Transdisciplinary Learning Lab to Eliminate Patient Harm and Reduce Waste

Johns Hopkins Medicine

Principal Investigator: Adam Sapirstein, MD

Project Team Members:

Ravi Aron, PhD

Noah Barasch, MS

Howard Carolan, MBA, MPH

Zoe Demko, BS

Cindy Dwyer, RN, BSN

Simon Mathews, MD

Philip Phan, PhD

Peter Pronovost, MD, PhD

Alan Ravitz, PhD

Mark Romig, MD

Michael Rosen, PhD

Grants Management Specialist: GALEN GREGOR
Email: galen.gregor@ahrq.hhs.gov Phone: 301-427-1457

Program Official: DAVID RODRICK

Supported by funding from Agency for Healthcare Research & Quality

Grant Award Number: 5P30HS023553-02

Period of Performance: 09/30/2014 - 03/29/2019

Project 1: Concepts and Requirements Thinking for an Ideal ICU

Abstract

Purpose: Patient safety issues lead to hundreds of thousands of patient deaths and many more preventable harms per year. We hypothesize that most of these adverse events could be prevented by designing a system to accomplish these goals.

Scope: The ICU was selected as the care setting of interest because of its complexity and the acute nature of patients. To further focus the research effort, the problem of pressure ulcers was selected because of the broad adverse impact this condition causes.

Methods: We outlined a systems engineering program known as the AI Patient Safety Learning Lab Framework based on previously successful models used by the military. We further applied these concepts to Pressure Ulcer Prevention (PUP) using established systems engineering tools. The PUP healthcare environment was analyzed in the presence of key healthcare stakeholders to devise a Concept of Operations (PUP CONOPs). The PUP CONOPs was used to guide design thinking activities and aiding in identification, prioritization, and evaluation of potential PUP solution concepts.

Results: Concepts for HRICU operational needs were outlined. A PUP CONOPs outlined key operational needs, metrics, and use cases for pressure ulcer prevention. Design thinking days yielded nine concepts for further exploration, which were scored in priority order. Finally, the PUP CONOPs evaluation provided recommendations for two existing solution concepts for integration into a system to predict, prevent, and treat pressure ulcers.

Key Words: High-reliability intensive care unit (HRICU), systems engineering, patient safety, pressure ulcers, concept of operations (CONOPS), Quality Function Deployment (QFD)

Purpose

Patient safety issues lead to hundreds of thousands of patient deaths and many more preventable harms per year. It is hypothesized that most of these adverse events could be prevented if a system that worked for the patients, providers, and other stakeholders was developed and implemented. We refer to this as the high-reliability ICU (HRICU). We hypothesize that systems engineering methods can provide healthcare with a model for iterating a broad objective of the HRICU. To this end, the objectives are:

Objectives:

1. Create a systems engineering (SE)-guided framework that can be used to create a high-reliability ICU in an iterative manner
2. Apply SE framework to address a common, preventable harm in the intensive care unit (ICU)
3. Draft a Systems Engineering Concept of Operations for the model harm case study

Scope

Healthcare has taken an approach to improvement that has been effective but lacks coordination and scalability. Systems engineering has proved valuable when systems become complex in technical, workflow, and information management. Although the HITech Act has moved the industry toward interoperability, for instance, solutions must be planned and must lead to integrated and affordable solution. Despite the recognition that systems engineering tools can help, healthcare leaders still struggle with many patient safety issues that lead to hundreds of thousands of patient deaths and many more preventable harms per year.[1]

Much of the problem lies in the way the current system was built and now operates. Healthcare systems have evolved with a high dependence on expertise, and harms are largely addressed individually (e.g., sepsis). This is expensive and ineffective. In many areas, standardization could replace expertise to great improvement. Our comprehensive program plan incorporates rapid prototyping and systems engineering methods, using a model developed for the US Navy's submarine force.[2] Our goal is to create a Systems Engineering guided framework to eventually create a high-reliability ICU (HRICU).

