, carbapenem-resistant Enterobacteriaceae screening, hepatitis viral load, cytomegalovirus, methicillin-resistant staphylococcus aureus screening, legionella antigen, sarcoptes scabiei, herpes simplex virus.

Processing of Clinical Notes and Machine Learning

Before using the unstructured clinical notes as predictors in the algorithm, I applied NLP to these notes, specifically using the Latent Dirichlet Allocation (LDA) topic modeling algorithm. The major topics found in the progress notes are extracted, numerically weighted, and combined with the structured variable data in Table 1. By applying the NLP LDA model to the clinical notes dataset, I identified 100 common text topics that are in turn classified under one of the following seven categories: 1) clinical status, 2) communication, 3) laboratory tests, 4) non-clinical status, 5) social relationships, 6) symptom and 7) treatment. The numerical loadings on the 100 topics, together with the structured data, are used as predictors in the diagnosis algorithm and the early prediction algorithm.

In the diagnosis algorithm, the data is subjected to a voting ensemble machine learning algorithm. For comparative purposes, I also used dagging, and gradient boosted trees (GBT) as two alternative classifiers. For consultations where the patient is classified as not having sepsis, the early prediction algorithm will then predict if the patient has a high risk of having sepsis in the next 4, 6, 12, 24, and 48 hours using the voting ensemble machine learning method. The dependent variables for the early prediction algorithm are

whether the patient will have sepsis in the next 4, 6, 12, 24, and 48 hours, respectively.

The trained and validated models are subsequently tested using the independent, hold-out test sample. Methodological details of the text-mining, diagnosis, and early prediction algorithms are provided in the Methods section.

Text mining technique

In addition to the structured data (Table 1) used in the predictive model, I utilized clinical notes data that are unstructured free-form text. Before the unstructured free-form text could be analyzed and used as part of a predictive model, I first utilize Latent Dirichlet Allocation (LDA) (Blei et al. 2003) to codify the unstructured free-form text into numerical representation. LDA, like other topic modeling algorithms, is an unsupervised technique that empirically creates topics based on patterns of (co-)occurrence of words found in the analyzed documents. Through LDA I generated topics from the clinical notes that are represented by a vector with individual topic loadings corresponding to the concentration of discussion for a particular topic.

I processed the unstructured free-form text using the following five steps. First, under HIPAA guidelines and standard for anonymization of data (Health Insurance Portability and Accountability Act of 1996, HIPPA), I removed all possible identifiers within the clinical notes. Identifiers that I parsed out included: names, geographic subdivisions, zip codes, ID numbers, all elements of date, telephone/ fax numbers, vehicle numbers, device serial numbers, and email addresses. Second, I tokenized the text contained in the

documents by removing all punctuation marks; lemmatizing the words by replacing words by their root form; applying part-of-speech tagging, and removing stop words, such as articles and prepositions. I also removed an extended list of medical-related stop words or phrases that are common in these texts but have no practical meaning. Examples of such are “report,” “progress,” “provide,” “lab unit” (A complete list of the words and phrases that are excluded from topic modeling is available upon request from the authors). After these two steps, I created a term-document matrix, where rows represented the occurrence (a measure of frequency) of a term in the documents, and columns represented the documents.

Third, I applied a text filter to the processed text to reduce the number of terms by eliminating rare terms and weighting the frequencies of terms with multiple occurrences. Fourth, I ran the LDA topic clustering algorithm on the terms to determine the various topics. Based on extant literature Schmiedel et al. (2018), the number of topics that can be developed from the text is highly subjective and is often a function of the number of observations or the expected variety of the topics in the dataset. Using these guidelines, and as part of my robustness checks, I attempted five different iterations with 25, 50, 75, 100, and 150 topics. Subsequent analysis of the results yielded qualitatively similar results for all five iterations in the final predictive model; thus, in this paper, I reported the 100-topic model for brevity. Drawing on the 100-topic model, I presented the 100 topics to three researchers to independently classify them into

seven topic categories (Table 2). All text-mining are processed using SAS Enterprise Miner 14.2.

Note that the step of developing the topics is required only during the model development and validation phase. When using the text-mining algorithm to assess new clinical notes, the 100 topics developed are used as the benchmark to help compute the fit between the new clinical note to these 100 topics. High fit measure represents a high similarity between the new clinical notes with the benchmark topics. From a practical implementation perspective, I found that the computation of the fit values is not computationally intensive. For example, a test-case patient with a lengthy clinical note of 1,806 words (the median length of the clinical notes in my sample is 840 words) requires only 0.17 seconds to estimate these fit values using an Intel i7 Processor 2.7 GHz, with 16.0G RAM in SAS Enterprise Miner 14.2.

Table 2. Topic Categories

Category	Count	Definition
clinical status	28	Routine updates of clinical conditions as well as diagnosis (e.g., vitals) excluding lab and radio-diagnostic tests
communication	3	Communication between staff
lab test	24	Orders and reports of lab or radio-diagnostic test results
non-clinical status	2	Routine updates of non-clinical conditions
social relationship	2	Information about family and social aspects of patient
symptom	10	Clinical symptoms
treatment	31	Treatment procedure or medication prescribed as well as the status of the treatment/ medication

Note: The 100 topics are classified into 7 different categories. The distribution of topics among categories is similar if 25, 50, 75, or 150 topics are extracted instead. Detailed results are available upon request from the corresponding authors.

Machine Learning Algorithm

Ensemble methods are machine-learning algorithms that utilize multiple classifiers to determine the predicted outcome by taking a (weighted) vote of their predictions. These methods often perform better than any single classifier (Abawajy et al. 2015; Zhang and Ma 2012). There are several different ensemble methods, such as voting, bagging, stacking, and boosting.

In my main estimation, I use a voting ensemble. Voting is an ensemble machine learning model that combines the predictions from multiple other models (base classifiers). Here, I use two base classifiers: a stochastic gradient descent (SGD) based logistics regression and a random forest algorithm. My combination rule is an average of probabilities, i.e., I calculate the average probability of the two base models as my voted probability.

The first base classifier, SGD, is an optimizing algorithm that seeks to minimize the error in prediction by learning iteratively from prior fitted estimates. The method iteratively draws random samples from the training sample to estimate the parameters of the model that is used to classify a patient as having sepsis or not having sepsis. It learns from each sampling iteration to determine the accuracy of the classification and adjust the parameter estimation until further improvements in prediction results are minimal.

For each iteration, the predicted parameter β is calculated, and the model is updated using the following logistic equation:

$$\beta^{new} = \beta^{old} + lr(y - \hat{y}) \hat{y} (1 - \hat{y}) x \quad (1)$$

where β is the optimized parameter, lr is a learning rate, $y - \hat{y}$ is the prediction error for the model in a particular iteration in the training data, $\hat{y}$ is the prediction made by the coefficients and x is the input value. In my case, the input variables were a combination of the structured variables (as indicated in table 1) and the topic loadings of each clinical note on the 100 topics I extracted in the text mining procedure.

The second classifier used here for voting is a random forest classifier, with the case of sepsis being the target variable. The probabilities of both classifiers are averaged out to arrive at the final probability used in my voting ensemble model.

Alternative estimators. For comparative purposes, I also used two different alternative estimators, namely, dagging and gradient boosted trees (GBT).

Dagging is an ensemble method that has been widely used in existing machine learning literature when the data is “noisy,” i.e., data with a large amount of additional irrelevant information. In dagging, I partition the training data sample into a set of disjoint stratified samples. I then select the base classifier Logistic Regression with Stochastic Gradient Descent (SGD) as described earlier for this procedure, and train the data using this base classifier within each of the disjointed samples. Next, the ensemble method applied the results from the trained copies of the base classifier to the validation data

sample. I compute the average across all sub-samples, and the predicted result is based on a vote across all sub-samples.

GBT uses an ensemble of multiple trees to create more powerful prediction models for classification and regression. The key idea is to build a series of trees, where each tree is trained, so that it attempts to correct the mistakes of the previous tree in the series. GBT involves three elements: a loss function to be optimized, a weak learner to make predictions, and an additive model to add weak learners to minimize the loss function. Decision trees are used as the weak learner in GBT and a gradient descent procedure is used to minimize the loss when adding trees. A fixed number of trees are added, or training stops once loss reaches an acceptable level or no longer improves on the validation dataset. All ensemble machine learning was conducted using KNIME Analytics Platform (version 4.1.6).

Other Model Diagnostics. To observe the trade-offs between PPV (Precision) and Sensitivity (Recall), refer to the Precision-Recall curves in Figure 3. Similarly, the SERA algorithm's calibration curves can be found in Figure 4.

Figure 3. Precision – Recall Curve (SERA algorithm). The vertical axis (precision) is truncated and starts at 0.5.

Figure 4. Calibration Curves for SERA Algorithm

The Performance of SERA Algorithm

I report the test results of the SERA algorithm for both oversampled (SMOTE) and non-oversampled (non-SMOTE) data. The SMOTE models present the results typically observed in machine learning predictive models where the prevalence of sepsis cases is high and equal to that of the non-sepsis cases (Liu et al. 2019) (Table 3). The non-SMOTE models present the performance of the model in typical clinical settings where the prevalence of sepsis is low (Table 4). For brevity, I focus on describing the results for the SMOTE models as in prior studies (Carnielli et al. 2018; Rennie et al. 2018; Xia et al. 2020).

For the diagnosis algorithm, my test sample yielded an AUC of 0.94, sensitivity 0.89, specificity 0.85, and positive predictive value (PPV) 0.85. The AUC of the diagnosis algorithm is higher than those of prior studies (Desautels et al. 2016; Dummitt et al. 2018; Henry et al. 2015b; Horng et al. 2017; Liu et al. 2019; Lukaszewski et al. 2008; Mani et al. 2014; Pereira et al. 2011; Shashikumar et al. 2017; Thiel et al. 2010), which range from 0.64 to 0.92.

The SERA algorithm predicts if a patient has a high risk for sepsis *before* being diagnosed with sepsis by physicians in the hospital. The algorithm predicting 48 hours ahead of the onset of sepsis has an AUC of 0.87. The AUC improves to 0.90, 24 hours prior to sepsis, and up to 0.94, 12 hours prior to sepsis. Notably, my algorithm's prediction at the 12-hour lead time has a higher AUC, sensitivity, specificity, and PPV than prior research (Horng et al. 2017; Islam et al. 2019; Liu et al. 2019; Nemati et al. 2018). My 24-hour lead time

early prediction also has a high AUC of 0.90 and at least 0.80 sensitivity, specificity, and PPV values. The additional hours for early prediction is crucial in improving clinical outcomes as every one hour delay in antimicrobial administration is associated with a decrease in patient's survival of 7.6% (Kumar et al. 2006). The additional 24 hours lead time ahead of similar predictions will provide clinicians additional lead time to prepare for highly time-sensitive sepsis management intervention.

Table 3. Statistics of Diagnosis and Early Prediction Algorithm (SMOTE)

Diagnosis Algorithm							
<i>Outcome</i>	Voting					Dagging	GBT
<i>Predict if patient has sepsis</i>	AUC	Sensitivity	Specificity	PPV	NPV	AUC	AUC
At present time	0.94	0.89	0.87	0.85	0.90	0.92	0.94
Early Prediction Algorithm							
<i>Outcome</i>	Voting					Dagging	GBT
<i>Predict if patient will have sepsis</i>	AUC	Sensitivity	Specificity	PPV	NPV	AUC	AUC
48 hours later	0.87	0.78	0.77	0.77	0.78	0.83	0.83
24 hours later	0.90	0.81	0.80	0.80	0.80	0.86	0.86
12 hours later	0.94	0.87	0.87	0.87	0.87	0.92	0.92
6 hours later	0.92	0.88	0.81	0.82	0.87	0.90	0.93

4 hours later	0.92	0.86	0.80	0.81	0.86	0.85	0.92
----------------------	------	------	------	------	------	------	------

Note: SMOTE applied to clinical notes to achieve a balanced sample of sepsis and non-sepsis case entries. SERA algorithm uses the Voting algorithm; Dagging and GBT algorithms are presented for comparative purposes.

Table 4. Statistics of Diagnosis and Early Prediction Algorithm (In Low Prevalence Condition without SMOTE)

Diagnosis Algorithm								
<i>Outcome</i>	Voting						Dagging	GBT
<i>Predict if patient has sepsis</i>	Prevalence	AUC	Sensitivit y	Specificit y	PPV	NPV	AUC	AUC
At present time	0.177	0.94	0.89	0.87	0.59	0.97	0.92	0.94
Early Prediction Algorithm								
<i>Outcome</i>	Voting						Dagging	GBT
<i>Predict if patient will have sepsis</i>	Prevalence	AUC	Sensitivit y	Specificit y	PPV	NPV	AUC	AUC
48 hours later	0.012	0.87	0.76	0.76	0.04	0.99	0.82	0.85
24 hours later	0.010	0.90	0.81	0.79	0.04	0.99	0.88	0.89
12 hours later	0.008	0.94	0.88	0.82	0.04	0.99	0.92	0.93
6 hours later	0.002	0.92	0.88	0.83	0.01	0.99	0.90	0.93
4 hours later	0.001	0.92	0.89	0.87	0.01	0.99	0.92	0.94

Note: No oversampled applied. Prevalence is computed at the clinical note level. For the same number of sepsis cases, the clinical note occurrences are different for a different time window. SERA algorithm uses the Voting algorithm; Dagging, and GBT algorithms are presented for comparative purposes.

SERA Algorithm vs. Human Prediction

I compared the performance of the SERA algorithm with other predictive scoring systems in the clinical context by juxtaposing the algorithm's prediction with standard clinical practice of diagnosing sepsis or predicting mortality due to infection. Clinical practice typically utilizes standardized scoring systems such as SIRS (Systemic Inflammatory Response Syndrome),

SOFA (Sequential organ-failure assessment), qSOFA (quickSOFA), and MEWS (Modified Early Warning System) to predict sepsis or mortality due to infection. Based on the meta-analysis study (Islam et al. 2019), the typical True Positive Rate (TPR, sensitivity) of these four scoring systems have AUC ranging from 0.50 to 0.78, TPR ranging from 0.56 to 0.8, and False Positive Rate (FPR, $1 - \text{specificity}$) ranging from 0.16 to 0.50 at 4 hours before the onset of sepsis. I plotted the ROC curves of my early prediction algorithm against these reported TPR and FPR scores (Figure 5: Panel a).

Further, I also estimated the TPR and FPR of the hospital physician's diagnosis of sepsis for all the patients' encounters within the independent, test sample. Based on medical guidelines for management of sepsis (Rhodes et al. 2017), physicians are required to normalize lactate in patients with elevated lactate levels for initial resuscitation. Further, physicians would, as part of diagnosis, obtain appropriate routine microbiologic cultures before starting antimicrobial therapy in patients with suspected sepsis to minimize delay in the start of antimicrobials. Hence, according to the hospital practice, when a physician suspects sepsis, she is to request one or more microbiologic culture test and lactate test as per above guidelines.

As such, in order to gauge the performance of the physician in predicting sepsis, I used the time when the physician orders both culture and lactate tests as the point in time when she suspects a patient has sepsis. The time between when orders are made and when the patient developed sepsis is determined as the prediction time window. Using this methodology, I computed

the TPR and FPR for physicians for five time-windows – 48, 24, 12, 6, and 4 hours before the onset of sepsis. In Figure 5 (Panel a), I can see that at 4 hours before the onset of sepsis, the SERA algorithm outperformed the hospital's physicians in predicting sepsis within the test sample. Additionally, SERA algorithm outperformed the typically reported accuracy rates of human-based scoring methods such as qSOFA, MEWS, SIRS and SOFA as reported in prior studies (Islam et al. 2019).

Figure 5. ROC Curves for 48, 24, 12, 6 and 4-hour Early Prediction. a and b The ROCs represent performance of early prediction algorithm at 4, 6, 12, 24 and 48 hours prior to onset of sepsis using the independent, test sample. “qSOFA”, “MEWS”, “SIRS”, and “SOFA” represent the TPR and FPR from these methods employed by physicians in prior studies at 0 to 4 hours prior to onset of sepsis. “Physicians” represent TPR and FPR of patients in the independent, test sample set that were suspected by hospital’s physicians to have sepsis at 4 hours prior to onset of sepsis. **b** “4h”, “6h”, “12h”, “24h”, “48h” represent TPR and FPR of patients in the independent, test sample set that were suspected by hospital’s physicians to have sepsis at the respective time prior to onset of sepsis.

Figure 5 (Panel b) presents the ROC of the early prediction algorithm for all the timings at 48, 24, 12, 6, and 4 hours before sepsis. I observe that the

ROC improves from 48 hours to 12 hours and maintains at a similar level beyond the 12-hour mark. The ROC of the SERA algorithm at all time periods outperformed hospital physicians' predictions. As a conservative comparison, I contrasted SERA algorithm prediction at 48, 24, 12, 6, and 4 hours against the prediction accuracy of MEWS, qSOFA, SIRS, and SOFA, performed at the 4-hour mark, as reported in the literature (Islam et al. 2019). I observe that SERA algorithm prediction for all four periods generally outperformed MEWS, qSOFA, SIRS, and SOFA prediction scorings even when they are assessed at the 4-hour mark before sepsis.

To examine the practical utility of the early prediction algorithm, I compute TPR and FPR of the early prediction algorithm at 48, 24, 12, 6, and 4-hour mark (Figure 6). I observed that 48-hour from the onset of sepsis, the SERA algorithm can pick up 0.78 of all patients that eventually have sepsis, and this prediction (TPR) improves to at least 0.86 for time periods less than 12 hours before the onset of sepsis. Hospital physicians, on the other hand, can detect approximately only 0.53 of all patients that eventually have sepsis at the 48-hour mark, and this proportion only increases marginally to 0.58 at the 6-hour mark. At the 4-hour mark, hospital physicians do observe a measurable increase in TPR to 0.65. For all five periods, the SERA algorithm has a higher TPR of predicting sepsis by 0.21 to 0.32 compared to hospital physicians in the same period. This suggests that the SERA algorithm has the potential to increase the number of early sepsis detection by 21% to 32% compared to relying only on hospital physicians' assessment.

Figure 6. Comparing Performance of the SERA algorithm vs. Physician. a

The bars represent the percentage of sepsis patient records correctly flagged as having a high risk of sepsis (likely to have sepsis) by either the algorithm or physicians. The chart compares the true positive rate of the algorithm's prediction at different lengths of time before the onset of sepsis against the true positive rate of physicians' prediction in the hospital. **b** The bars represent the percentage of non-sepsis patient records erroneously flagged as having a high risk of sepsis (likely to have sepsis) by either the algorithm or physicians. The chart compares the false positive rate of the algorithm's prediction at different lengths of time before the onset of sepsis against the false positive rate of physicians' prediction in the hospital.

The FPR of SERA algorithm varies from 0.23 at the 48-hour mark to 0.13 at the 12-hour mark (Figure 6: Panel b). The FPR of hospital physicians' predictions, however, is considerably higher than SERA algorithm, and it varies between 0.34 to 0.27 from the 48-hour to 4-hour mark. Thus, the use of SERA algorithm has the potential to reduce false positives by 0.07 to 0.17. Figure 6 suggests that the TPR and FPR of hospital physicians peaked only at the 4-hour mark – this means that physicians have a much shorter lead time for medical intervention. Conversely, the SERA algorithm is able to achieve considerably earlier prediction rates of up to 48 hours prior to sepsis onset, which means that physicians have much earlier warnings of sepsis onset and thus more time for more effective interventions.

Benefits of Unstructured Clinical Text in Early Sepsis Prediction

To quantify the added predictive value of unstructured clinical notes in the SERA algorithm, I use only the structured variables, as seen in Table 1, to diagnose and predict sepsis. I then compare the algorithm's performance (with only structure variables) against the algorithm with structured and unstructured clinical notes (Figure 7).

On the one hand, I observed that while the added use of clinical text did improve the performance of the SERA algorithm for diagnosis and early prediction (up to 6 hours), the improvement was marginal. On the other hand, when I consider time frames between 12 to 48 hours ahead of sepsis, I noted that the use of clinical text for early prediction had considerable benefits. Specifically, I observed that for the SERA algorithm in the 12 to 48-hour early

prediction window, the added clinical text improved 1) the AUC by between 0.10 to 0.15, 2) the sensitivity by 0.07 to 0.13, and 3) the specificity by 0.08 to 0.14.

Figure 7: Comparing SERA Algorithm Performance with Models without Clinical Text. I measure the change in performance for each diagnosis and prediction model when clinical text is added as additional predictors. Structured

variables represent the models that only include vitals, investigations, and treatment data as predictors. Structured variables + Text represent the models have additional clinical notes as predictors. **a** Compares the AUC of models with and without clinical notes. **b** Compares the sensitivity of models with and without clinical notes. **c** Compares the specificity of models with and without clinical notes.

The results suggest that for time periods closer to the onset of sepsis, the measurable symptoms of sepsis are manifested in the structured variables such as a drop in blood pressure values. In this instance, the added use of clinical text provides only marginal predictive gains to the SERA algorithm as the structured variables captured most of the sepsis symptoms. The unstructured clinical text, however, provides greater value in prediction when I am focusing on early prediction between 12 to 48 hours prior to sepsis as the patient has yet to manifest observable symptoms that can be measured by the structured sepsis variables. My model suggests that the physician's judgment and qualitative inputs of the patient's prognosis at that time provide additional crucial data that can be used for predicting sepsis.

Application of SERA Algorithm in Low Sepsis Prevalence

Environment

In a retrospective cohort study of 409 United States hospitals from 2009 and 2014, sepsis prevalence ranges from 1.8% to 12% of all hospitalized patients. The mean prevalence is 6%, and this remains relatively consistent over time (Rhee et al. 2017). Despite the naturally low prevalence of sepsis in clinical settings, most studies develop sepsis prediction algorithms using oversampled datasets with a significantly higher prevalence of up to 50%

(Desautels et al. 2016; Dummitt et al. 2018; Henry et al. 2015b; Horng et al. 2017; Liu et al. 2019; Lukaszewski et al. 2008; Mani et al. 2014; Pereira et al. 2011; Shashikumar et al. 2017; Thiel et al. 2010). In my study, I develop the models for both the oversampled environment, as seen in prior research, and for situations where the prevalence is low, as seen in a typical clinical environment.

As reported in Table 4, the models developed with the typical low clinical prevalence context of 6.15 percent can achieve high sensitivity, albeit with naturally low PPV. For purposes of clinical application, I simulate how changes in the prevalence of sepsis impact the PPV (Table 5). For example, for a 12-hour early prediction window, hospitals experiencing the lower the end of 1.8% sepsis prevalence will have an estimated algorithm PPV of 8.3% (95% CI: .89% to 8.73%). If the prevalence of sepsis increases to the higher end of 12%, the estimated PPV increases to 40.25% (95% CI: 38.94% to 41.58%). The simulated results illustrate the applicability of the SERA algorithm in natural clinical settings where the prevalence of sepsis varies depending on the type of clinical specialty and/or location of the institutions.

Table 5. Simulated PPV at Different Sepsis Prevalence Levels

		PPV (95% CI)					
	Prevalence	1.8%	6%	12%	20%	30%	50%
Algorithm	Diagnosis	0.11 (0.11-0.11)	0.30 (0.29-0.31)	0.48 (0.47-0.49)	0.63 (0.62-0.63)	0.74 (0.74-0.75)	0.87 (0.87-0.87)
	4hr	0.11 (0.10-0.12)	0.30 (0.28-0.33)	0.48 (0.45-0.51)	0.63 (0.60-0.66)	0.75 (0.72-0.77)	0.87 (0.86-0.89)
	6hr	0.09 (0.08-0.09)	0.25 (0.22-0.27)	0.41 (0.39-0.44)	0.56 (0.54-0.59)	0.69 (0.67-0.71)	0.84 (0.82-0.85)

	12hr	0.08 (0.08- 0.09)	0.24 (0.23- 0.25)	0.40 (0.39- 0.42)	0.55 (0.54- 0.57)	0.68 (0.67- 0.69)	0.83 (0.82- 0.84)
	24hr	0.07 (0.06- 0.07)	0.20 (0.19- 0.21)	0.35 (0.33- 0.36)	0.49 (0.48- 0.51)	0.63 (0.61- 0.64)	0.80 (0.79- 0.81)
	48hr	0.06 (0.05- 0.06)	0.17 (0.16- 0.18)	0.30 (0.29- 0.32)	0.44 (0.43- 0.46)	0.58 (0.56- 0.59)	0.77 (0.75- 0.77)

Note: 1.8% (12%) represents the low (high) end of the typical sepsis prevalence observed in hospitals based on prior research¹⁹. 6% represents the average sepsis prevalence of hospitals in the United States. Other prevalence percentages are for comparison purposes as they represent oversampled prevalence percentages typically observed in prior research.

Proposed Setup of SERA Algorithm Within the Clinical Environment

I propose two possible modes in which the SERA algorithm can operate within the clinical environment.

Background mode: In this first mode, the SERA algorithm is designed to run in the background. Specifically, it is configured to run during key events using the latest patient's clinical data available, e.g., during ward shift handovers. If the risk score exceeds the designated cut-off level, the EMR system will alert the physician. Alternatively, if there are more computing resources available, hospitals can choose to run it in fixed hourly-time intervals. For a large 500-bed hospital, if I run all cases individually, the SERA algorithm will take approximately 90 seconds to score all 500 patients. This approach ensures an ongoing, regular time-based sepsis risk assessment for patients within the hospital. (See Figure 8: Panel a for the workflow for this mode)

Figure 8. Workflow for the setup of SERA algorithm within the clinical environment. **a** A proposed workflow for the SERA algorithm to operate in the background and provide alerts when key events (such as ward shift handovers) are triggered. **b** A proposed workflow for the SERA algorithm to run immediately after a physician updates a patient's clinical notes in the EMR system.

Ad-hoc mode: In the second mode, the algorithm can be designed to run immediately after a physician updates her clinical notes in the EMR system.

In this case, the SERA algorithm will run in an ad-hoc manner since the score is only calculated after a physician has updated the patient's status. The SERA's score then acts as decision support mechanism to flag suspected sepsis cases. As observed from the study, the SERA algorithm outperforms physicians in early prediction of sepsis and thus may be an important early warning indicator to assist physicians in their diagnosis and patient care (See Figure 8: Panel b for the workflow).

Discussion

The SERA algorithm can potentially provide the early detection of sepsis that clinicians need to effectively intervene and manage sepsis patients. Medical practice and prior studies (Rhodes et al. 2017) have shown that early detection of sepsis is challenging, and sepsis patients can deteriorate rapidly; thus, every minute counts in the diagnosis of sepsis. I have shown the effectiveness of the SERA algorithm in that it is able to flag 21% to 32% more patients at risk of sepsis compared to hospital physicians in a clinical setting between 4 to 48 hours before the onset of sepsis. This additional lead time in sepsis alert provides greater opportunities for physicians to commence treatment, thereby lowering mortality.

As in every medical alert system, false positive in the alert system is a cost to the healthcare organization. Unacceptable levels of false positives can lead to wastage of medical resources as physicians need to perform additional follow-up diagnosis and treatment. Further, frequent false positives in a clinical setting will lead to a loss of confidence in these alerts among attending

physicians, who will, in turn, be less likely to act on future alerts. A high number of false alerts may also reduce their confidence in other similar best practice alerts embedded within EMR systems. The SERA algorithm, however, has achieved high TPR while maintaining a reasonable low FPR – a FPR considerably lower than hospital physicians’ early assessment of sepsis. Further, based on the simulation of different sepsis prevalence levels, I show that the PPV of my model is also appropriate for clinical applications, where the prevalence of sepsis is naturally low.

This study thus presents the potential of using AI as a “canary in a coal mine” for clinicians. I could envision running the SERA algorithm in the background to continuously monitor patients and act as an early warning system for patients who are at risk of having sepsis. At the same time, using the SERA algorithm in this manner could also potentially flag sepsis patients and thus prevent them from being overlooked; a situation that is common in stressful hospital settings where physicians have to continually make multiple medical decisions and judgment while managing high patient loads. I have provided details of how to practically set up this system in the methods section.

Through this study, I provide further empirical evidence for the value of data from EMR systems. Notably, I show that the data from EMR systems can be applied for more advanced healthcare applications. Unlike prior studies that utilized only structured EMR data, I demonstrated the value of using the vast amounts of unstructured data embedded in progress notes. For a typical EMR system, the volume of unstructured data dwarfs the volume of structured data,

and I believe that the value from unstructured data could be unlocked in a systematic fashion using NLP (see below). Using it for sepsis prediction as per my study is one case in point. NLP-enabled analysis would be naturally suited for managing other complex, equivocal, and critical clinical challenges as these challenges would benefit from expert insights embedded within EMR's unstructured data.

My study also extends the use of NLP in healthcare research. Within the medical informatics literature, NLP has been widely used to extract concepts, entities, and events, as well as their relations and associated attributes from free-text (Sheikhalishahi et al. 2019; Small and Medsker 2014), given that free-text is a common form of data in EMR systems (Jensen et al. 2017). My study suggests that the use of NLP could be extended by integrating NLP processed text with traditional machine learning tools to assist in diagnosing and early detection of diseases. My NLP results suggest that clinicians' progress notes, as well as other clinical reports (e.g., radiologist reports) stored in the EMR systems, may contain patterns of themes and topics that can be extracted and used productively for clinical analysis. Mining these notes are especially useful in clinical settings where the symptoms of ailments are not easily observed in structured data captured in the EMR. My study also highlights the value of mining and analyzing progress notes, notwithstanding the use of medical abbreviations and the idiosyncratic manner in which clinicians record their diagnoses and treatment plans.

Like all studies, there are limitations to my study. While I have validated and tested the models with an independent, test sample from a later time period, future research should conduct external validation in different hospital settings to improve the SERA algorithm and make my findings more robust.

In conclusion, if predictive algorithms similar to SERA algorithm can be designed to access valuable structured and unstructured data patient data continuously, hospital systems could potentially provide a 24 by 7, continual monitoring of patients' condition, thereby improving the ability for early sepsis detection and intervention. Though the SERA algorithm can achieve higher sensitivity and specificity rates compared to some physician's diagnosis and prior machine learning algorithms, I believe its primary role is to complement, not substitute, the clinical team's existing work. I further argue that other similar NLP-enabled algorithms could be developed to augment healthcare workers' knowledge and improve their decision making (Ohno-Machado 2011).

CHAPTER 4 EXPERT JUDGMENTS IN OVERRIDING ARTIFICIAL INTELLIGENCE: A HEALTHCARE CONTEXT

Introduction

Artificial Intelligence (AI) algorithms have it harder. Even as AI algorithms become more ubiquitous in business settings, Gartner predicts that 80% of business leaders would override alerts or prescriptions made by AI (Peters and Duncan 2020). This phenomenon of overriding AI algorithm alerts is one aspect of the challenges related to the intersection of humans and machines – an area which is gaining more research attention in recent years (Lee et al. 2015; Pasquale 2015; Rudin 2019). Not only does overriding AI algorithms limit their potential benefits, but there are also inherent risks associated with such behavior. Overriding of AI systems in heavy machinery, semi-autonomous vehicles, and equipment can lead to accidents and injuries as human judgment is fallible and biased (Cheatham et al. 2019). Similarly, recent studies have shown that the introduction of AI creates “breakages in social structures” (Elish and Watkins 2020) where experts are less likely to adhere to AI prescriptions – even those with critical consequences (Elish and Watkins 2020; Sendak et al. 2020). With the increasing application of AI technologies in the healthcare industry, electronic medical record (EMR) systems are now equipped with a variety of AI algorithms for diagnosis and alerts (Bardhan et al.

2015; Helm et al. 2016; Henry et al. 2015a; Montani and Striani 2019). These AI algorithm alerts are designed to warn clinicians when acceptable operating procedures are not followed or need to be executed and as such, they can help physicians avoid medical errors. For example, EMR systems have different prescriptive AI algorithm alerts that inform clinicians on medical allergy review, potential medical contraindications, and recommended clinical protocols (Ancker et al. 2017; Daniel et al. 2019). While the use of some AI algorithm alerts and other automated alerts have produced promising results (Kucher et al. 2005; Schedlbauer et al. 2009; Strom et al. 2010b), many other alerts continue to face significant challenges in actual implementation and use (Garg et al. 2005; Jaspers et al. 2011; Strom et al. 2010b; Wilson et al. 2015).

Like those settings discussed above, overriding AI alerts is also common in clinical settings, and physicians had wrongly dismissed certain AI algorithm alerts that led to significant health consequences (Russ et al. 2014; Shah et al. 2006). For example, Slight et al. (2018) estimated that in the United States alone, about 5.5 million were inappropriately overridden, resulting in approximately 196,600 adverse drug events. As such, recent work has pointed to the need to unpack the interactions between AI algorithms and experts by understanding how experts perceive, process, and act on the alerts from AI algorithms (Ahmad et al. 2018; Kelly et al. 2019; Lee et al. 2015). This is especially salient in the healthcare domain, given that physicians' overriding of AI algorithm alerts is a common widespread problem (Elish and Watkins 2020; Van Der Sijs et al. 2006). Research suggests that there is a need to understand

the conditions under which expert judgment (i.e., physicians) overrides AI algorithm alerts. Understanding those conditions will enable us to develop more effective ways to mitigate the overriding behavior, especially when they can potentially lead to dire consequences.

It is worth noting that the AI-based systems I am focusing on in this work is different from traditional clinical decision support systems (DSS) in the following two ways: first, traditional clinical DSS are usually knowledge-based and limited in scale (Chen et al. 2017). This is because traditional clinical DSS usually start in a top-down fashion from human knowledge and formalize and operationalize this knowledge into software. AI-based clinical systems, however, usually start in a bottom-up fashion from large volume of data of human activities, which are processed with AI techniques to extract useful information that can aid clinical decision. Therefore, AI-based systems are generally more powerful and can lead to more effective outcomes in clinical practice. Second, AI-based systems lack in transparency and explainability while traditional clinical DSS are generally easy for users to understand the basis for the advice and recommendations (Shortliffe and Sepúlveda 2018). As a result of the difference between traditional DSS and AI-based systems, it is necessary to examine expert judgement specifically in the context of AI-based systems although traditional DSS have been well explored in existing literatures.

Extant research suggests that the contextual factors are one set of conditions that influence how expert judgments are formed, that in turn,

impacts their behavior (Kahneman 2011). As contextual factors change, such as workflow or time constraints, they impact how the expert construe the AI algorithm alerts, which influences their beliefs, reliance, and adherence to the AI algorithm alerts (Van Der Sijs et al. 2006). Studies also have found that a physician's decision to override alerts is related to individual physician factors (e.g., physicians perceive alerts as less knowledgeable). While these empirical studies have suggested some salient conditions, there are not many large-scale studies that have empirically examined how these individual and contextual factors influence the interaction between AI algorithm alerts and physicians' behavior. Although it may be difficult to change physicians' mindset, an in-depth understanding of the influence of these factors on physician and AI algorithm alerts would give us theoretically grounded methods to modify these factors to improve physicians' adherence to AI algorithm alerts.

My study thus addresses this gap by answering the following research questions: "What are the factors that influence an expert's AI algorithm alert overriding behavior?" More specifically, in a healthcare context, "How do these factors influence physicians' decision to override AI algorithm alerts and what I can do to reduce such overriding behavior?" My field study of 1,152 physicians' responses to over 66,000 automated clinical alerts over a 22-month period provides theoretically grounded insights to *when* and *why* physicians override AI algorithm alerts. These findings improve my current theoretical understanding of human-AI interaction and have practical implications on how I design and deploy AI algorithm alerts. By tracing how and why physicians

decide to override AI algorithm alerts, my study's findings provide healthcare managers and clinicians a better grasp on how and when to trigger AI algorithm alerts and thereby create a more conducive environment for effective expert-AI interaction. The advantage of using a field study to examine physician's overriding behavior in actual clinical settings is that it is more realistic and provides concrete validation to my findings. However, this added realism comes at a cost, where, unlike controlled experiments, I am unable to entirely isolate the process – which is not within the scope of this study.

The remainder of the study is organized as follows: in the following section, I describe my field study's research context. Then, I discuss previous studies on expert judgment versus AI algorithms. I then propose the research model and hypotheses. Following, I describe the data, the model specification, and estimation, and I discuss the results as well as further analysis. Finally, I present my discussion and conclusions.

Research Context: Field Study

The field study site is a 700-bed hospital (HIT Hospital¹) that had achieved the highest level (Stage 7) on Healthcare Information and Management Systems Society's (HIMSS) EMR Adoption Model. Hospitals that are awarded the Analytics Stage 7 status are recognized as organizations who are successfully leveraging health information technology and data analytics to analyze clinical patterns to improve quality of care (Pettit 2013). As

¹ HIT hospital is a pseudonym.

part of its data analytics capabilities, HIT Hospital's medical informatics team had implemented AI-enabled CDSS within the EMR in the form of data-driven automated clinical alerts in clinical settings. Specifically, the hospital EMR system will fire off automated clinical alerts to physicians when predictive conditions that require attention from physicians are met.

AI-enabled CDSS and automated clinical alerts have been used extensively to influence physician behaviour, especially in point-of-care clinical information systems such as computerized provider order entry, bar-coded medication administration, pharmacy-based medication management and electronic prescribing systems (Baysari et al. 2017; McCoy et al. 2011). Increasingly, AI-enabled CDSS can also trigger prescriptive alerts to inform physicians of preventive clinical interventions based on evidence-based care guidelines (Lysaght et al. 2019). Such automated clinical alerts are designed to warn physicians about potential errors such as inappropriate dosing and drug-drug interactions, or advise them concerning clinical guidelines for specific medical diagnoses, thus either enabling or supporting them to make more accurate therapeutic decisions (Russ et al. 2014).