To accomplish this larger goal, we began by applying the Systems Engineering guided framework to address a common preventable harm in the ICU. The preventable harm we selected is that of pressure ulcers because of its large adverse impact. Pressure ulcers are an epidemic problem in the United States, affecting 2.5 million people; they are associated with 60,000 deaths and are estimated to cost $9.1-$11.6 billion annually.[3,4] Care costs and patient length of stay increase and quality of life decreases with each occurrence of a pressure ulcer.[5] Thus, pressure ulcers represented a meaningful challenge area to demonstrate how rapid prototyping, human-centered design, and systems engineering tools could collectively facilitate the solution process in healthcare.

Methods

Development of the Learning Lab Framework - We developed a comprehensive program plan, identifying the sequence of activities required for integration of a healthcare system of systems. The program plan is based on a similar plan developed by the US Navy's submarine force. This Navy program capitalizes on design thinking and an open business model in a constrained budget environment.[6] We outline a simplified programmatic framework to serve as the AI Patient Safety Learning Lab Framework in Figure 1, which served as a system development lifecycle (SDLC) to engineer an HRICU. We used human-centered design concepts, including a concept poster and storyboarding, to create an HRICU concept plan. To then evaluate this framework SDLC process, including its associated SE tools, we applied it to a well-known, yet enduring, patient safety issue: pressure ulcer prevention. We pursued framework steps 1 through 4 in Figure 1 in the context of pressure ulcer prevention. In the ideation phase (Figure 1, Step 2), we solicited the input of multiple stakeholders, including doctors, nurses, administrators, patients and families, patient advocates, engineers, researchers, specialty services, support services, payers, and others familiar with academic, industrial, and regulatory work pertaining to pressure ulcers, to develop relevant content for the Project 1 products.

Establishing a CONOPS for a system to predict, prevent, identify, and treat pressure ulcers in the HRICU - An initial CONOPS template tailored for healthcare was derived by performing a literature search to identify existing 1) research on standardized approaches to CONOPS generation, 2) CONOPS development templates, and 3) healthcare-specific CONOPS activities. Once completed, stakeholder engagement was performed, consisting of four primary efforts: 1) a recurring bi-weekly transdisciplinary team meeting that spanned 3 months, 2) individual engagements with stakeholders to further document technical and operational context, 3) review of the literature on pressure ulcer prevention to gather additional perspectives from stakeholders beyond our immediate reach, and 4) design thinking day to solicit candidate solutions ideas. As a final step, methods and findings from the pressure injury use case were fed back into the CONOPS (i.e., Step 3 and 4 below) to outline its application to Pressure Ulcer Prevention.

Call for Innovation - We held an IDEO-style design thinking event to brainstorm new ideas for pressure ulcer prediction, prevention, or mitigation.[7] Evaluative criteria for candidate solutions were developed based upon operational end user needs described in the CONOPS.

Quality Function Deployment (QFD) - A second series of workshops was organized to solicit input from the stakeholders to establish the quantitative metrics that would be used to score candidate solutions using a standard, SE industry-accepted method known as Quality Function Deployment (QFD). We selected the QFD process, a standard Systems Engineering method, to score candidate solutions for pressure ulcer prevention. The QFD is a seven-step process for the selection of candidate solutions based on effectiveness, performance, and resource constraints.[8] It results in a solution score that allows ranking of potential solutions. The SE Quality Function Deployment (QFD) technique was conducted over the course of several workshops, utilizing the clinical expertise key stakeholders in wound care nursing and pressure ulcer prevention, and was administered by career Systems Engineers at the Johns Hopkins Applied Physics Laboratory.

Evaluation of candidate Pressure Ulcer Solution Space using the Learning Lab Model - We performed a market survey of existing products and current research for pressure ulcer prevention, but no formal industry solicitation was made. We created a test plan and held a simulation event to evaluate proposed solutions and their integration in the existing system of care using the CONOPS.