The automated clinical alert examined in this context is the indwelling catheter (IDC) alert. The main purpose of the IDC alert is to draw physicians' attention to patients with an indwelling urinary catheter predicted to be at risk of Catheter-Associated Urinary Tract Infection (CAUTI) given existing patient data, and the IDC alert will advise the removal of the catheter to prevent CAUTI. CAUTI is among the most common healthcare-associated infections

and accounts for approximately 10% of all infections in acute care hospitals (Magill et al. 2018). Given that about 15% of hospital patients will have an indwelling catheter at some point during their hospitalization, the risk of CAUTI is significant (Lo et al. 2008). Urinary tract infection can lead to numerous clinical complications and 13,000 deaths every year in the United States (Klevens et al. 2007).

The physician encounters the IDC alert when she opens the electronic medical record of a patient predicted to be at risk of CAUTI. The IDC alert will provide the physician with the clinical guideline: *“UTI causes 30% of all healthcare associated infections, and approximately 75% are related to urinary catheters”* and the physician has the option to either accept or dismiss the IDC alert. See Figure 9 for a screenshot of the alert. If the physician accepts the IDC alert, they will be prompted to indicate the clinical reasons why the patient’s catheter should remain in place or be removed. If the physician dismisses the IDC alert, then the IDC alert window will be immediately closed. The medical informatics team designed the IDC alert with the option to dismiss it so that physicians can move forward with their work and not compromise patient care. However, from the hospital perspective, when the physician dismisses the IDC alert, it represents a deviation from clinical best practices. Based on prevailing hospital guidelines, the IDC alert is presented to physicians only and not to other healthcare workers.

Note: Screenshot has been modified to remove patient identifiers.

Figure 9. IDC Alert Interface Snapshot

Theory and Hypothesis Development

Construal of AI Algorithm Alerts and Overriding Behavior

How do experts perceive AI algorithm alerts and their prescriptions?

The perceived level of abstractness versus the concreteness of an AI algorithm alert can be best explained with Construal Level Theory (CLT). CLT has been applied in many IS studies to explain phenomena such as user contribution motivations in online feedback forums (Phang et al. 2015), consumers' acceptance of online recommendations (Köhler et al. 2011), IT project risk management activities (Lee et al. 2019), information search and decision-making (Shani et al. 2009), and IT adoption tendencies (Shen and Chiou 2010).

According to CLT, an expert's construal of an AI algorithm alert is determined by the psychology distance of the alert. Psychological distance is a

“subjective experience that something is close to or far away from the self, here and now” (Trope and Liberman 2010, p. 440). The different ways in which an alert might be removed from that egocentric reference point (of self, here, and now) represent different aspects of distance. CLT literature has identified four dimensions of psychological distance: spatial, social distance, hypotheticality, and temporal (see Trope and Liberman (2010) for a comprehensive review). For example, an AI algorithm alert for a long-term investment opportunity – a temporally distant event – is associated with high psychological distance. Conversely, an alert in automated driving system to avoid a collision – a spatially near event – is associated with low psychological distance.

CLT posits that the greater a person’s psychological distance from a referent target (e.g., person, event, object), the more likely the target will invoke a high-level construal. Conversely, the smaller a person’s psychological distance is to a target, the more likely the target will invoke a low-level construal (Trope et al. 2007). High-level construal is viewed as “[more] abstract, [less] coherent, and superordinate mental representation” compared with low-level construal (Trope and Liberman 2010). High-level construal also has fewer details; for example, the word “illness” invokes a higher construal level than “skin cancer,” which invokes a low-level construal with more concrete details.

The construal level, in turn, affects an individual’s prediction of outcomes and evaluation of tasks (Trope et al. 2007). Prior literature suggests the following way by which construal level is associated with experts’

overriding behavior. Specifically, a high-level construal of target will lead to an increased reliance on high-level information (e.g., self-beliefs, generic trends and theories) and a corresponding decrease in reliance of low-level information (e.g., idiosyncratic events, irregular outcomes, situational and task characteristics) on prediction (Trope and Liberman 2010). The *overweighting of certainty* associated with high-level information and *underweighting of uncertainty* associated with low-level information generally leads to an upward bias in an individual's confidence in predicting psychologically distant events. Although one has less detailed information regarding high-level construal events, the asymmetric weighting between high-level versus low-level construal leads to the individual's unwarranted overconfidence in situations of high-level construal (Trope and Liberman 2010). This resonates with Nussbaum et al.'s (2006) study, where they found that individuals are more likely to be *more confident* in their judgments in higher compared to lower construal events. Applying this principle to AI algorithm alerts, it suggests that an individual who has greater confidence in his own judgment will typically reduce his reliance on warnings from AI algorithms, which results in the individual overriding AI algorithm alerts. In this way, extant literature suggests that a physician's construal level of AI algorithm alerts plays a role in their overriding behavior. Under conditions of high (low)-level construal, physicians are likely to be more (less) optimistic and overweigh positive (negative) aspects of the situation and thereby choose to override (accept) the alerts (Henderson et al. 2006; Nussbaum et al. 2006; Trope and Liberman 2010).

Environmental Characteristics and Overriding Behavior

The clinical environment in which the AI algorithm alerts appear influences how physicians construe the alerts. In my context, the IDC alert pops up when the physician accesses a patient's electronic medical records. Based on my interviews with HIT hospital physicians, alerts occur when the physician is directly examining the patient during the ward rounds, or when the physician is reviewing patient notes and test results in their offices. In both instances, the physician can choose to 1) adhere to the alert by reviewing the patient records and update the records in the EMR system or 2) to override the IDC alert and continue with their current task. One key difference in the context of the IDC alert is whether it occurs in the patient's physical presence. IDC alerts triggered in the physical presence of the patient represents a closer spatial distance, which I argue impacts the physician's IDC alert overriding behavior. Note that the review of the patient records can be performed in the same EMR system, which triggers the alert and can be done with or without the patient's presence.

Spatial distance represents the physical distance between the target and the focal individual who is experiencing the stimulus. Targets at a larger spatial distance are associated with higher psychological distance, and hence, an individual is more likely to perceive them as high-construal (Trope and Liberman 2010). Conversely, targets that occur in close physical proximity of the focal individual has low psychological distance. A physician who sees the IDC alert while reviewing a patient's record in the presence of the patient experiences a lower psychological distance (low spatial distance) as compared

to a physician who sees the IDC alert while reviewing the patient's electronic medical record in their office. When a patient is physically near to the physician, the physician experiences a closer and more immediate perception of the patient's condition with greater situational and specific details, which leads to the lower construal of the alert (Fujita et al. 2006).

As such, I argue that *where* the physician encounters the IDC alert in relation to the patient will lead to variances in the psychological distance of the alert. When the IDC alert is presented to a physician, it warns the physician about a potential adverse event—the possibility of CAUTI infection—that might occur to the patient. Given the many potential reasons why the patient may or may not require the indwelling catheter and the probability of the patient acquiring CAUTI because of the catheter, physicians have the option to either dismiss or accept the IDC alert. From my earlier arguments, however, it has been established that high-level construal is likely to lead to over-confidence, increased optimism, and less cautious predictions; hence, I posit that:

H1 – Physical Environment Overriding Hypothesis: AI algorithm alerts appearing in the physical absence of the patient are more likely to be overridden compared to alerts appearing in the physical presence of the patient.

Alert Exposure and Overriding Behavior

As healthcare systems directly affect clinical outcomes with life-or-death consequences, AI algorithms for early detection of critical illnesses are

designed to maximize *specificity* (i.e., true positive rate) by capturing a large proportion of positive cases at the expense of lower *sensitivity* (i.e., with higher false positive rate). Because this approach tends to err on the side of caution, AI algorithm alerts tend to lead to more frequent false alarms.

Extant literature suggests that when physicians experience a high volume of AI algorithm alerts, they may become desensitized to the alert over time (Cash 2009; Horsky et al. 2005; Shah et al. 2006). More importantly, research suggests that physicians who encounter a high rate of *non-critical* alerts inadvertently suffer from false alarm or Aesop's "cry wolf" effect (Breznitz 1984; Jones 2014). In other words, the repeated exposure of alarms without adverse events produces a downward bias on one's perceived probability of an adverse event occurring for subsequent alarms (Breznitz 1985). Applied to the IDC alert context, the "cry wolf" effect may reduce the credibility of subsequent IDC alerts for the physician. This cry wolf effect can be better understood by the concept of hypotheticality of events – one of the four dimensions of psychological distance discussed in CLT.

Individuals assess the hypotheticality of an event as the perceived probability of the occurrence of the event (Trope and Liberman 2010). Improbable events are more psychologically distant to the observer compared to probable events; thus, the probability of event occurrence is negatively correlated to the psychological distance (Liberman et al. 2007; Trope et al. 2007). Put in another way, Wakslak et al. (2006) found that independent of spatial and social distance, a task which has a high perceived probability of

occurrence will have a low psychological distance. Following the CLT, these low psychological distance tasks are construed as low-level construal and are described in richer details compared to other high psychological distance tasks.

In line with CLT and extant literature findings on cry wolf effect, I propose that in my context, the more a physician is exposed to IDC alerts for the same patient that do not lead to adverse events will cause the physician to perceive subsequent IDC alerts to be of higher hypotheticality, which leads to high-level construal of these subsequent IDC alerts. In turn, the high-level construal of these subsequent IDC alerts is likely to lead to over-confidence, increased optimism, and less cautious predictions. Thus, I posit that:

H2 – “Cry Wolf” Overriding Hypothesis: A larger number of prior AI algorithm alerts encountered by a specific physician for a particular patient is associated with a higher likelihood of the physician overriding the subsequent alert.

Domain Expertise and Overriding Behavior

The medical profession consists of different medical specialties to ensure specialization of skills among physicians. Although all physicians attending to hospital patients are qualified experts, there is still heterogeneity in their expertise (Wang and Cole 2016). Like other professions (e.g., lawyers and engineers), physicians identify themselves with their respective specialty ingroup such as cardiologists, oncologists, and neurologists. As such, physicians typically identify themselves more with issues relating to their ingroup medical specialty compared to the issues of outgroup medical

specialties. For example, a neurosurgeon will relate more with clinical neurological issues compared to clinical cardiac issues. As such, I expect physicians to experience varying *social distances* depending on how related their own referent medical specialty is to the alert's medical context (Amit et al. 2009a; Trope et al. 2007).

Social distance is the perceived distinction between self and others, familiar versus unfamiliar, ingroup members versus outgroup members, and so on (Liberman et al. 2007). Specifically, in this context, the social distance results from ingroup-outgroup medical specialty comparisons. Hence, the ingroup represents physicians whose specialty is closely related to the condition of the IDC alert while the outgroup represents physicians whose specialty not related to the condition of the IDC alert.

Notably, social distance can be associated with inanimate objects (e.g., soccer balls, coins (Amit et al. 2009a)), pictures (Amit et al. 2009b), and words (Amit et al. 2009a). Thus, I apply the social distance arguments for physicians to AI algorithm alerts where I note that alerts, like other inanimate objects, are associated with specific groups, ranging from general clinical alerts to alerts that are closely linked to specific medical specialties or ailments. General alerts may include the patient's short-term readmission risk alerts or medication duplication alerts that are generalized, easily understood, and as such are presented to a wide range of healthcare workers. On the other hand, there are alerts specific to medical specialties or ailments, e.g., alerts to prompt for an additional cancer screening test, which require more in-depth medical-specific

knowledge of the issues surrounding the alert. These are usually presented to physicians working in that specialty.

Put together, I argue that physicians are more likely to construe alerts associated with clinical matters close to their medical specialty as low-level construal with more details. Conversely, physicians are more likely to construe alerts associated with clinical matters not related to their medical specialty as high-level construal with fewer details. Thus, when IDC alerts are presented to a physician whose medical specialty is not related to the CAUTI domain, the physician will perceive the alert as high-level construal, which is likely to lead to over-confidence, heighten optimism, and less conservative predictions. Thus, I posit that:

H3 – Domain Expertise Overriding Hypothesis: AI algorithm alerts associated with clinical matters that are less related to the physician’s medical specialty are more likely to be overridden than alerts associated with clinical matters that are related to the physician’s medical specialty.

Notably, an alternative explanation here is that the physician exposed to the IDC alert might not be responsible for the clinical specialty that is presented in the alert and hence dismisses the alert. To address this alternative explanation, I note that in HIT Hospital, the IDC alert has to be addressed by ***all*** physicians regardless of their specialty. This is because the IDC alert seeks to reduce the risk of patient infection – which is the responsibility of ***all*** physicians and not just infectious disease specialists.

Mitigating “Cry Wolf” Overriding Behavior

Given such an environment of high alert exposure and potential “cry wolf” effect as hypothesized earlier, physicians may have high tendencies to override these alerts. A practical follow up question is exploring boundary conditions that this “cry wolf” effect may be mitigated.

Earlier, I consider how environmental or expert characteristics can ameliorate the way physicians construe the hypotheticality of AI algorithm alerts, and hopefully leading to lower construal of these alerts. A logical choice would be to consider the environmental context in which the alert appears as this is within the control of alert designers and is a relatively easy intervention to implement (compared to expert characteristic). This notion is also found in the CLT literature. Trope and Liberman (2010) argued that the four dimensions of the psychological distance vary in terms of controllability. *Environmental* spatial distance is the easiest to control, manipulate and observe (i.e., spatial distance varies as one moves closer or further from the object). Social distance (expert characteristics) and hypotheticality (alert exposure) are only partially controllable. Further, CLT literature has suggested that while the four dimensions of psychological distance are correlated and can simultaneously impact each other (Pronin et al. 2008; Wakslak et al. 2008), the relative impact of each dimension on the level of construal varies with some dimensions being more significant than others. Trope and Liberman (2010), found that *environmental* spatial distance is more significant and fundamental than the three other dimensions of distance as it is among the first to be learned (during

childhood), more perceivable (e.g. easier for one to perceive an event that occurred nearby than to assess the probability of an event occurring), more unequivocal, and easiest to communicate to the observer. Spatial distance is also shown to be more fundamental, as it can influence other dimensions such as temporal distance in an individuals' perception (Boroditsky 2000; Boroditsky and Ramscar 2002).

Given that environmental spatial distance is the most controllable and has a more significant impact on other dimensions of psychological distance, I propose that the environmental context in which the AI algorithm alert appears can potentially mitigate the “cry wolf” effect, which would lower the level of alert construal and lead to less overriding behavior. This logic is based on applying the CLT literature and juxtaposing hypothesis 1 and hypothesis 2, i.e., I propose that a spatial distance condition of alert presentation is likely to moderate an increased hypotheticality effect caused by high alert exposure rate. Formally, I posit:

***H4 - Mitigating the “Cry Wolf” Effect Hypothesis:** A larger number of prior AI algorithm alerts in the physical presence of the patient is associated with a lower likelihood of the physician overriding the subsequent alert.*

Data Collection and Methodology

I collect the EMR log data for the automated clinical alerts and physicians' actions in the hospital from March 2017 till December 2018. During this period, I record a total of 66,230 alert events and this sample represents all instances in which the alert was triggered. The alerts originate

from 1,877 patients, with each patient triggering an average of about 35 alerts during his/her hospitalization stay. This dataset tracks a total of 1,152 different physicians and their actions after receiving the automated clinical alert. Table 6 presents the distribution of the type of physicians within the sample. In my dataset, only 5,849 alerts (8.83%) were accepted by physicians, concurring with the existing literatures that alert acceptance rate is very low (Van Der Sijs et al. 2006). A large majority of alerts (91.17%) were dismissed by the physicians without further action. In addition to collecting the alert events, I also captured additional information surrounding the alerts. The additional information includes details about the physician receiving the alert, the temporal history of the alerts received for the patient, basic patient and patient medical details, etc.

Table 6. Distribution of Physician Type

Physician ranking	Freq.	Percent	Cum.
Intern (first year post-medical school training)	163	14.15	14.15
Resident (two to 6 years)	584	50.69	64.84
Fellow (complete residency and in specialist training)	161	13.98	78.82
Attending (specialists)	244	21.18	100
Total	1,152	100	

In the following parts of this section, I outline the unit of analysis, the variables as well as the model specification employed to test the hypotheses.

Key Variable Measurement and Description

The unit of analysis for this study is each automated clinical alert, and the dependent variable is a binary indicator, *override*, that measures whether the physician accepts or dismisses the alert.

I measured various key independent variables related to the hypothesized individual and contextual factors. To operationalize the environmental factor, I used a flag indicator, *ward*, to represent if the alert was presented in the ward (in the presence of the patient) or out of the ward. To operationalize alert exposure, I captured the physicians' prior exposure with alerts in the following way. Given that various physicians can consult a patient and physicians can have multiple patients, I first computed patient-physician (P_C) dyads². Having this dyad will allow us to precisely measure each physician's alert exposure concerning a particular patient. I computed P_C_{total} , which is the *total* number of alerts a particular physician received for a particular patient, to measure her alert exposure. More alerts received by a particular physician for a particular patient during his hospitalization stay increases the “cry wolf” effect (increases hypotheticality).

Physicians may choose to override alerts because they trust more in their own clinical knowledge, and this is typically seen in more experienced physicians (Magnus et al. 2002; Taylor and Tamblyn 2004). To represent the experience and seniority of the physician, I proxied the physician's expertise level using the rank of the physician, C_rank , which captures the seniority of the physician (i.e., Intern, Resident, Fellow, and Attending).

² I use C to represent physician (Clinician) to create clarity in my nomenclature and not confuse it with Patient, which is denoted as P .

To operationalize domain expertise, I first captured the medical specialty of the physician receiving the alert. Given that the automated alert relates to CAUTI, medical specialties related to infectious diseases and urology are specialties that are closely related to the nature of the alert. As reasoned in the H4, specialties that are more related to the medical context of the alert will have lower social distance than specialties less related to the context of the alert. I created a binary variable, *infectious*, to capture whether the medical specialty is closely related to the CAUTI alerts. Based on my interviews with physicians, I chose three specialties, (i.e., IV antibiotic, urology, and infectious disease), as specialties that are related to CAUTI, and the rest of the specialties (e.g., neurology, spine surgery) as non-infectious specialties.

To measure how environmental context can mitigate the “cry wolf” effect, I computed P_C_{ward} , which represents the total number of alerts received in the ward for the patient-physician pair. One unit increase of this measure has two effects on physicians: 1) it will increase the alert exposure for the physician and thus reinforce the “cry wolf” effect; 2) it will alert the physician a low low-spatial distance condition.

Controlling for Alternative Explanations

There are various alternative reasons documented in the literature that argues for the overriding of clinical alerts. Here, I outline these alternative explanations and introduce the proxies used to control for these other alternative explanations. I use these proxies as controls in my model specification.

Habitual overriding of alerts. The medical literature suggests that when physicians are overexposed to AI alerts, they tend to develop a habit of dismissing the alerts without paying additional consideration to these alerts (Baysari et al. 2017). The authors argue that physicians who are exposed to the most alerts – are most likely to develop habits in overriding these alerts. To control for possible habitual overriding of alerts, I compute the total number of alerts a physician historically overrides (*override_ct*). Physicians who have developed a habit of overriding are likely to be represented by high override count. I took the natural logarithm of the total number of overrides as this variable is highly skewed.