Results

The stepwise AI Patient Safety Learning Lab framework outlines a path to accomplish our goal of re-engineering healthcare; our project plan includes a series of interdependent stages depicted in Figure 1. These stages provide a rich innovation pipeline that couples concept and requirements development through laboratory-based assessment and transition of requirements to industry. Through this pipeline, best-of-breed solutions can be provided to industry for scaling and dissemination for wider impact. As depicted in Figure 1, seven major stages are associated with this approach.

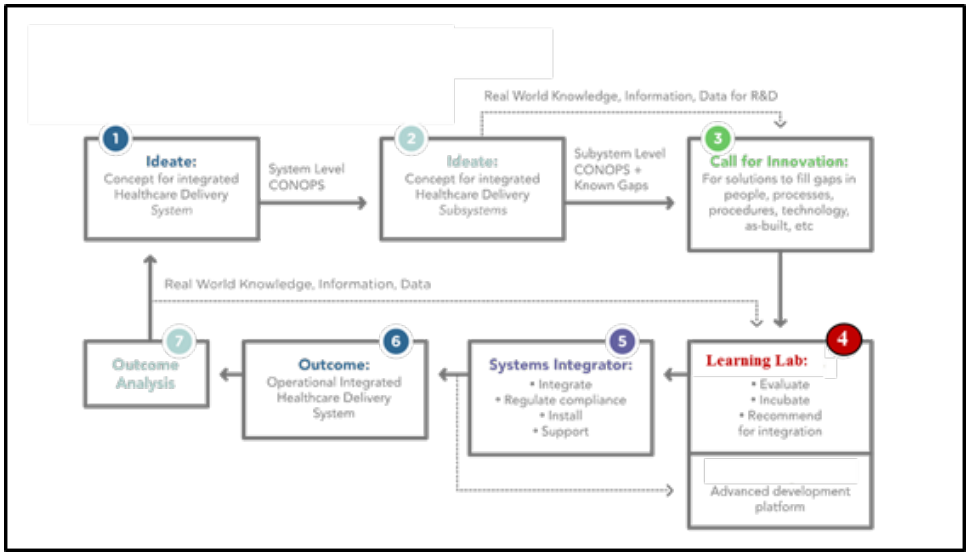


Figure 1: Development Process and the Role of the AI Patient Safety Learning Lab

Concept for Integrated Healthcare Delivery System: HRICU - We drafted a storyboard that captured 12 common tasks in the ICU. These are (1) admit the patient; (2) develop patient plan for diagnosis, treatment, and recovery; (3) communicate the plan, roles and responsibilities, and expectations to the care team; (4) monitor and assess progress; (5) coordinate care; (6) promote a healing environment; (7) prevent harms; (8) respond to patient needs; (9) discharge or transfer patient; (10) measure outcomes; (11) update processes and workflows; and (12) nurture a culture of learning and accountability. Using the storyboard, concept operational needs for the HRICU were identified and included (1) patient-centered care; (2) a healing environment; (3) actionable information; (4) coordination of care; (5) harm prevention; (6) resilient resources; and (7) learning and accountability.

Establishing a CONOPS for a system to predict, prevent, identify, and treat pressure ulcers in the HRICU - A comprehensive CONOPS was created as a communications tool for the team. A CONOPS defines the problem to be solved, lists high-level requirements, identifies the relevant metrics, and includes notable use cases. The CONOPS serves as a guide, providing the necessary information to ensure stakeholder concepts are evaluated through the lens of the system requirements and operational needs.

Call for Innovation

Design Thinking Results - We conducted a design thinking event with stakeholder input to brainstorm solution spaces for the pressure ulcer problem in the ICU. We used individual group voting, rapid prototyping, and cost versus impact mapping to distill the brainstorming content into candidate ideas for implementation. The output included nine concepts, six prototypes, and five distilled ideas that were later used to exercise the evaluative QFD method selected for the project. The ideas selected did not provide any defensible justification regarding why they were selected. The CONOPS was applied to how solutions would be evaluated using the QFD process defined below.