Physician fatigue or lack of time. Research shows that many physicians override alerts because of alert fatigue due to the high volume of repeated exposure to similar alerts that causes cognitive overload in physicians, who eventually stop responding to these alerts (Horsky et al. 2005; Shah et al. 2006). At the same time, physicians also override alerts because alerts occur when they are busy (Overhage et al. 1997) or when they interrupt the physicians' workflow (Ash et al. 2004). I controlled the number of unique patients with indwelling catheters the particular physician encountered for the particular workday (*CPatNum*) to control for physician's exposure to patients with similar ailments as well as a control for their workload for the day. Further, I recording the day of the week to control for duty shifting (Diestre et al. 2019), the month of the year to control for seasonality that impacts general patient load.

Learning effects and prior experience of alerts. As physicians consult a large number of patients, they have heterogeneous experiences and exposure to these alerts. To control prior experience or exposure of alerts that are *not related* to a particular patient, I computed physician-level controls for alert exposure rate i.e., the total number of IDC alerts a particular physician has received within the hospital since the beginning of the IDC alert program (C_{total}). This variable controls for physician's familiarity with the system. I took the natural logarithm of the total number of IDC alerts as this variable is highly skewed.

Repeat consultation for the day. At times, a physician might consult a patient for multiple times within a single day. Given that dismissed alerts will all keep reappearing each time the physician accesses the patient's medical records, to control for repeated consultations for the same day, I use a binary flag P_C_{first} , to record the first alert for that day that appear to a particular patient-physician pair. Additionally, I also control for number of days since the physician saw the first alert of the patient with the variable, *day_lag*.

Patient's condition. The patient's condition might influence the physician's decision to override an alert. I use the length of stay (*LOS*) as control for severity of the patient's ailment (Yeow and Goh 2015) and to

capture the type of ailment the patient is suffering, I captured the department in which the patient is warded (e.g. cardiology or orthopedic)³.

Table 7. Definition and description of variables

Variable	Description	Definition
<i>accept</i>	Respond to alerts	Whether the physician accepted the alert
<i>ward</i>	Spatial context	Whether the alert is shown during the ward round
<i>infectious</i>	Infectious diseases	Whether the patient is in an infectious specialty
<i>P_C_{total}</i>	No. of alerts for the patient	The number of alerts encountered by the physician for the particular patient
<i>P_C_{ward}</i>	No. of alerts for the patient during ward round	The number of alerts encountered by the physician for the particular patient during the ward round
<i>C_{total}</i>	No. of alerts encountered by the physician	The total number of alerts encountered by the physician
<i>C_{PatNum}</i>	No. of patients encountered by the physician on the day	The number of unique patients encountered by the physician on the day
<i>day_lag</i>	No. of days since the first alert for the patient	The number of days since the physician saw the first alert of the patient
<i>C_rank</i>	Physician level	The expertise level of the physician
<i>age</i>	Patient age	The age of the patient
<i>gender</i>	Patient gender	0 for female, and 1 for male
<i>race</i>	Race of patient	Dummy variables for different races
<i>los</i>	Patient length-of-stay	The length-of-stay for the patient in this hospitalization encounter
<i>day</i>	Day of the week	Monday through Sunday as dummy variables
<i>dept</i>	Department	The ward the patient is in. Different specialties have different wards
<i>spec</i>	Specialty	The speciality of the physician who receives the alert
<i>month</i>	Month	Jan through Dec, the month in which the alert appears

The descriptive statistics are shown in Table 8.

Table 8. Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
<i>accept</i>	66,230	0.088	0.284	0	1

³ Both the medical specialty of the physician and department of the Patient's ward are required as a patient can be placed in one department ward and be attended by physicians from different medical specialties. E.g., a geriatric patient in a cardiology department ward may be attended to by both a cardiologist (specialty) and a geriatrician (specialty).

ward	66,230	0.456	0.498	0	1
infectious	66,230	0.057	0.232	0	1
P_C_{total}	66,230	8.208	11.949	1	120
P_C_{ward}	66,230	3.874	5.986	0	60
P_C_{ward}²	66,230	50.833	178.879	0	3600
P_C_{nonward}	66,230	4.335	6.817	0	79
C_{total}	66,230	84.764	93.791	1	669
P_C_{today}	66,230	2.939	2.728	1	32
C_{PatNum}	66,230	1.303	0.620	1	7
day_lag	66,230	3.167	8.822	0	240
C_rank=1	66,230	0.611	0.488	0	1
C_rank=2	66,230	0.057	0.232	0	1
C_rank=3	66,230	0.082	0.275	0	1
age	66,230	67.894	16.120	16	105
gender	66,230	0.558	0.497	0	1
los	66,230	38.464	48.467	1	357

Note: C_rank = 0 if physician level is an intern (first year post-medical school), C_rank = 1 if physician level is a resident (2 to 6 years of work experience), C_rank = 2 if physician level is a fellow (completed residency and in specialist training), and C_rank = 3 if physician level is an attending physician (specialists).

Model Specification and Estimation

I propose that physician, i , for patient, j , decision-making process, D , of accepting (versus dismissing) the alerts, is a function of the psychological distance of the alerts, B , physician characteristics, Ψ , patient characteristics, A , and other exogenous variables, Ω , where the probability of accepting the alert, $P(D)$, is expressed as:

$$P(D) = f(B, \Psi, A, \Omega) \quad (3)$$

Since my dependent variable is a binary outcome of whether the physician accepts the alert, I estimate the following equation using fixed-effects logit regression:

$$\begin{aligned} \text{Logit}[P(D_{ij})] = & \beta_1 \text{ward}_{ij} + \beta_2 \text{infectious}_{ij} + \beta_3 P_C_{total,ij} + \beta_4 P_C_{ward,ij} \\ & + \psi_1 C_{total,i} + \psi_2 C_{PatNum,i} + \psi_3 C_rank_i + \lambda_1 \text{day_lag}_{ij} \\ & + \lambda_2 \text{age}_j + \lambda_3 \text{gender}_j + \lambda_4 \text{los}_j + \lambda_5 \text{race}_j + \omega_1 \text{dept}_j \\ & + \omega_2 \text{spec}_{ij} + \omega_5 \text{day}_{ij} + \omega_3 \text{mth}_{ij} + v_i + \epsilon_{ij} \end{aligned} \quad (4)$$

where β , ψ , λ , ω are parameter estimates of the variables represented by vectors B , Ψ , A , and Ω , respectively. During the data collection period, a total of 1,152 different physicians received an average of 59 alerts each. To account for unobservable heterogeneity across physicians, I include physician fixed effects (Wooldridge 2002), v_i . ϵ is the error term.

Results

In this section, I present the main results from my econometric models, several robustness checks, and some further analysis.

Main Results

Table 9 below presents the estimation results. Column 3 represents the fixed-effects logit model. For comparative purposes, I also present alternative estimations – column 1 (estimation without exogenous controls, Ω , and without physician fixed effects, v_i) and column 2 (estimation without physician fixed effects, v_i).

Table 9. Regression Results

Variable	Model (1) accept	Model (2) accept	Model (3) accept
<i>ward</i>	0.233*** (0.031)	0.200*** (0.032)	0.205*** (0.040)
<i>infectious</i>	0.445*** (0.051)	0.699*** (0.127)	0.601*** (0.187)
<i>P_C_{total}</i>	-0.016*** (0.004)	-0.024*** (0.004)	-0.021*** (0.006)
<i>P_C_{ward}</i>	-0.002 (0.008)	0.011 (0.008)	0.035*** (0.012)

<i>C_{total}</i>	-0.005*** (0.000)	-0.002*** (0.000)	-0.004*** (0.000)
<i>C_{PatNum}</i>	0.070*** (0.025)	0.042 (0.026)	-0.023 (0.033)
<i>C_{rank=1}</i>	0.533*** (0.040)	0.171*** (0.043)	Invariant ⁴
<i>C_{rank=2}</i>	0.521*** (0.064)	0.314*** (0.068)	Invariant
<i>C_{rank=3}</i>	0.812*** (0.054)	0.755*** (0.057)	Invariant
<i>day_{lag}</i>	0.017*** (0.002) (0.054)	0.016*** (0.002) (0.057)	0.008*** (0.003)
<i>age</i>	-0.000 (0.001)	0.002 (0.001)	0.000 (0.001)
<i>gender</i>	0.039 (0.029)	0.053* (0.032)	0.098** (0.041)
<i>los</i>	-0.003*** (0.000)	-0.002*** (0.001)	-0.001 (0.001)
<i>race dummy</i>	suppressed for brevity ⁵	suppressed for brevity	suppressed for brevity
<i>dept dummy</i>	No	suppressed for brevity	suppressed for brevity
<i>spec dummy</i>	No	suppressed for brevity	suppressed for brevity
<i>day dummy</i>	No	suppressed for brevity	suppressed for brevity
<i>month dummy</i>	No	suppressed for brevity	suppressed for brevity
<i>Physician fixed effects</i>	No	No	Yes
Constant	-2.641*** (0.089)	-1.509** (0.703)	
<i>AIC</i>	37890.630	34650.573	21083.347
<i>BIC</i>	38063.495	35624.080	21990.178
<i>Log likelihood</i>	-18926.315	-17218.286	-10438.674

Note: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

From Model (3) of Table 9, I can see that the coefficient for the variable *ward* ($\beta_1 = 0.205, p < 0.01$) is positive and significant. This suggests that alerts that are presented within the ward (low spatial distance) are more likely to be

⁴ Physicians' rank did not vary during this time period. Fixed effects estimation omits invariant variables within the group.

⁵ Due to the large number of dummy categories, estimates are suppressed for brevity. Full results can be requested from the authors.

accepted and not dismissed, supporting H1. Additionally, the variable, *infectious* is positive and significant ($\beta_2 = 0.601, p < 0.01$), indicating that physicians whose specialties are related to the nature of the alert (low social distance) are more likely to accept the alerts instead of dismissing them, thus supporting H2.

In terms of the “cry wolf” effect, I observed that the relationship between the total number of alerts the physician receives for a particular patient (P_C_{total}) and their acceptance of the alert is negative and significant ($\beta_3 = -0.021, p < 0.01$). This finding suggests that all else being constant, incremental instances of alerts (increased hypotheticality) might be counter-productive in promoting affirmative action by physicians. Further, I also observed that the total number of alerts a physician receives in general (C_{total}) is negatively associated with affirmative action ($\psi_1 = -0.004, p < 0.01$). This suggests that the “cry wolf” effect is probably more prevalent and possibly extends to other alerts occurring beyond those in each patient-physician pair. These findings are in line with H3.

Finally, I also found support for H4. The total number of alerts received by a physician for a particular patient during the ward round is positively associated with the likelihood that the physician accepts the alert ($\beta_4 = 0.035, p < 0.01$). This finding confirms my arguments on how to mitigate the “cry wolf” effect caused by a high volume of automated alerts. Here, I observe that a high volume of alerts presented in the context of low spatial distance can mitigate the effects of increased hypotheticality.

Robustness Checks and Further Analysis

In this section, I discuss alternative operationalizations, different estimations, robustness checks, and additional analyses that were conducted to support and extend my earlier findings.

Alternative Measurements of Hypotheticality

Chandran and Menon (2004) argued that individuals' risk perception is influenced by how I aggregate events. They show that when risky events are framed and processed in a disaggregated fashion (i.e., daily), they are perceived to be riskier. For example, stating that "every *day*, two people die from heart disease" will cause individuals to view heart disease as a more serious problem compared to framing the statement as "every *year*, 730 people die from heart disease." In my earlier estimation to test the "cry wolf" hypothesis, I use the total number of alerts a physician receives for a particular patient (P_C_{total}) as a monotonic measure for hypotheticality. This is a cumulative value and represents all the alerts starting from the day the patient is warded. Empirically, it is not possible to know what mental frames were employed by physicians when they assess the earlier alerts. Do physicians construe the alerts received in an aggregated fashion for each patient, or do physicians only consider alerts received within the same workday for the patient? To take into consideration alternative mental frames used by physicians, I compute the total number of alerts received by each patient-physician pair for a *particular* workday, P_C_{today} , as an alternative operationalization for hypotheticality. Results from

this alternative operationalization present similar support for H3 (see Models 4 and 5, Table 10). In Model 4, I observe that with a daily mental frame for the number of alerts (P_C_{today}), physicians are more likely to dismiss an alert with an increasing daily number of alerts for a particular patient ($\beta_6 = -0.055, p < 0.05$). This effect is also consistent for Model 5, where I test both aggregated ($P_C_{total}, \beta_2 = -0.016, p < 0.01$) as well as daily operationalization ($P_C_{today}, \beta_6 = -0.053, p < 0.01$).

In the main model estimation, I use the total number of in-ward alerts for each patient-physician dyad to operationalize the mitigating effect of low spatial distance on hypotheticality, and I use the number of *all* alerts for each patient-physician dyad to operationalize the hypotheticality of the alerts. An alternative operationalization for hypotheticality will be to use the number of non-ward alerts, $P_C_{nonward}$. The presentation of these alerts to the physicians occurs in non-ward locations that are physically distant from the patient, thereby heightening hypotheticality. P_C_{total} is the sum of $P_C_{nonward}$ and P_C_{ward} ; hence, the use of all three variables $P_C_{total}, P_C_{ward}, P_C_{nonward}$ simultaneously in the earlier specification is not possible due to full rank violation. Results from Model 6, as shown in Table 10, provide additional support for H1 and H4. Using a different operationalization for hypotheticality ($P_C_{nonward}$) I observed that increased hypotheticality leads to an increased likelihood of alert dismissal ($P_C_{nonward}$, coef. = -0.016, $p < 0.01$. H3 supported). Additionally, an increase in low spatial distance alerts tempers the increased hypotheticality of the alerts ($P_C_{ward}, \beta_4 = 0.023, p < 0.01$).

Diminishing Returns to Scale of Spatial Distance

Hypothesis 4 shows that spatial distance mitigates the impact of hypotheticality and defuses the “cry wolf” effect. A natural follow-up question for this effect is whether the mitigating impact of spatial distance is effective across all ranges of values. Prior literature suggests that psychological distance, especially spatial distance, follows a concave function and experiences diminishing returns to scale (Trope and Liberman 2010). Similarly, Holyoak and Mah (1982) argued that individuals are more sensitive to psychological distance nearer to the egocentric self, and this sensitivity decreases over larger value ranges. The arguments of diminishing returns of psychological distance are in line with the Weber-Fechner Law (Trope and Liberman 2010; Zauberman et al. 2009) and empirically, I examine if the mitigating impact of spatial distance experiences similar diminishing rates of return. For example, I observe that more in-ward alerts are associated with more affirmative action by the physicians accepting the alerts, but does this positive relationship diminish if physicians encounter excessive instances of in-ward alerts for the specific patient?

To operationalize a possible change in returns to scale, I model a quadratic relationship between in-ward alert encounters and physician actions using a squared-term, as shown in the following model:

$$\begin{aligned} \text{Logit}[P(D_{ij})] = & \beta_1 \text{ward}_{ij} + \beta_2 \text{infectious}_{ij} + \beta_3 P_C_{total,ij} + \beta_4 P_C_{ward,ij} \\ & + \beta_5 P_C_{ward,ij}^2 + \beta_6 P_C_{today,ij} + \psi_1 C_{total,i} + \psi_2 C_{PatNum,i} \\ & + \lambda_1 \text{day_lag}_{ij} + \lambda_2 \text{age}_j + \lambda_3 \text{gender}_j + \lambda_4 \text{los}_j + \lambda_5 \text{race}_j \\ & + \omega_1 \text{dept}_j + \omega_2 \text{spec}_{ij} + \omega_3 \text{day}_{ij} + \omega_4 \text{mth}_{ij} + v_i + \epsilon_{ij} \end{aligned} \quad (5)$$

where $P_C_{ward}^2$ represents the squared term and the rest of the terms are the same as in Equation (2).

Model 8 in Table 10 presents the estimation of equation 3 above, and Model 7 is an alternative estimation of equation (3) where I use P_C_{total} , as presented in my original specification of equation (2), in place of $P_C_{nonward}$. Both Models 7 and 8 showed a negative quadratic relationship ($P_C_{ward}^2$, $\beta_5 = -0.002$, $p < 0.01$), suggesting that although increased instances of in-ward alerts are associated with increased acceptance of the alerts, this relationship experiences diminishing returns to scale where *excessive* in-ward alerts are associated with an increased rate of dismissal.

Table 10. Robustness Checks and Further Analysis

Variables	Model (4) accept	Model (5) accept	Model (6) accept	Model (7) accept	Model (8) accept
<i>Ward</i>	0.186*** (0.038)	0.153*** (0.040)	0.153*** (0.040)	0.088** (0.042)	0.088** (0.042)
<i>Infectious</i>	0.600*** (0.188)	0.584*** (0.188)	0.584*** (0.188)	0.588*** (0.188)	0.588*** (0.188)
P_C_{total}		-0.016*** (0.006)		-0.013** (0.006)	
P_C_{ward}	0.012** (0.005)	0.039*** (0.012)	0.023*** (0.007)	0.091*** (0.015)	0.078*** (0.012)
$P_C_{ward}^2$				-0.002*** (0.000)	-0.002*** (0.000)
P_C_{today}	-0.055*** (0.009)	-0.053*** (0.010)	-0.053*** (0.010)	-0.072*** (0.010)	-0.072*** (0.010)
$P_C_{nonward}$			-0.016*** (0.006)		-0.013** (0.006)
C_{total}	-0.004*** (0.000)	-0.004*** (0.000)	-0.004*** (0.000)	-0.004*** (0.000)	-0.004*** (0.000)
C_{PatNum}	-0.005 (0.034)	-0.006 (0.034)	-0.006 (0.034)	-0.009 (0.034)	-0.009 (0.034)

<i>day_lag</i>	0.005*	0.006**	0.006**	0.004	0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
<i>age</i>	0.000	0.000	0.000	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
<i>gender</i>	0.089**	0.091**	0.091**	0.091**	0.091**
	(0.041)	(0.042)	(0.042)	(0.042)	(0.042)
<i>los</i>	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
<i>race dummy</i>	suppressed for brevity	suppressed for brevity	suppressed for brevity	suppressed for brevity	suppressed for brevity
<i>dept dummy</i>	suppressed for brevity	suppressed for brevity	suppressed for brevity	suppressed for brevity	suppressed for brevity
<i>spec dummy</i>	suppressed for brevity	suppressed for brevity	suppressed for brevity	suppressed for brevity	suppressed for brevity
<i>day dummy</i>	suppressed for brevity	suppressed for brevity	suppressed for brevity	suppressed for brevity	suppressed for brevity
<i>month dummy</i>	suppressed for brevity	suppressed for brevity	suppressed for brevity	suppressed for brevity	suppressed for brevity
<i>Physician fixed effects</i>	Yes	Yes	Yes	Yes	Yes
<i>AIC</i>	21058.223	21053.466	21053.466	21021.147	21021.147
<i>BIC</i>	21965.054	21969.101	21969.101	21945.586	21945.586
<i>Log Likelihood</i>	-10426.112	-10422.733	-10422.733	-10405.574	-10405.574

Note: Standard errors in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Discussion and Conclusions

By drawing on CLT, my study provides a novel perspective to why physicians override automated clinical alerts. My study investigated how differences in psychological distance, which influence the physicians' construal of automated clinical alerts, explain why physicians tend to override alerts in practice. As hypothesized, my findings show that *low* spatial distance and *low* social distance is associated with a *higher* likelihood of physicians accepting clinical alerts, while *increased* hypotheticality is associated with a *lower* likelihood of physicians accepting clinical alerts. Specifically, my study shows that every additional alert a physician receives during non-ward time is associated with a 2% drop in likelihood to accept the subsequent alerts for the

patient. Given that an ICU patient, on average, sets off a total of 771 alerts per day (Jones 2014), and in my study, an average patient sets off 35 *CAUTI* alerts, a 2% change in likelihood per additional alert is cumulatively significant. I also find that a low level of spatial distance can mitigate the effect of increased hypotheticality on alert acceptance to some extent. Empirically, I observed that every additional alert a physician receives in the ward for the patient is associated with a 4% increase in the likelihood of accepting the subsequent alert. These findings are important as more AI-enabled CDSS and automated clinical alerts are being introduced and implemented in healthcare organizations. While it is critical that I am able to build more explainable and dynamic algorithms to assist with physicians' work of providing evidence-based care, it is also equally important that physicians are willing and able to attend to these inputs and interventions as part of their daily work routines. My theoretically grounded findings provide valuable insight as to how contextual and individual factors may influence those decisions and thereby help in the development of better ways of alerting physicians. As such, my study has several important theoretical and practical implications.

Theoretical Contributions

My findings provide a theoretically grounded explanation about the nature of a clinician's decision process towards clinical alerts, which is under-explored in the existing literature. Specifically, using CLT, I unpack how contextual factors (i.e., space), individual factors (i.e., physician specialty), and alert design (i.e., frequency of alerts) impact physicians' perception (i.e.,

construal of the alert). In turn, I show how the level of construal through various psychological biases (overconfidence, heightened optimism, shallow processing, etc.) may shape their decision with respect to the alert.

I contribute to healthcare IT research by providing a theoretical framework to explain physicians' reactions to AI-based alerts. This framework not only offers insights to the cognitive and psychological processes of physicians, but also how contextual dimensions are theoretically linked to individual mental construal in influencing their ultimate decision (whether to override or not override the alerts). Notably, my study found how spatial distance (a more fundamental and controllable psychological distance) could mitigate the more equivocal psychological distance of hypotheticality. This provides one specific case of how one dimension of psychological distance works in conjunction with another dimension. At the same time, I also established the boundary condition for their joint effects in that the mitigating effect of spatial distance on hypotheticality suffers from diminishing returns to scale such that up to a certain point excessive in-ward alerts will still lead to dismissal of alerts. I argue that future research could consider other ways perceptions of alerts are shaped, as well as the decision biases that may influence the physician's overriding of alerts. My findings on the relationship between the physicians' psychological distance from the patient and how this impacts their follow-up actions can be generalized to other AI-enabled CDSS, which have been increasingly common in my healthcare practice.

I also contribute to existing CLT research in several ways. First, I extend the application of CLT to a new empirical context (i.e., healthcare IT) to understand how an expert user perceives automated clinical alerts, as well as their subsequent decisions vis-à-vis the alerts. My findings show that CLT's main principles are useful in explaining my focal phenomenon (Liberman et al. 2002; Liberman et al. 2007; Trope and Liberman 2010). This extension is also above and beyond existing IS research that applied CLT in online recommendations or IT adoption (Köhler et al. 2011; Shen and Chiou 2010), as most of these contexts do not involve expert users or decisions that have significant consequences. My study shows that CLT principles are equally applicable to intensive, critical decision-making scenarios, where decisions may have severe implications for a patient's health. My context of physicians and automated clinical alerts is also novel for CLT, in that of the four dimensions of psychological distance, only three were applicable (as explained above, the temporal distance dimension was less relevant). As such, CLT research may consider the relative applicability of different psychological dimensions in the healthcare context.

Second, while CLT has looked at how social distance may apply to the construal of inanimate objects, my study shows how it may also be applicable to digital artifacts like automated clinical alerts (Eisenhardt and Graebner 2007). Whereas a small stream of research has looked at how the digital medium shapes the level of construal among users (Kaufman and Flanagan 2016), my study shifts the focus to how digital artifacts are targets of construal,

given their association with other dimensions of psychological distance like social distance. Thus, CLT research should consider extending this new angle to understand how other digital artifacts are linked with different dimensions of psychological distance.

Lastly, my study contributes to CLT research in that my study is based on actual field data, while most prior CLT studies are lab-based, controlled studies with student subjects. The field study allows us to collect data relating to different psychological distances experienced by physicians and associate them with the corresponding actual medical decisions made in real-life medical situations. Compared to lab-based studies with student subjects that employ hypothetical scenarios, my results show how CLT can be applied in user decision making in real-life conditions.

Contribution to Medical Practice and EMR design

My theoretical insights provide important considerations for how I can design a more effective alerting system and other AI-enabled CDSS. While prior studies focus more on improving the design artifact of the alerts (e.g., improving specificity of alerts, designing user interface that are unambiguous (Glassman et al. 2002; Kawamoto et al. 2005; Van Der Sijs et al. 2006)), my findings show that focusing only on the artifact design is insufficient. Physicians do not act or behave based on the actual artifact but their mental representations of the alert artifact (construal). In other words, the same alert when presented in a different context for the same physician may lead to a different decision outcome. This is especially applicable in recent times, as AI-

enabled CDSS are more ubiquitous, and they are embedded in different medical, technological platforms such as laptops, portable medical devices, and other mobile applications. The increased mobility of medical devices suggests that they can be triggered in many different contexts and under different modes. Hence, the alert design needs to take the interaction of the alert's frequency into account, with consideration of the where, and when of the user's egocentric self.

Given that the interaction occurs as a psychological process, the design should take physician's biases into account. Specifically, in my study, I show how decisions to override an alert could be mitigated by reducing the alert's level of construal, which in turn causes the physician to use less biased and more involved information processing in deciding whether to override the alert or not. One alert design principle would be to use spatial distance to counter the hypotheticality distance as discussed above – in this case, a higher number of alerts could be activated if physicians are located near to their patients. Another design could be to ensure the alert's specialty be related to the physician's specialty, as this reduces the social distance. At the same time, I could reduce hypotheticality and alert fatigue by increasing the duration between alerts. This design principle is different from the existing literature's suggestion of suppressing alerts in general, since the physician's decision is not only due to cognitive overload but is also affected by the physician's construal of alert. Moreover, if I were to broadly suppress all alerts, I may inadvertently

compromise patient safety if it was done without careful error management (Hsieh et al. 2004; Van Der Sijs et al. 2006).

In sum, my study suggests that I can approach the design of alerts delivery in a more nuanced manner. With a greater understanding of physicians' perception of and decision making towards alerts, I suggest I need better targeting of alerts. I should design alerts to be pushed to the right physician (instead of every physician), at the right time (instead of all instances where the patient records are viewed), and in the right place (in the ward) where the physician has best access to patient status to make the decision. This approach to alert targeting is similar to how online ads targets potential customers based on the contextual factors surrounding the customers receiving the online ad.

Separately, a better understanding of physicians' biases when attending to automated clinical alerts will assist hospitals and other healthcare institutions provide better EMR training programs that highlights how physicians' biases affect their work with EMR. This will require some changes in how hospitals conduct their EMR training and induction programs for new physicians. My findings suggest that EMR training and operating guidelines should provide strategies to help physicians minimize decision biases while attending to automated alerts in an EMR system.

Limitations and Future Work

As with any research, my study is not without limitations. First, my study was conducted in a single hospital. As institutional contexts vary, this

may limit the generalizability of my findings. However, my data were extracted from a widely used EMR platform, EPIC™ EMR, which holds medical records of more than 64% of patients in the United States and 2.5% of patients worldwide (Glaze 2015), making this limitation less critical. Nonetheless, future research may investigate automatic clinical alerts in different hospitals to provide external validity to my findings.

Second, my study focusses on only one type of alert. However, there may be different types of clinical alerts working simultaneously in most medical settings (Phansalkar et al. 2010). Thus, future research should investigate whether my findings hold for other types of automatic clinical alert. Finally, while I attempted to control as many factors as possible, such as physician fixed effects and other exogenous factors, future research may take a further step in examining physicians' reactions to different alerts to rule out potential interaction among different alerts.

REFERENCE

- Ahmad, M. A., Eckert, C., and Teredesai, A. 2018. "Interpretable Machine Learning in Healthcare," *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 559-560.
- Amit, E., Algom, D., and Trope, Y. 2009a. "Distance-Dependent Processing of Pictures and Words," *Journal of Experimental Psychology: General* (138:3), pp. 400-415.
- Amit, E., Algom, D., Trope, Y., and Liberman, N. 2009b. "Thou Shalt Not Make Unto Thee Any Graven Image": The Distance-Dependence of Representation," in *The Handbook of Imagination and Mental Simulation*. pp. 53-68.
- Ancker, J. S., Edwards, A., Nosal, S., Hauser, D., Mauer, E., and Kaushal, R. 2017. "Effects of Workload, Work Complexity, and Repeated Alerts on Alert Fatigue in a Clinical Decision Support System," *BMC Medical Informatics and Decision Making* (17:1), pp. 1-9.
- Ash, J. S., Berg, M., and Coiera, E. 2004. "Some Unintended Consequences of Information Technology in Health Care: The Nature of Patient Care Information System-Related Errors," *Journal of the American Medical Informatics Association* (11:2), pp. 104-112.
- Bardhan, I., Oh, J.-h., Zheng, Z., and Kirksey, K. 2015. "Predictive Analytics for Readmission of Patients with Congestive Heart Failure," *Information Systems Research* (26:1), pp. 19-39.
- Baskerville, R. L., and Myers, M. D. 2002. "Information Systems as a Reference Discipline," *Mis Quarterly* (26:1), pp. 1-14.
- Bates, D. W., Kuperman, G. J., Wang, S., Gandhi, T., Kittler, A., Volk, L., Spurr, C., Khorasani, R., Tanasijevic, M., and Middleton, B. 2003. "Ten Commandments for Effective Clinical Decision Support: Making the Practice of Evidence-Based Medicine a Reality," *Journal of the American Medical Informatics Association* (10:6), pp. 523-530.
- Baysari, M. T., Tariq, A., Day, R. O., and Westbrook, J. I. 2017. "Alert Override as a Habitual Behavior—a New Perspective on a Persistent Problem," *Journal of the American Medical Informatics Association* (24:2), pp. 409-412.
- Bobb, A., Gleason, K., Husch, M., Feinglass, J., Yarnold, P. R., and Noskin, G. A. 2004. "The Epidemiology of Prescribing Errors: The Potential Impact of Computerized Prescriber Order Entry," *Archives of Internal Medicine* (164:7), pp. 785-792.
- Boroditsky, L. 2000. "Metaphoric Structuring: Understanding Time through Spatial Metaphors," *Cognition* (75:1), pp. 1-28.
- Boroditsky, L., and Ramscar, M. 2002. "The Roles of Body and Mind in Abstract Thought," *Psychological Science* (13:2), pp. 185-189.

- Breznitz, S. 1984. *Cry Wolf: The Psychology of False Alarms*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Breznitz, S. 1985. "False Alarms: Their Effects on Fear and Adjustment," *Issues in Mental Health Nursing* (7:1-4), pp. 335-348.
- Carspecken, C. W., Sharek, P. J., Longhurst, C., and Pageler, N. M. 2013. "A Clinical Case of Electronic Health Record Drug Alert Fatigue: Consequences for Patient Outcome," *Pediatrics* (131:6), pp. 1970-1973.
- Cash, J. J. 2009. "Alert Fatigue," *American Journal of Health-System Pharmacy* (66:23), pp. 2098-2101.
- Chandran, S., and Menon, G. 2004. "When a Day Means More Than a Year: Effects of Temporal Framing on Judgments of Health Risk," *Journal of Consumer Research* (31:2), pp. 375-389.
- Cubitt, R., McDonald, R., and Read, D. 2018. "Time Matters Less When Outcomes Differ: Unimodal Vs. Cross-Modal Comparisons in Intertemporal Choice," *Management Science* (64:2), pp. 873-887.
- Daniel, G., Silcox, C., Sharma, I., and Wright, M. 2019. "Current State and near-Term Priorities for Ai-Enabled Diagnostic Support Software in Health Care," Duke-Margolis Center of Health Policy.
- Dexter, P. R., Perkins, S., Overhage, J. M., Maharry, K., Kohler, R. B., and McDonald, C. J. 2001. "A Computerized Reminder System to Increase the Use of Preventive Care for Hospitalized Patients," *New England Journal of Medicine* (345:13), pp. 965-970.
- Diestre, L., Barber IV, B., and Santaló, J. 2019. "The Friday Effect: Firm Lobbying, the Timing of Drug Safety Alerts, and Drug Side Effects," *Management Science* (Forthcoming).
- Durieux, P., Trinquart, L., Colombet, I., Nies, J., Walton, R., Rajeswaran, A., Rège-Walther, M., Harvey, E., and Burnand, B. 2008. "Computerized Advice on Drug Dosage to Improve Prescribing Practice," *Cochrane Database of Systematic Reviews* (2008:3), pp. 1-56.
- Eisenhardt, K. M., and Graebner, M. E. 2007. "Theory Building from Cases: Opportunities and Challenges," *Academy of Management Journal* (50:1), pp. 25-32.
- Embi, P. J., and Leonard, A. C. 2012. "Evaluating Alert Fatigue over Time to Ehr-Based Clinical Trial Alerts: Findings from a Randomized Controlled Study," *Journal of the American Medical Informatics Association* (19:1), pp. 145-148.
- Fang, X., Gao, Y., and Hu, P. J. 2019. "A Prescriptive Analytics Method for Cost Reduction in Clinical Decision Making," *MIS Quarterly* (Forthcoming).
- Feldstein, A., Simon, S. R., Schneider, J., Krall, M., Laferriere, D., Smith, D. H., Sittig, D. F., and Soumerai, S. B. 2004. "How to Design Computerized Alerts to Ensure Safe Prescribing Practices," *The Joint Commission Journal on Quality and Safety* (30:11), pp. 602-613.

- Fujita, K., Henderson, M. D., Eng, J., Trope, Y., and Liberman, N. 2006. "Spatial Distance and Mental Construal of Social Events," *Psychological Science* (17:4), pp. 278-282.
- Garg, A. X., Adhikari, N. K., McDonald, H., Rosas-Arellano, M. P., Devereaux, P., Beyene, J., Sam, J., and Haynes, R. B. 2005a. "Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes: A Systematic Review," *Jama* (293:10), pp. 1223-1238.
- Garg, A. X., Adhikari, N. K., McDonald, H., Rosas-Arellano, M. P., Devereaux, P. J., Beyene, J., Sam, J., and Haynes, R. B. 2005b. "Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes: A Systematic Review," *Jama* (293:10), pp. 1223-1238.
- Gilovich, T., Kerr, M., and Medvec, V. H. 1993. "Effect of Temporal Perspective on Subjective Confidence," *Journal of Personality and Social Psychology* (64:4), pp. 552-560.
- Glassman, P. A., Simon, B., Belperio, P., and Lanto, A. 2002. "Improving Recognition of Drug Interactions: Benefits and Barriers to Using Automated Drug Alerts," *Medical Care* (40:12), pp. 1161-1171.
- Glaze, J. 2015. "Epic Systems Draws on Literature Greats for Its Next Expansion." *Wisconsin State Journal*, from https://madison.com/news/local/govt-and-politics/epic-systems-draws-on-literature-greats-for-its-next-expansion/article_4d1cf67c-2abf-5cfd-8ce1-2da60ed84194.html
- Helm, J. E., Alaeddini, A., Stauffer, J. M., Bretthauer, K. M., and Skolarus, T. A. 2016. "Reducing Hospital Readmissions by Integrating Empirical Prediction with Resource Optimization," *Production and Operations Management* (25:2), pp. 233-257.
- Henderson, M. D., Fujita, K., Trope, Y., and Liberman, N. 2006. "Transcending the "Here": The Effect of Spatial Distance on Social Judgment," *Journal of Personality and Social Psychology* (91:5), pp. 845-856.
- Henry, K. E., Hager, D. N., Pronovost, P. J., and Saria, S. 2015. "A Targeted Real-Time Early Warning Score (Trewscore) for Septic Shock," *Science Translational Medicine* (7:299), pp. 1-9.
- Holyoak, K. J., and Mah, W. A. 1982. "Cognitive Reference Points in Judgments of Symbolic Magnitude," *Cognitive Psychology* (14:3), pp. 328-352.
- Horsky, J., Zhang, J., and Patel, V. L. 2005. "To Err Is Not Entirely Human: Complex Technology and User Cognition," *Journal of Biomedical Informatics* (38:4), pp. 264-266.
- Hsieh, T. C., Kuperman, G. J., Jaggi, T., Hojnowski-Diaz, P., Fiskio, J., Williams, D. H., Bates, D. W., and Gandhi, T. K. 2004. "Characteristics and Consequences of Drug Allergy Alert Overrides in a Computerized Physician Order Entry System," *Journal of the American Medical Informatics Association* (11:6), pp. 482-491.

- Jaspers, M. W., Smeulders, M., Vermeulen, H., and Peute, L. W. 2011. "Effects of Clinical Decision-Support Systems on Practitioner Performance and Patient Outcomes: A Synthesis of High-Quality Systematic Review Findings," *Journal of the American Medical Informatics Association* (18:3), pp. 327-334.
- Jones, K. 2014. "Alarm Fatigue a Top Patient Safety Hazard," *Canadian Medical Association Journal* (186:3), pp. 178-178.
- Kaufman, G., and Flanagan, M. 2016. "High-Low Split: Divergent Cognitive Construal Levels Triggered by Digital and Non-Digital Platforms," *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 2773-2777.
- Kawamoto, K., Houlihan, C. A., Balas, E. A., and Lobach, D. F. 2005. "Improving Clinical Practice Using Clinical Decision Support Systems: A Systematic Review of Trials to Identify Features Critical to Success," *BMJ* (330:7494), pp. 1-8.
- Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., and King, D. 2019. "Key Challenges for Delivering Clinical Impact with Artificial Intelligence," *BMC Medicine* (17:1), pp. 1-9.
- Kesselheim, A. S., Cresswell, K., Phansalkar, S., Bates, D. W., and Sheikh, A. 2011. "Clinical Decision Support Systems Could Be Modified to Reduce 'Alert Fatigue' while Still Minimizing the Risk of Litigation," *Health Affairs* (30:12), pp. 2310-2317.
- Klevens, R. M., Edwards, J. R., Richards Jr, C. L., Horan, T. C., Gaynes, R. P., Pollock, D. A., and Cardo, D. M. 2007. "Estimating Health Care-Associated Infections and Deaths in Us Hospitals, 2002," *Public Health Reports* (122:2), pp. 160-166.
- Köhler, C. F., Breugelmans, E., and Dellaert, B. G. 2011. "Consumer Acceptance of Recommendations by Interactive Decision Aids: The Joint Role of Temporal Distance and Concrete Versus Abstract Communications," *Journal of Management Information Systems* (27:4), pp. 231-260.
- Krall, M. A., and Sittig, D. F. 2002. "Clinician's Assessments of Outpatient Electronic Medical Record Alert and Reminder Usability and Usefulness Requirements," *Proceedings of the AMIA Symposium: American Medical Informatics Association*, p. 400.
- Kucher, N., Koo, S., Quiroz, R., Cooper, J. M., Paterno, M. D., Soukonnikov, B., and Goldhaber, S. Z. 2005. "Electronic Alerts to Prevent Venous Thromboembolism among Hospitalized Patients," *New England Journal of Medicine* (352:10), pp. 969-977.
- Lee, J. S., Keil, M., and Shalev, E. 2019. "Seeing the Trees or the Forest? The Effect of IT Project Managers' Mental Construal on IT Project Risk Management Activities," *Information Systems Research* (30:3), pp. 1051-1072.