Quality Function Deployment (QFD) - Though design thinking days are a productive means of brainstorming innovations, they often yield numerous outputs that require rigorous analysis to filter and prioritize solutions for which to devote precious resources. To go from the divergent solution set generated in design thinking to a convergent solution set that would fulfill the CONOPs requirements, we performed QFD.

We followed the seven-step process to generate Measures of Effectiveness (MOEs) and Measures of Performance (MOP), to identify a list of nine different candidate solutions. These solutions that ranged in maturity from capabilities presently available at our institution to those requiring new research and development. We scored the candidate solutions using and an MOE/MOP scoring matrix and were able to create a decision support ranking grid seen in Figure 2.

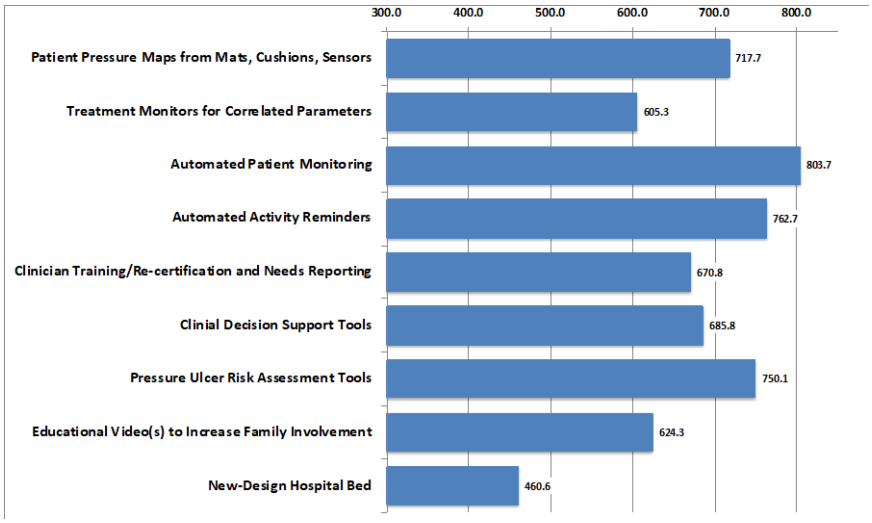


Figure 2 - Candidate pressure ulcer solutions and their associated QFD solution scores. The nine solutions that were evaluated using QFD are shown on the vertical axis. The overall impact score of each solution based on the user-defined measures is displayed on the horizontal axis.

The solutions spanned the categories of people, processes, and technology. Based on this analysis, the following solution concepts were identified in priority order for further use and/or development:

(1) Automated activity reminders;
(2) Pressure ulcer risk assessment tools;
(3) Clinician training/re-certification and needs reporting;
(4) Educational videos; and
(5) Patient pressure maps.

Evaluation of candidate Pressure Ulcer Solution Space using the Learning Lab Model - We conducted a limited market survey and institutional outreach and learned of several initiatives to prevent pressure ulcers. The first effort was an educational program assembled for Johns Hopkins Hospital. The educational program was evaluated using the CONOPs requirements, and this evaluation was provided to the program leaders. The educational program meets many of the requirements in the CONOPs and can be integrated into a larger Pressure Ulcer Prevention system.

A second initiative was an invention to detect and alert providers of pressure ulcer risks on an individual patient. This capability was viewed as a representative solution that might be received as part of a Learning Lab framework. The project team evaluated the device per CONOPs and provided guidance to the vendor on improvements to ensure its optimal integration into healthcare workflows. Though desirable, the novel sensor technology remains unproven and has not reached a point of maturity where it can be incorporated into the Pressure Ulcer Prevention system. The CONOPs could be used by the developers to expedite commercialization.