- Liberman, N., Sagristano, M. D., and Trope, Y. 2002. "The Effect of Temporal Distance on Level of Mental Construal," *Journal of Experimental Social Psychology* (38:6), pp. 523-534.
- Liberman, N., and Trope, Y. 1998. "The Role of Feasibility and Desirability Considerations in near and Distant Future Decisions: A Test of Temporal Construal Theory," *Journal of Personality and Social Psychology* (75:1), pp. 5-18.
- Liberman, N., Trope, Y., and Stephan, E. 2007. "Psychological Distance,").
- Lo, E., Nicolle, L., Classen, D., Arias, K. M., Podgorny, K., Anderson, D. J., Burstin, H., Calfee, D. P., Coffin, S. E., and Dubberke, E. R. 2008. "Strategies to Prevent Catheter-Associated Urinary Tract Infections in Acute Care Hospitals," *Infection Control & Hospital Epidemiology* (29:1), pp. 41-50.
- Luo, X., Andrews, M., Fang, Z., and Phang, C. W. 2014. "Mobile Targeting," *Management Science* (60:7), pp. 1738-1756.
- Lysaght, T., Lim, H. Y., Xafis, V., and Ngiam, K. Y. 2019. "Ai-Assisted Decision-Making in Healthcare," *Asian Bioethics Review* (11:3), pp. 299-314.
- Magill, S. S., O'Leary, E., Janelle, S. J., Thompson, D. L., Dumyati, G., Nadle, J., Wilson, L. E., Kainer, M. A., Lynfield, R., and Greissman, S. 2018. "Changes in Prevalence of Health Care–Associated Infections in Us Hospitals," *New England Journal of Medicine* (379:18), pp. 1732-1744.
- Magnus, D., Rodgers, S., and Avery, A. 2002. "Gps' Views on Computerized Drug Interaction Alerts: Questionnaire Survey," *Journal of Clinical Pharmacy and Therapeutics* (27:5), pp. 377-382.
- McCoy, A. B., Waitman, L. R., Lewis, J. B., Wright, J. A., Choma, D. P., Miller, R. A., and Peterson, J. F. 2011. "A Framework for Evaluating the Appropriateness of Clinical Decision Support Alerts and Responses," *Journal of the American Medical Informatics Association* (19:3), pp. 346-352.
- Mitchell, T. R., Thompson, L., Peterson, E., and Cronk, R. 1997. "Temporal Adjustments in the Evaluation of Events: The "Rosy View"," *Journal of Experimental Social Psychology* (33:4), pp. 421-448.
- Montani, S., and Striani, M. 2019. "Artificial Intelligence in Clinical Decision Support: A Focused Literature Survey," *Yearbook of Medical Informatics* (28:1), pp. 120-127.
- Nanji, K. C., Slight, S. P., Seger, D. L., Cho, I., Fiskio, J. M., Redden, L. M., Volk, L. A., and Bates, D. W. 2014. "Overrides of Medication-Related Clinical Decision Support Alerts in Outpatients," *Journal of the American Medical Informatics Association* (21:3), pp. 487-491.
- Nisan, M. 1972. "Dimension of Time in Relation to Choice Behavior and Achievement Orientation," *Journal of Personality and Social Psychology* (21:2), pp. 175-182.

- Nussbaum, S., Liberman, N., and Trope, Y. 2006. "Predicting the near and Distant Future," *Journal of Experimental Psychology: General* (135:2), pp. 152-161.
- Nussbaum, S., Trope, Y., and Liberman, N. 2003. "Creeping Dispositionism: The Temporal Dynamics of Behavior Prediction," *Journal of Personality and Social Psychology* (84:3), pp. 485-497.
- Oppenheim, M. I., Vidal, C., Velasco, F. T., Boyer, A. G., Cooper, M. R., Hayes, J. G., and Frayer, W. W. 2002. "Impact of a Computerized Alert During Physician Order Entry on Medication Dosing in Patients with Renal Impairment," *Proceedings of the AMIA Symposium: American Medical Informatics Association*, pp. 577-581.
- Overhage, J. M., Tierney, W. M., Zhou, X.-H., and McDonald, C. J. 1997. "A Randomized Trial of "Corollary Orders" to Prevent Errors of Omission," *Journal of the American Medical Informatics Association* (4:5), pp. 364-375.
- Park, B., and Judd, C. M. 1990. "Measures and Models of Perceived Group Variability," *Journal of Personality and Social Psychology* (59:2), pp. 173-191.
- Park, B., Ryan, C. S., and Judd, C. M. 1992. "Role of Meaningful Subgroups in Explaining Differences in Perceived Variability for in-Groups and out-Groups," *Journal of Personality and Social Psychology* (63:4), pp. 553-567.
- Peterson, J. F., and Bates, D. W. 2001. "Preventable Medication Errors: Identifying and Eliminating Serious Drug Interactions," *Journal of the American Pharmaceutical Association* (41:2), pp. 159-160.
- Pettit, L. 2013. "Understanding Emram and How It Can Be Used by Policy-Makers, Hospital Cios and Their It Teams," *World hospitals and health services: the official journal of the International Hospital Federation* (49:3), pp. 7-9.
- Phang, C. W., Kankanhalli, A., and Tan, B. C. 2015. "What Motivates Contributors Vs. Lurkers? An Investigation of Online Feedback Forums," *Information Systems Research* (26:4), pp. 773-792.
- Phansalkar, S., Edworthy, J., Hellier, E., Seger, D. L., Schedlbauer, A., Avery, A. J., and Bates, D. W. 2010. "A Review of Human Factors Principles for the Design and Implementation of Medication Safety Alerts in Clinical Information Systems," *Journal of the American Medical Informatics Association* (17:5), pp. 493-501.
- Phansalkar, S., Van der Sijs, H., Tucker, A. D., Desai, A. A., Bell, D. S., Teich, J. M., Middleton, B., and Bates, D. W. 2013. "Drug—Drug Interactions That Should Be Non-Interruptive in Order to Reduce Alert Fatigue in Electronic Health Records," *Journal of the American Medical Informatics Association* (20:3), pp. 489-493.
- Pronin, E., Olivola, C. Y., and Kennedy, K. A. 2008. "Doing Unto Future Selves as You Would Do Unto Others: Psychological Distance and

- Decision Making," *Personality and Social Psychology Bulletin* (34:2), pp. 224-236.
- Romanow, D., Rai, A., and Keil, M. 2018. "Cpoe-Enabled Coordination: Appropriation for Deep Structure Use and Impacts on Patient Outcomes," *MIS Quarterly* (42:1), pp. 189-212.
- Russ, A. L., Zillich, A. J., Melton, B. L., Russell, S. A., Chen, S., Spina, J. R., Weiner, M., Johnson, E. G., Daggy, J. K., and McManus, M. S. 2014. "Applying Human Factors Principles to Alert Design Increases Efficiency and Reduces Prescribing Errors in a Scenario-Based Simulation," *Journal of the American Medical Informatics Association* (21:e2), pp. e287-e296.
- Schedlbauer, A., Prasad, V., Mulvaney, C., Phansalkar, S., Stanton, W., Bates, D. W., and Avery, A. J. 2009. "What Evidence Supports the Use of Computerized Alerts and Prompts to Improve Clinicians' Prescribing Behavior?," *Journal of the American Medical Informatics Association* (16:4), pp. 531-538.
- Shah, N. R., Seger, A. C., Seger, D. L., Fiskio, J. M., Kuperman, G. J., Blumenfeld, B., Recklet, E. G., Bates, D. W., and Gandhi, T. K. 2006. "Improving Acceptance of Computerized Prescribing Alerts in Ambulatory Care," *Journal of the American Medical Informatics Association* (13:1), pp. 5-11.
- Shani, Y., Igou, E. R., and Zeelenberg, M. 2009. "Different Ways of Looking at Unpleasant Truths: How Construal Levels Influence Information Search," *Organizational Behavior and Human Decision Processes* (110:1), pp. 36-44.
- Shen, C.-C., and Chiou, J.-S. 2010. "The Impact of Perceived Ease of Use on Internet Service Adoption: The Moderating Effects of Temporal Distance and Perceived Risk," *Computers in Human Behavior* (26:1), pp. 42-50.
- Smith, P. K., and Trope, Y. 2006. "You Focus on the Forest When You're in Charge of the Trees: Power Priming and Abstract Information Processing," *Journal of Personality and Social Psychology* (90:4), pp. 578-596.
- Stephan, E., Liberman, N., and Trope, Y. 2010. "Politeness and Psychological Distance: A Construal Level Perspective," *Journal of Personality and Social Psychology* (98:2), pp. 268-280.
- Stern, E. R., and Rotello, C. M. 2000. "Memory Characteristics of Recently Imagined Events and Real Events Experienced Previously," *The American Journal of Psychology* (113:4), pp. 569-590.
- Strom, B. L., Schinnar, R., Aberra, F., Bilker, W., Hennessy, S., Leonard, C. E., and Pifer, E. 2010a. "Unintended Effects of a Computerized Physician Order Entry Nearly Hard-Stop Alert to Prevent a Drug Interaction: A Randomized Controlled Trial," *Archives of Internal Medicine* (170:17), pp. 1578-1583.

- Strom, B. L., Schinnar, R., Bilker, W., Hennessy, S., Leonard, C. E., and Pifer, E. 2010b. "Randomized Clinical Trial of a Customized Electronic Alert Requiring an Affirmative Response Compared to a Control Group Receiving a Commercial Passive Cpoe Alert: Nsaid—Warfarin Co-Prescribing as a Test Case," *Journal of the American Medical Informatics Association* (17:4), pp. 411-415.
- Taylor, L., and Tamblyn, R. 2004. "Reasons for Physician Non-Adherence to Electronic Drug Alerts," *Medinfo* (11), pp. 1101-1105.
- Timmermans, S., and Mauck, A. 2005. "The Promises and Pitfalls of Evidence-Based Medicine," *Health Affairs* (24:1), pp. 18-28.
- Trope, Y., and Liberman, N. 2003. "Temporal Construal," *Psychological Review* (110:3), pp. 403-421.
- Trope, Y., and Liberman, N. 2010. "Construal-Level Theory of Psychological Distance," *Psychological Review* (117:2), pp. 440-463.
- Trope, Y., Liberman, N., and Wakslak, C. 2007. "Construal Levels and Psychological Distance: Effects on Representation, Prediction, Evaluation, and Behavior," *Journal of Consumer Psychology* (17:2), pp. 83-95.
- Van Der Sijs, H., Aarts, J., Van Gelder, T., Berg, M., and Vulto, A. 2008. "Turning Off Frequently Overridden Drug Alerts: Limited Opportunities for Doing It Safely," *Journal of the American Medical Informatics Association* (15:4), pp. 439-448.
- Van Der Sijs, H., Aarts, J., Vulto, A., and Berg, M. 2006. "Overriding of Drug Safety Alerts in Computerized Physician Order Entry," *Journal of the American Medical Informatics Association* (13:2), pp. 138-147.
- Wakslak, C., and Trope, Y. 2008. "The Who, Where, and When of Low and High Probability Events: Probability as Distance and Everyday Decision-Making." New York University.
- Wakslak, C. J., Nussbaum, S., Liberman, N., and Trope, Y. 2008. "Representations of the Self in the near and Distant Future," *Journal of Personality and Social Psychology* (95:4), pp. 757-773.
- Wakslak, C. J., Trope, Y., Liberman, N., and Alony, R. 2006. "Seeing the Forest When Entry Is Unlikely: Probability and the Mental Representation of Events," *Journal of Experimental Psychology: General* (135:4), pp. 641-653.
- Wang, J., and Cole, C. A. 2016. "The Effects of Age and Expertise on Product Evaluations: Does the Type of Information Matter?," *Management Science* (62:7), pp. 2039-2053.
- Weingart, S. N., Toth, M., Sands, D. Z., Aronson, M. D., Davis, R. B., and Phillips, R. S. 2003. "Physicians' Decisions to Override Computerized Drug Alerts in Primary Care," *Archives of Internal Medicine* (163:21), pp. 2625-2631.
- Wilson, F. P., Shashaty, M., Testani, J., Aqeel, I., Borovskiy, Y., Ellenberg, S. S., Feldman, H. I., Fernandez, H., Gitelman, Y., and Lin, J. 2015. "Automated, Electronic Alerts for Acute Kidney Injury: A Single-Blind,

- Parallel-Group, Randomised Controlled Trial," *The Lancet* (385:9981), pp. 1966-1974.
- Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT press.
- Zauberman, G., Kim, B. K., Malkoc, S. A., and Bettman, J. R. 2009. "Discounting Time and Time Discounting: Subjective Time Perception and Intertemporal Preferences," *Journal of Marketing Research* (46:4), pp. 543-556.
- Abawajy, J., Chowdhury, M., and Kelarev, A. 2015. "Hybrid Consensus Pruning of Ensemble Classifiers for Big Data Malware Detection," *IEEE Transactions on Cloud Computing* (8:2), pp. 398-407.
- Agarwal, R., and Dhar, V. 2014. "Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research," *Information Systems Research* (25:3), pp. 443-448.
- Agarwal, R., Gao, G., DesRoches, C., and Jha, A. K. 2010. "Research Commentary—the Digital Transformation of Healthcare: Current Status and the Road Ahead," *Information Systems Research* (21:4), pp. 796-809.
- Ahmad, M. A., Eckert, C., and Teredesai, A. 2018. "Interpretable Machine Learning in Healthcare," *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 559-560.
- Amarasingham, R., Patzer, R. E., Huesch, M., Nguyen, N. Q., and Xie, B. 2014. "Implementing Electronic Health Care Predictive Analytics: Considerations and Challenges," *Health Affairs* (33:7), pp. 1148-1154.
- Amit, E., Algom, D., and Trope, Y. 2009a. "Distance-Dependent Processing of Pictures and Words," *Journal of Experimental Psychology: General* (138:3), pp. 400-415.
- Amit, E., Algom, D., Trope, Y., and Liberman, N. 2009b. "Thou Shalt Not Make Unto Thee Any Graven Image": The Distance-Dependence of Representation," in *The Handbook of Imagination and Mental Simulation*. pp. 53-68.
- Ancker, J. S., Edwards, A., Nosal, S., Hauser, D., Mauer, E., and Kaushal, R. 2017. "Effects of Workload, Work Complexity, and Repeated Alerts on Alert Fatigue in a Clinical Decision Support System," *BMC Medical Informatics and Decision Making* (17:1), pp. 1-9.
- Ash, J. S., Berg, M., and Coiera, E. 2004. "Some Unintended Consequences of Information Technology in Health Care: The Nature of Patient Care Information System-Related Errors," *Journal of the American Medical Informatics Association* (11:2), pp. 104-112.
- Bardhan, I., Oh, J.-h., Zheng, Z., and Kirksey, K. 2015. "Predictive Analytics for Readmission of Patients with Congestive Heart Failure," *Information Systems Research* (26:1), pp. 19-39.

- Batista, G. E., Prati, R. C., and Monard, M. C. 2004. "A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data," *ACM SIGKDD Explorations Newsletter* (6:1), pp. 20-29.
- Baysari, M. T., Tariq, A., Day, R. O., and Westbrook, J. I. 2017. "Alert Override as a Habitual Behavior—a New Perspective on a Persistent Problem," *Journal of the American Medical Informatics Association* (24:2), pp. 409-412.
- Blei, D. M. 2012. "Probabilistic Topic Models," *Communications of the ACM* (55:4), pp. 77-84.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. "Latent Dirichlet Allocation," *Journal of Machine Learning Research* (3:Jan), pp. 993-1022.
- Bobb, A., Gleason, K., Husch, M., Feinglass, J., Yarnold, P. R., and Noskin, G. A. 2004. "The Epidemiology of Prescribing Errors: The Potential Impact of Computerized Prescriber Order Entry," *Archives of Internal Medicine* (164:7), pp. 785-792.
- Boroditsky, L. 2000. "Metaphoric Structuring: Understanding Time through Spatial Metaphors," *Cognition* (75:1), pp. 1-28.
- Boroditsky, L., and Ramscar, M. 2002. "The Roles of Body and Mind in Abstract Thought," *Psychological Science* (13:2), pp. 185-189.
- Breznitz, S. 1984. *Cry Wolf: The Psychology of False Alarms*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Breznitz, S. 1985. "False Alarms: Their Effects on Fear and Adjustment," *Issues in Mental Health Nursing* (7:1-4), pp. 335-348.
- Brynjolfsson, E., and McAfee, A. 2012. "Winning the Race with Ever-Smarter Machines," *MIT Sloan Management Review* (53:2), pp. 53-60.
- Carnielli, C. M., Macedo, C. C. S., De Rossi, T., Granato, D. C., Rivera, C., Domingues, R. R., Pauletti, B. A., Yokoo, S., Heberle, H., Busso-Lopes, A. F., Cervigne, N. K., Sawazaki-Calone, I., Meirelles, G. V., Marchi, F. A., Telles, G. P., Minghim, R., Ribeiro, A. C. P., Brandão, T. B., de Castro, G., González-Arriagada, W. A., Gomes, A., Pentead, F., Santos-Silva, A. R., Lopes, M. A., Rodrigues, P. C., Sundquist, E., Salo, T., da Silva, S. D., Alaoui-Jamali, M. A., Graner, E., Fox, J. W., Coletta, R. D., and Paes Leme, A. F. 2018. "Combining Discovery and Targeted Proteomics Reveals a Prognostic Signature in Oral Cancer," *Nature Communications* (9:1), p. 3598.
- Carspecken, C. W., Sharek, P. J., Longhurst, C., and Pageler, N. M. 2013. "A Clinical Case of Electronic Health Record Drug Alert Fatigue: Consequences for Patient Outcome," *Pediatrics* (131:6), pp. 1970-1973.
- Cash, J. J. 2009. "Alert Fatigue," *American Journal of Health-System Pharmacy* (66:23), pp. 2098-2101.
- Chandran, S., and Menon, G. 2004. "When a Day Means More Than a Year: Effects of Temporal Framing on Judgments of Health Risk," *Journal of Consumer Research* (31:2), pp. 375-389.

- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. 2002. "Smote: Synthetic Minority over-Sampling Technique," *Journal of Artificial Intelligence Research* (16), pp. 321-357.
- Cheatham, B., Javanmardian, K., and Samandari, H. 2019. "Confronting the Risks of Artificial Intelligence," *McKinsey Quarterly*, pp. 1-9.
- Chen, H., Chiang, R. H., and Storey, V. C. 2012. "Business Intelligence and Analytics: From Big Data to Big Impact," *MIS Quarterly* (36:4), pp. 1165-1188.
- Chen, J. H., Alagappan, M., Goldstein, M. K., Asch, S. M., and Altman, R. B. 2017. "Decaying Relevance of Clinical Data Towards Future Decisions in Data-Driven Inpatient Clinical Order Sets," *International journal of medical informatics* (102), pp. 71-79.
- Cohen, I. G., Amarasingham, R., Shah, A., Xie, B., and Lo, B. 2014. "The Legal and Ethical Concerns That Arise from Using Complex Predictive Analytics in Health Care," *Health affairs* (33:7), pp. 1139-1147.
- Croskerry, P. 2005. "The Theory and Practice of Clinical Decision-Making," *Canadian Journal of Anesthesia* (52), pp. R1-R8.
- Cubitt, R., McDonald, R., and Read, D. 2018. "Time Matters Less When Outcomes Differ: Unimodal Vs. Cross-Modal Comparisons in Intertemporal Choice," *Management Science* (64:2), pp. 873-887.
- Dane, E., Rockmann, K. W., and Pratt, M. G. 2012. "When Should I Trust My Gut? Linking Domain Expertise to Intuitive Decision-Making Effectiveness," *Organizational Behavior and Human Decision Processes* (119:2), pp. 187-194.
- Daniel, G., Silcox, C., Sharma, I., and Wright, M. 2019. "Current State and near-Term Priorities for Ai-Enabled Diagnostic Support Software in Health Care," Duke-Margolis Center of Health Policy.
- Desautels, T., Calvert, J., Hoffman, J., Jay, M., Kerem, Y., Shieh, L., Shimabukuro, D., Chettipally, U., Feldman, M. D., and Barton, C. 2016. "Prediction of Sepsis in the Intensive Care Unit with Minimal Electronic Health Record Data: A Machine Learning Approach," *JMIR Medical Informatics* (4:3), p. e28.
- Dexter, P. R., Perkins, S., Overhage, J. M., Maharry, K., Kohler, R. B., and McDonald, C. J. 2001. "A Computerized Reminder System to Increase the Use of Preventive Care for Hospitalized Patients," *New England Journal of Medicine* (345:13), pp. 965-970.
- Diestre, L., Barber IV, B., and Santaló, J. 2019. "The Friday Effect: Firm Lobbying, the Timing of Drug Safety Alerts, and Drug Side Effects," *Management Science* (Forthcoming).
- Dummitt, B., Zeringue, A., Palagiri, A., Veremakis, C., Burch, B., and Yount, B. 2018. "Using Survival Analysis to Predict Septic Shock Onset in Icu Patients," *Journal of Critical Care* (48), pp. 339-344.
- Durieux, P., Trinquart, L., Colombet, I., Nies, J., Walton, R., Rajeswaran, A., Rège-Walther, M., Harvey, E., and Burnand, B. 2008. "Computerized

- Advice on Drug Dosage to Improve Prescribing Practice," *Cochrane Database of Systematic Reviews* (2008:3), pp. 1-56.
- Eisenhardt, K. M., and Graebner, M. E. 2007. "Theory Building from Cases: Opportunities and Challenges," *Academy of Management Journal* (50:1), pp. 25-32.
- Elish, M. C., and Watkins, E. A. 2020. "Repairing Innovation: A Study of Integrating Ai in Clinical Care," Data & Society Research Institute.
- Ferrer, R., Martin-Loeches, I., Phillips, G., Osborn, T. M., Townsend, S., Dellinger, R. P., Artigas, A., Schorr, C., and Levy, M. M. 2014. "Empiric Antibiotic Treatment Reduces Mortality in Severe Sepsis and Septic Shock from the First Hour: Results from a Guideline-Based Performance Improvement Program," *Critical Care Medicine* (42:8), pp. 1749-1755.
- Fichman, R. G., Kohli, R., and Krishnan, R. 2011. "Editorial Overview—the Role of Information Systems in Healthcare: Current Research and Future Trends," *Information Systems Research* (22:3), pp. 419-428.
- Fujita, K., Henderson, M. D., Eng, J., Trope, Y., and Liberman, N. 2006. "Spatial Distance and Mental Construal of Social Events," *Psychological Science* (17:4), pp. 278-282.
- Garg, A. X., Adhikari, N. K., McDonald, H., Rosas-Arellano, M. P., Devereaux, P., Beyene, J., Sam, J., and Haynes, R. B. 2005. "Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes: A Systematic Review," *Jama* (293:10), pp. 1223-1238.
- Glassman, P. A., Simon, B., Belperio, P., and Lanto, A. 2002. "Improving Recognition of Drug Interactions: Benefits and Barriers to Using Automated Drug Alerts," *Medical Care* (40:12), pp. 1161-1171.
- Glaze, J. 2015. "Epic Systems Draws on Literature Greats for Its Next Expansion." *Wisconsin State Journal*, from https://madison.com/news/local/govt-and-politics/epic-systems-draws-on-literature-greats-for-its-next-expansion/article_4d1cf67c-2abf-5cfd-8ce1-2da60ed84194.html
- Guszcza, J., Lewis, H., and Evans-Greenwood, P. 2017. "Cognitive Collaboration: Why Humans and Computers Think Better Together." *Deloitte Insights*, from <https://www2.deloitte.com/insights/us/en/deloitte-review/issue-20/augmented-intelligence-human-computer-collaboration.html>
- Han, W., Ada, S., Sharman, R., and Rao, H. R. 2015. "Campus Emergency Notification Systems: An Examination of Factors Affecting Compliance with Alerts," *Mis Quarterly* (39:4), pp. 909-930.
- Helm, J. E., Alaeddini, A., Stauffer, J. M., Bretthauer, K. M., and Skolarus, T. A. 2016. "Reducing Hospital Readmissions by Integrating Empirical Prediction with Resource Optimization," *Production and Operations Management* (25:2), pp. 233-257.

- Henderson, M. D., Fujita, K., Trope, Y., and Liberman, N. 2006. "Transcending the "Here": The Effect of Spatial Distance on Social Judgment," *Journal of Personality and Social Psychology* (91:5), pp. 845-856.
- Henry, K. E., Hager, D. N., Pronovost, P. J., and Saria, S. 2015a. "A Targeted Real-Time Early Warning Score (Trewscore) for Septic Shock," *Science Translational Medicine* (7:299), pp. 1-9.
- Henry, K. E., Hager, D. N., Pronovost, P. J., and Saria, S. 2015b. "A Targeted Real-Time Early Warning Score (Trewscore) for Septic Shock," *Science Translational Medicine* (7:299), p. 299ra122.
- Hoffman, R. 2016. "Using Artificial Intelligence to Set Information Free," *MIT Sloan Management Review* (58:1), p. 20.
- Holyoak, K. J., and Mah, W. A. 1982. "Cognitive Reference Points in Judgments of Symbolic Magnitude," *Cognitive Psychology* (14:3), pp. 328-352.
- Hornig, S., Sontag, D. A., Halpern, Y., Jernite, Y., Shapiro, N. I., and Nathanson, L. A. 2017. "Creating an Automated Trigger for Sepsis Clinical Decision Support at Emergency Department Triage Using Machine Learning," *PloS One* (12:4), p. e0174708.
- Horsky, J., Zhang, J., and Patel, V. L. 2005. "To Err Is Not Entirely Human: Complex Technology and User Cognition," *Journal of Biomedical Informatics* (38:4), pp. 264-266.
- Hsieh, T. C., Kuperman, G. J., Jaggi, T., Hojnowski-Diaz, P., Fiskio, J., Williams, D. H., Bates, D. W., and Gandhi, T. K. 2004. "Characteristics and Consequences of Drug Allergy Alert Overrides in a Computerized Physician Order Entry System," *Journal of the American Medical Informatics Association* (11:6), pp. 482-491.
- Hsu, W., Han, S. X., Arnold, C. W., Bui, A. A., and Enzmann, D. R. 2015. "A Data-Driven Approach for Quality Assessment of Radiologic Interpretations," *Journal of the American Medical Informatics Association* (23:e1), pp. e152-e156.
- Hung, S.-Y. 2003. "Expert Versus Novice Use of the Executive Support Systems: An Empirical Study," *Information and Management* (40:3), pp. 177-189.
- Islam, M. M., Nasrin, T., Walther, B. A., Wu, C.-C., Yang, H.-C., and Li, Y.-C. 2019. "Prediction of Sepsis Patients Using Machine Learning Approach: A Meta-Analysis," *Computer Methods and Programs in Biomedicine* (170), pp. 1-9.
- Jacucci, G., Spagnolli, A., Freeman, J., and Gamberini, L. 2015. "Symbiotic Interaction: A Critical Definition and Comparison to Other Human-Computer Paradigms," *International Workshop on Symbiotic Interaction*: Springer, pp. 3-20.
- Jain, H., Padmanabhan, B., Pavlou, P. A., and Santanam, R. T. 2018. "Call for Papers—Special Issue of Information Systems Research—Humans, Algorithms, and Augmented Intelligence: The Future of Work,

- Organizations, and Society," *Information Systems Research* (29:1), pp. 250-251.
- Jaspers, M. W., Smeulders, M., Vermeulen, H., and Peute, L. W. 2011. "Effects of Clinical Decision-Support Systems on Practitioner Performance and Patient Outcomes: A Synthesis of High-Quality Systematic Review Findings," *Journal of the American Medical Informatics Association* (18:3), pp. 327-334.
- Jensen, K., Soguero-Ruiz, C., Mikalsen, K. O., Lindsetmo, R.-O., Kouskoumvekaki, I., Girolami, M., Skrovseth, S. O., and Augestad, K. M. 2017. "Analysis of Free Text in Electronic Health Records for Identification of Cancer Patient Trajectories," *Scientific reports* (7), p. 46226.
- Johnston, A. C., and Warkentin, M. 2010. "Fear Appeals and Information Security Behaviors: An Empirical Study," *MIS quarterly*, pp. 549-566.
- Jones, K. 2014. "Alarm Fatigue a Top Patient Safety Hazard," *Canadian Medical Association Journal* (186:3), pp. 178-178.
- Jones, S. L., Ashton, C. M., Kiehne, L. B., Nicolas, J. C., Rose, A. L., Shirkey, B. A., Masud, F., and Wray, N. P. 2016. "Outcomes and Resource Use of Sepsis-Associated Stays by Presence on Admission, Severity, and Hospital Type," *Medical care* (54:3), p. 303.
- Kahneman, D. 2011. *Thinking, Fast and Slow*. Macmillan.
- Kamar, E., Hacker, S., and Horvitz, E. 2012. "Combining Human and Machine Intelligence in Large-Scale Crowdsourcing," *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1: International Foundation for Autonomous Agents and Multiagent Systems*, pp. 467-474.
- Kaufman, G., and Flanagan, M. 2016. "High-Low Split: Divergent Cognitive Construal Levels Triggered by Digital and Non-Digital Platforms," *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 2773-2777.
- Kawamoto, K., Houlihan, C. A., Balas, E. A., and Lobach, D. F. 2005. "Improving Clinical Practice Using Clinical Decision Support Systems: A Systematic Review of Trials to Identify Features Critical to Success," *BMJ* (330:7494), pp. 1-8.
- Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., and King, D. 2019. "Key Challenges for Delivering Clinical Impact with Artificial Intelligence," *BMC Medicine* (17:1), pp. 1-9.
- Kesselheim, A. S., Cresswell, K., Phansalkar, S., Bates, D. W., and Sheikh, A. 2011. "Clinical Decision Support Systems Could Be Modified to Reduce 'Alert Fatigue' while Still Minimizing the Risk of Litigation," *Health Affairs* (30:12), pp. 2310-2317.
- Klevens, R. M., Edwards, J. R., Richards Jr, C. L., Horan, T. C., Gaynes, R. P., Pollock, D. A., and Cardo, D. M. 2007. "Estimating Health Care-Associated Infections and Deaths in Us Hospitals, 2002," *Public Health Reports* (122:2), pp. 160-166.

- Köhler, C. F., Breugelmans, E., and Dellaert, B. G. 2011. "Consumer Acceptance of Recommendations by Interactive Decision Aids: The Joint Role of Temporal Distance and Concrete Versus Abstract Communications," *Journal of Management Information Systems* (27:4), pp. 231-260.
- Kohli, R., and Tan, S. S.-L. 2016. "Electronic Health Records: How Can Is Researchers Contribute to Transforming Healthcare?," *MIS Quarterly* (40:3), pp. 553-573.
- Komiak, S. Y., and Benbasat, I. 2006. "The Effects of Personalization and Familiarity on Trust and Adoption of Recommendation Agents," *MIS quarterly*, pp. 941-960.
- Komorowski, M., Celi, L. A., Badawi, O., Gordon, A. C., and Faisal, A. A. 2018. "The Artificial Intelligence Clinician Learns Optimal Treatment Strategies for Sepsis in Intensive Care," *Nature Medicine* (24:11), pp. 1716–1720.
- Krall, M. A., and Sittig, D. F. 2002. "Clinician's Assessments of Outpatient Electronic Medical Record Alert and Reminder Usability and Usefulness Requirements," *Proceedings of the AMIA Symposium: American Medical Informatics Association*, p. 400.
- Kucher, N., Koo, S., Quiroz, R., Cooper, J. M., Paterno, M. D., Soukonnikov, B., and Goldhaber, S. Z. 2005. "Electronic Alerts to Prevent Venous Thromboembolism among Hospitalized Patients," *New England Journal of Medicine* (352:10), pp. 969-977.
- Kumar, A., Roberts, D., Wood, K. E., Light, B., Parrillo, J. E., Sharma, S., Suppes, R., Feinstein, D., Zanotti, S., and Taiberg, L. 2006. "Duration of Hypotension before Initiation of Effective Antimicrobial Therapy Is the Critical Determinant of Survival in Human Septic Shock," *Critical Care Medicine* (34:6), pp. 1589-1596.
- Lee, J. S., Keil, M., and Shalev, E. 2019. "Seeing the Trees or the Forest? The Effect of It Project Managers' Mental Construal on It Project Risk Management Activities," *Information Systems Research* (30:3), pp. 1051-1072.
- Lee, M. K., Kusbit, D., Metsky, E., and Dabbish, L. 2015. "Working with Machines: The Impact of Algorithmic and Data-Driven Management on Human Workers," *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pp. 1603-1612.
- Lee, Y. E., and Benbasat, I. 2011. "Research Note—the Influence of Trade-Off Difficulty Caused by Preference Elicitation Methods on User Acceptance of Recommendation Agents across Loss and Gain Conditions," *Information Systems Research* (22:4), pp. 867-884.
- Liberman, N., Sagristano, M. D., and Trope, Y. 2002. "The Effect of Temporal Distance on Level of Mental Construal," *Journal of Experimental Social Psychology* (38:6), pp. 523-534.
- Liberman, N., and Trope, Y. 1998. "The Role of Feasibility and Desirability Considerations in near and Distant Future Decisions: A Test of

- Temporal Construal Theory," *Journal of Personality and Social Psychology* (75:1), pp. 5-18.
- Liberman, N., Trope, Y., and Stephan, E. 2007. "Psychological Distance,").
- Lin, Y.-K., Chen, H., Brown, R. A., Li, S.-H., and Yang, H.-J. 2017. "Healthcare Predictive Analytics for Risk Profiling in Chronic Care: A Bayesian Multitask Learning Approach," *MIS Quarterly* (41:2), pp. 476-495.
- Lipshitz, R., and Strauss, O. 1997. "Coping with Uncertainty: A Naturalistic Decision-Making Analysis," *Organizational Behavior and Human Decision Processes* (69:2), pp. 149-163.
- Liu, R., Greenstein, J. L., Sarma, S. V., and Winslow, R. L. 2019. "Natural Language Processing of Clinical Notes for Improved Early Prediction of Septic Shock in the Icu," *IEEE*, pp. 6103-6108.
- Liu, Y., Gadepalli, K., Norouzi, M., Dahl, G. E., Kohlberger, T., Boyko, A., Venugopalan, S., Timofeev, A., Nelson, P. Q., and Corrado, G. S. 2017. "Detecting Cancer Metastases on Gigapixel Pathology Images," *arXiv preprint arXiv:1703.02442*).
- Lo, E., Nicolle, L., Classen, D., Arias, K. M., Podgorny, K., Anderson, D. J., Burstin, H., Calfee, D. P., Coffin, S. E., and Dubberke, E. R. 2008. "Strategies to Prevent Catheter-Associated Urinary Tract Infections in Acute Care Hospitals," *Infection Control & Hospital Epidemiology* (29:1), pp. 41-50.
- Lukaszewski, R. A., Yates, A. M., Jackson, M. C., Swingler, K., Scherer, J. M., Simpson, A., Sadler, P., McQuillan, P., Titball, R. W., and Brooks, T. J. 2008. "Presymptomatic Prediction of Sepsis in Intensive Care Unit Patients," *Clinical and Vaccine Immunology* (15:7), pp. 1089-1094.
- Luo, X., Andrews, M., Fang, Z., and Phang, C. W. 2014. "Mobile Targeting," *Management Science* (60:7), pp. 1738-1756.
- Lysaght, T., Lim, H. Y., Xafis, V., and Ngiam, K. Y. 2019. "Ai-Assisted Decision-Making in Healthcare," *Asian Bioethics Review* (11:3), pp. 299-314.
- Magill, S. S., O'Leary, E., Janelle, S. J., Thompson, D. L., Dumyati, G., Nadle, J., Wilson, L. E., Kainer, M. A., Lynfield, R., and Greissman, S. 2018. "Changes in Prevalence of Health Care–Associated Infections in Us Hospitals," *New England Journal of Medicine* (379:18), pp. 1732-1744.
- Magnus, D., Rodgers, S., and Avery, A. 2002. "Gps' Views on Computerized Drug Interaction Alerts: Questionnaire Survey," *Journal of Clinical Pharmacy and Therapeutics* (27:5), pp. 377-382.
- Mani, S., Ozdas, A., Aliferis, C., Varol, H. A., Chen, Q., Carnevale, R., Chen, Y., Romano-Keeler, J., Nian, H., and Weitkamp, J.-H. 2014. "Medical Decision Support Using Machine Learning for Early Detection of Late-Onset Neonatal Sepsis," *Journal of the American Medical Informatics Association* (21:2), pp. 326-336.
- McCoy, A. B., Waitman, L. R., Lewis, J. B., Wright, J. A., Choma, D. P., Miller, R. A., and Peterson, J. F. 2011. "A Framework for Evaluating

- the Appropriateness of Clinical Decision Support Alerts and Responses," *Journal of the American Medical Informatics Association* (19:3), pp. 346-352.
- Meyer, G., Adomavicius, G., Johnson, P. E., Elidrisi, M., Rush, W. A., Sperl-Hillen, J. M., and O'Connor, P. J. 2014. "A Machine Learning Approach to Improving Dynamic Decision Making," *Information Systems Research* (25:2), pp. 239-263.
- Montani, S., and Striani, M. 2019. "Artificial Intelligence in Clinical Decision Support: A Focused Literature Survey," *Yearbook of Medical Informatics* (28:1), pp. 120-127.
- Muir, B. M., and Moray, N. 1996. "Trust in Automation. Part II. Experimental Studies of Trust and Human Intervention in a Process Control Simulation," *Ergonomics* (39:3), pp. 429-460.
- Murdoch, T. B., and Detsky, A. S. 2013. "The Inevitable Application of Big Data to Health Care," *Jama* (309:13), pp. 1351-1352.
- Murray, S. G., Yim, J. W., Croci, R., Rajkomar, A., Schmajuk, G., Khanna, R., and Cucina, R. J. 2017. "Using Spatial and Temporal Mapping to Identify Nosocomial Disease Transmission of *Clostridium Difficile*," *JAMA internal medicine* (177:12), pp. 1863-1865.
- Nanji, K. C., Slight, S. P., Seger, D. L., Cho, I., Fiskio, J. M., Redden, L. M., Volk, L. A., and Bates, D. W. 2014. "Overrides of Medication-Related Clinical Decision Support Alerts in Outpatients," *Journal of the American Medical Informatics Association* (21:3), pp. 487-491.
- Nemati, S., Holder, A., Razmi, F., Stanley, M. D., Clifford, G. D., and Buchman, T. G. 2018. "An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the Icu," *Critical Care Medicine* (46:4), pp. 547-553.
- Norman, G. R. 2000. "The Epistemology of Clinical Reasoning: Perspectives from Philosophy, Psychology, and Neuroscience," *Academic Medicine* (75:10), pp. S127-S133.
- Nussbaum, S., Liberman, N., and Trope, Y. 2006. "Predicting the near and Distant Future," *Journal of Experimental Psychology: General* (135:2), pp. 152-161.
- Nussbaum, S., Trope, Y., and Liberman, N. 2003. "Creeping Dispositionism: The Temporal Dynamics of Behavior Prediction," *Journal of Personality and Social Psychology* (84:3), pp. 485-497.
- Ohno-Machado, L. 2011. "Realizing the Full Potential of Electronic Health Records: The Role of Natural Language Processing," *Journal of the American Medical Informatics Association* (18:5), pp. 539-539.
- Oppenheim, M. I., Vidal, C., Velasco, F. T., Boyer, A. G., Cooper, M. R., Hayes, J. G., and Frayer, W. W. 2002. "Impact of a Computerized Alert During Physician Order Entry on Medication Dosing in Patients with Renal Impairment," *Proceedings of the AMIA Symposium: American Medical Informatics Association*, pp. 577-581.

- Overhage, J. M., Tierney, W. M., Zhou, X.-H., and McDonald, C. J. 1997. "A Randomized Trial of "Corollary Orders" to Prevent Errors of Omission," *Journal of the American Medical Informatics Association* (4:5), pp. 364-375.
- Pasquale, F. 2015. *The Black Box Society*. Harvard University Press.
- Payne, J. W. 1976. "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis," *Organizational Behavior and Human Performance* (16:2), pp. 366-387.
- Pereira, R. D., Fialho, A. S., Cismondi, F., Vieira, S. M., Sousa, J. M., Almeida, R. J., Kaymak, U., Reti, S. R., Howell, M. D., and Finkelstein, S. N. 2011. "Predicting Septic Shock Outcomes in a Database with Missing Data Using Fuzzy Modeling: Influence of Pre-Processing Techniques on Real-World Data-Based Classification," *IEEE*, pp. 2507-2512.
- Peters, G., and Duncan, A. D. 2020. "100 Data and Analytics Predictions through 2024," *Gartner Report* (G00721868).
- Pettit, L. 2013. "Understanding Emram and How It Can Be Used by Policy-Makers, Hospital Cios and Their It Teams," *World hospitals and health services: the official journal of the International Hospital Federation* (49:3), pp. 7-9.
- Phang, C. W., Kankanhalli, A., and Tan, B. C. 2015. "What Motivates Contributors Vs. Lurkers? An Investigation of Online Feedback Forums," *Information Systems Research* (26:4), pp. 773-792.
- Phansalkar, S., Edworthy, J., Hellier, E., Seger, D. L., Schedlbauer, A., Avery, A. J., and Bates, D. W. 2010. "A Review of Human Factors Principles for the Design and Implementation of Medication Safety Alerts in Clinical Information Systems," *Journal of the American Medical Informatics Association* (17:5), pp. 493-501.
- Phansalkar, S., Van der Sijs, H., Tucker, A. D., Desai, A. A., Bell, D. S., Teich, J. M., Middleton, B., and Bates, D. W. 2013. "Drug—Drug Interactions That Should Be Non-Interruptive in Order to Reduce Alert Fatigue in Electronic Health Records," *Journal of the American Medical Informatics Association* (20:3), pp. 489-493.
- Platt, M. L., and Huettel, S. A. 2008. "Risky Business: The Neuroeconomics of Decision Making under Uncertainty," *Nature Neuroscience* (11:4), pp. 398-403.
- Poon, E. G., Keohane, C. A., Bane, A., Featherstone, E., Hays, B. S., Dervan, A., Woolf, S., Hayes, J., Newmark, L. P., and Gandhi, T. K. 2008. "Impact of Barcode Medication Administration Technology on How Nurses Spend Their Time Providing Patient Care," *Journal of Nursing Administration* (38:12), pp. 541-549.
- Popejoy, L. L., Khalilia, M. A., Popescu, M., Galambos, C., Lyons, V., Rantz, M., Hicks, L., and Stetzer, F. 2014. "Quantifying Care Coordination Using Natural Language Processing and Domain-Specific Ontology," *Journal of the American Medical Informatics Association* (22:e1), pp. e93-e103.

- Pronin, E., Olivola, C. Y., and Kennedy, K. A. 2008. "Doing Unto Future Selves as You Would Do Unto Others: Psychological Distance and Decision Making," *Personality and Social Psychology Bulletin* (34:2), pp. 224-236.
- Qian, Y., Hu, Y., Cui, J., Zheng, Q., and Nie, Z. 2011. "Combining Machine Learning and Human Judgment in Author Disambiguation," *Proceedings of the 20th ACM international conference on Information and knowledge management: ACM*, pp. 1241-1246.
- Rader, E., Cotter, K., and Cho, J. 2018. "Explanations as Mechanisms for Supporting Algorithmic Transparency," *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1-13.
- Ransbotham, S. 2016. "Can Artificial Intelligence Replace Executive Decision Making?" *MIT Sloan Management Review*, from <https://sloanreview.mit.edu/article/can-artificial-intelligence-replace-executive-decision-making/>
- Rennie, S., Dalby, M., van Duin, L., and Andersson, R. 2018. "Transcriptional Decomposition Reveals Active Chromatin Architectures and Cell Specific Regulatory Interactions," *Nature Communications* (9:1), p. 487.
- Rhee, C., Dantes, R., Epstein, L., Murphy, D. J., Seymour, C. W., Iwashyna, T. J., Kadri, S. S., Angus, D. C., Danner, R. L., and Fiore, A. E. 2017. "Incidence and Trends of Sepsis in Us Hospitals Using Clinical Vs Claims Data, 2009-2014," *Jama* (318:13), pp. 1241-1249.
- Rhodes, A., Evans, L. E., Alhazzani, W., Levy, M. M., Antonelli, M., Ferrer, R., Kumar, A., Sevransky, J. E., Sprung, C. L., and Nunnally, M. E. 2017. "Surviving Sepsis Campaign: International Guidelines for Management of Sepsis and Septic Shock: 2016," *Intensive Care Medicine* (43:3), pp. 304-377.
- Rocker, G., Cook, D., Sjökvist, P., Weaver, B., Finfer, S., McDonald, E., Marshall, J., Kirby, A., Levy, M., and Dodek, P. 2004. "Clinician Predictions of Intensive Care Unit Mortality," *Critical Care Medicine* (32:5), pp. 1149-1154.
- Romanow, D., Cho, S., and Straub, D. 2012. "Editor's Comments: Riding the Wave: Past Trends and Future Directions for Health It Research," *MIS quarterly*, pp. iii-x.
- Rudin, C. 2019. "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead," *Nature Machine Intelligence* (1:5), pp. 206-215.
- Russ, A. L., Zillich, A. J., Melton, B. L., Russell, S. A., Chen, S., Spina, J. R., Weiner, M., Johnson, E. G., Daggy, J. K., and McManus, M. S. 2014. "Applying Human Factors Principles to Alert Design Increases Efficiency and Reduces Prescribing Errors in a Scenario-Based Simulation," *Journal of the American Medical Informatics Association* (21:e2), pp. e287-e296.

- Sadler-Smith, E., and Shefy, E. 2004. "The Intuitive Executive: Understanding and Applying 'Gut Feel' in Decision-Making," *Academy of Management Perspectives* (18:4), pp. 76-91.
- Schedlbauer, A., Prasad, V., Mulvaney, C., Phansalkar, S., Stanton, W., Bates, D. W., and Avery, A. J. 2009. "What Evidence Supports the Use of Computerized Alerts and Prompts to Improve Clinicians' Prescribing Behavior?," *Journal of the American Medical Informatics Association* (16:4), pp. 531-538.
- Schmiedel, T., Müller, O., and vom Brocke, J. 2018. "Topic Modeling as a Strategy of Inquiry in Organizational Research: A Tutorial with an Application Example on Organizational Culture," *Organizational Research Methods*, pp. 1-28.
- Sendak, M., Elish, M. C., Gao, M., Futoma, J., Ratliff, W., Nichols, M., Bedoya, A., Balu, S., and O'Brien, C. 2020. "" The Human Body Is a Black Box" Supporting Clinical Decision-Making with Deep Learning," *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 99-109.
- Seymour, C. W., Gesten, F., Prescott, H. C., Friedrich, M. E., Iwashyna, T. J., Phillips, G. S., Lemeshow, S., Osborn, T., Terry, K. M., and Levy, M. M. 2017. "Time to Treatment and Mortality During Mandated Emergency Care for Sepsis," *New England Journal of Medicine* (376:23), pp. 2235-2244.
- Shah, N. R., Seger, A. C., Seger, D. L., Fiskio, J. M., Kuperman, G. J., Blumenfeld, B., Recklet, E. G., Bates, D. W., and Gandhi, T. K. 2006. "Improving Acceptance of Computerized Prescribing Alerts in Ambulatory Care," *Journal of the American Medical Informatics Association* (13:1), pp. 5-11.
- Shani, Y., Igou, E. R., and Zeelenberg, M. 2009. "Different Ways of Looking at Unpleasant Truths: How Construal Levels Influence Information Search," *Organizational Behavior and Human Decision Processes* (110:1), pp. 36-44.
- Shashikumar, S. P., Stanley, M. D., Sadiq, I., Li, Q., Holder, A., Clifford, G. D., and Nemati, S. 2017. "Early Sepsis Detection in Critical Care Patients Using Multiscale Blood Pressure and Heart Rate Dynamics," *Journal of electrocardiology* (50:6), pp. 739-743.
- Sheikhalishahi, S., Miotto, R., Dudley, J. T., Lavelli, A., Rinaldi, F., and Osmani, V. 2019. "Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review," *JMIR Medical Informatics* (7:2), p. e12239.
- Shen, C.-C., and Chiou, J.-S. 2010. "The Impact of Perceived Ease of Use on Internet Service Adoption: The Moderating Effects of Temporal Distance and Perceived Risk," *Computers in Human Behavior* (26:1), pp. 42-50.
- Shortliffe, E. H., and Sepúlveda, M. J. 2018. "Clinical Decision Support in the Era of Artificial Intelligence," *Jama* (320:21), pp. 2199-2200.

- Sinuff, T., Adhikari, N. K., Cook, D. J., Schünemann, H. J., Griffith, L. E., Rocker, G., and Walter, S. D. 2006. "Mortality Predictions in the Intensive Care Unit: Comparing Physicians with Scoring Systems," *Critical Care Medicine* (34:3), pp. 878-885.
- Slight, S. P., Seger, D. L., Franz, C., Wong, A., and Bates, D. W. 2018. "The National Cost of Adverse Drug Events Resulting from Inappropriate Medication-Related Alert Overrides in the United States," *Journal of the American Medical Informatics Association* (25:9), pp. 1183-1188.
- Small, S. G., and Medsker, L. 2014. "Review of Information Extraction Technologies and Applications," *Neural computing and applications* (25:3-4), pp. 533-548.
- Smith, P. K., and Trope, Y. 2006. "You Focus on the Forest When You're in Charge of the Trees: Power Priming and Abstract Information Processing," *Journal of Personality and Social Psychology* (90:4), pp. 578-596.
- Sox, H. C., Higgins, M. C., and Owens, D. K. 2013. *Medical Decision Making*. UK: John Wiley & Sons.
- Strom, B. L., Schinnar, R., Abera, F., Bilker, W., Hennessy, S., Leonard, C. E., and Pifer, E. 2010a. "Unintended Effects of a Computerized Physician Order Entry Nearly Hard-Stop Alert to Prevent a Drug Interaction: A Randomized Controlled Trial," *Archives of Internal Medicine* (170:17), pp. 1578-1583.
- Strom, B. L., Schinnar, R., Bilker, W., Hennessy, S., Leonard, C. E., and Pifer, E. 2010b. "Randomized Clinical Trial of a Customized Electronic Alert Requiring an Affirmative Response Compared to a Control Group Receiving a Commercial Passive Cpoe Alert: Nsaid—Warfarin Co-Prescribing as a Test Case," *Journal of the American Medical Informatics Association* (17:4), pp. 411-415.
- Sweeney, T. E., Perumal, T. M., Henao, R., Nichols, M., Howrylak, J. A., Choi, A. M., Bermejo-Martin, J. F., Almansa, R., Tamayo, E., and Davenport, E. E. 2018. "A Community Approach to Mortality Prediction in Sepsis Via Gene Expression Analysis," *Nature Communications* (9:1), p. 694.
- Tam, K. Y., and Ho, S. Y. 2005. "Web Personalization as a Persuasion Strategy: An Elaboration Likelihood Model Perspective," *Information systems research* (16:3), pp. 271-291.
- Taylor, L., and Tamblyn, R. 2004. "Reasons for Physician Non-Adherence to Electronic Drug Alerts," *Medinfo* (11), pp. 1101-1105.
- Thiel, S. W., Rosini, J. M., Shannon, W., Doherty, J. A., Micek, S. T., and Kollef, M. H. 2010. "Early Prediction of Septic Shock in Hospitalized Patients," *Journal of Hospital Medicine* (5:1), pp. 19-25.
- Timmermans, S., and Mauck, A. 2005. "The Promises and Pitfalls of Evidence-Based Medicine," *Health Affairs* (24:1), pp. 18-28.
- Trope, Y., and Liberman, N. 2010. "Construal-Level Theory of Psychological Distance," *Psychological Review* (117:2), pp. 440-463.

- Trope, Y., Liberman, N., and Wakslak, C. 2007. "Construal Levels and Psychological Distance: Effects on Representation, Prediction, Evaluation, and Behavior," *Journal of Consumer Psychology* (17:2), pp. 83-95.
- van der Poll, T., van de Veerdonk, F. L., Scicluna, B. P., and Netea, M. G. 2017. "The Immunopathology of Sepsis and Potential Therapeutic Targets," *Nature Reviews Immunology* (17:7), pp. 407-420.
- Van Der Sijs, H., Aarts, J., Vulto, A., and Berg, M. 2006. "Overriding of Drug Safety Alerts in Computerized Physician Order Entry," *Journal of the American Medical Informatics Association* (13:2), pp. 138-147.
- Wakslak, C. J., Nussbaum, S., Liberman, N., and Trope, Y. 2008. "Representations of the Self in the near and Distant Future," *Journal of Personality and Social Psychology* (95:4), pp. 757-773.
- Wakslak, C. J., Trope, Y., Liberman, N., and Alony, R. 2006. "Seeing the Forest When Entry Is Unlikely: Probability and the Mental Representation of Events," *Journal of Experimental Psychology: General* (135:4), pp. 641-653.
- Wallach, H. M. 2006. "Topic Modeling: Beyond Bag-of-Words," pp. 977-984.
- Wang, J., and Cole, C. A. 2016. "The Effects of Age and Expertise on Product Evaluations: Does the Type of Information Matter?," *Management Science* (62:7), pp. 2039-2053.
- Wang, W., and Benbasat, I. 2007. "Recommendation Agents for Electronic Commerce: Effects of Explanation Facilities on Trusting Beliefs," *Journal of Management Information Systems* (23:4), pp. 217-246.
- Weick, K. E., and Roberts, K. H. 1993. "Collective Mind in Organizations: Heedful Interrelating on Flight Decks," *Administrative science quarterly*, pp. 357-381.
- Weingart, S. N., Toth, M., Sands, D. Z., Aronson, M. D., Davis, R. B., and Phillips, R. S. 2003. "Physicians' Decisions to Override Computerized Drug Alerts in Primary Care," *Archives of Internal Medicine* (163:21), pp. 2625-2631.
- Wilson, F. P., Shashaty, M., Testani, J., Aqeel, I., Borovski, Y., Ellenberg, S. S., Feldman, H. I., Fernandez, H., Gitelman, Y., and Lin, J. 2015. "Automated, Electronic Alerts for Acute Kidney Injury: A Single-Blind, Parallel-Group, Randomised Controlled Trial," *The Lancet* (385:9981), pp. 1966-1974.
- Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT press.
- Xia, B., Zhao, D., Wang, G., Zhang, M., Lv, J., Tomoiaga, A. S., Li, Y., Wang, X., Meng, S., Cooke, J. P., Cao, Q., Zhang, L., and Chen, K. 2020. "Machine Learning Uncovers Cell Identity Regulator by Histone Code," *Nature Communications* (11:1), p. 2696.
- Xiao, B., and Benbasat, I. 2007. "E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact," *MIS quarterly* (31:1), pp. 137-209.

- Xu, H., Jiang, M., Oetjens, M., Bowton, E. A., Ramirez, A. H., Jeff, J. M., Basford, M. A., Pulley, J. M., Cowan, J. D., and Wang, X. 2011. "Facilitating Pharmacogenetic Studies Using Electronic Health Records and Natural-Language Processing: A Case Study of Warfarin," *Journal of the American Medical Informatics Association* (18:4), pp. 387-391.
- Yang, H., Spasic, I., Keane, J. A., and Nenadic, G. 2009. "A Text Mining Approach to the Prediction of Disease Status from Clinical Discharge Summaries," *Journal of the American Medical Informatics Association* (16:4), pp. 596-600.
- Yeow, A., and Goh, K. H. 2015. "Work Harder or Work Smarter? Information Technology and Resource Allocation in Healthcare Processes," *MIS Quarterly* (39:4).
- Zauberman, G., Kim, B. K., Malkoc, S. A., and Bettman, J. R. 2009. "Discounting Time and Time Discounting: Subjective Time Perception and Intertemporal Preferences," *Journal of Marketing Research* (46:4), pp. 543-556.
- Zhang, C., and Ma, Y. 2012. *Ensemble Machine Learning: Methods and Applications*. Springer.