Conclusion - Quality improvement efforts are largely siloed, addressing individual harms while often pursuing solution spaces that do not consider downstream impact to other healthcare subsystems. Systems Engineering has been used successfully in other industries to deal with such complexities. Our program plan was defined in order to outline and demonstrate the application of a healthcare-specific SDLC (the AI Patient Safety Learning Lab Framework).

The AI Patient Safety Learning Lab proved a feasible path for applying design thinking and rapid prototyping methods in tandem with rigorous SE methods to brainstorm and analyze system products and capabilities for implementation. Future work will demand additional resources to scale the Learning Lab framework in order to address all harms on the path to an HRICU.

Project 2: Leverage Open Application Programming Interfaces (OAPI) to Eliminate Patient Harm, Optimize Patient Outcomes and Experience, and Reduce Waste

Structured Abstract

Purpose: The administration of high-alert medications requires the use of systems to prevent errors. The common system that requires independent verification of dose changes by a second clinician is error prone and inefficient. We postulate that the ability to integrate interoperable medication infusion pumps and medical records systems with existing dose adjustment algorithms can improve the safety and efficiency in the delivery of high-alert and other medications.

Scope: A nurse-managed insulin infusion protocol used at the Johns Hopkins Hospital, a test instance of the Epic Medical Record, and a commercially available infusion pump with open control access were selected to test and demonstrate our hypothesis.

Methods: We followed stepwise systems engineering practices with a plan to develop a novel system, the Smart Agent, to semi-autonomously administer intravenous insulin according to our hospital's protocol for nurse-managed insulin infusion in the intensive care units (ICUs); 20 critical care nurses were presented with 12 simulated scenarios (six each between manual and Smart Agent) to assess accuracy, efficiency, workload, usability, and trust in the new system.

Results: Out of 120 scenarios, no errors were made with the Smart Agent system, and 20 errors were made with the manual system (16.6% of the 120 scenarios completed using standard of care). Participants rated the usability and workload of the Smart Agent significantly higher than the manual practice, and they rated trust similarly.

Key Words: interoperability, infusion pumps, medication error, medication administration, independent double-check, integration, insulin infusion

Purpose

This study compares the safety and efficiency of standard-of-care (SOC)/manual management of a modified Yale insulin infusion protocol with a bidirectional pump communication (Smart Agent) support tool. The Smart Agent tool supports bedside management of the insulin infusion protocol to improve safety and efficiency and reduces workload while increasing perceptions of usability of the technologies involved in the process. This study evaluated the effectiveness of this intervention at meeting those design criteria and generating additional formative input from frontline nurses, the ultimate end users of the system in development. This research will help characterize the potential benefits from these technologies and identify any latent risks to their introduction into actual care.

Objectives:

1. Develop a human in-the-loop infusion pump system using clinical decision support and command and control functionality to conduct a nurse-managed insulin infusion protocol
2. Compare automated algorithm results with clinical examples to validate software
3. Measure time and compliance with current, nurse-managed protocol for insulin infusion
4. Demonstrate that automation can be used to eliminate the need for an independent nurse double check.

Scope

Medication administration errors remain the leading category of adverse patient safety events. In particular, high-risk medication errors may be life threatening.[9,10,11] The majority of medication errors in the ICU occur during administration, and the leading causes include errors in documentation, failure to follow protocol, and communication.[12] Current risk mitigation strategies to prevent medication administration errors often require human double checks, enhanced supervision by clinicians, increasing workload and putting a strain on productivity.[13] A safe, productive system is possible today through clever use of technology. Open application programmer interfaces are increasingly available, though many systems still have proprietary interfaces or lack thoughtful integration; thus, devices are not interoperable. It is difficult to create decision support, predict patient risk, and monitor and improve performance. The potential of systems enabled through open interfaces to contribute to the elimination of patient harm is profound. From previous design thinking sessions, our team envisioned a bidirectional infusion pump system that would leverage open application programmer interfaces to improve the safety and efficiency of high-risk medication administration workflows. We hypothesized that such a system would reduce errors and workload, thus eliminating the need for an independent nurse double check.

We convened clinical stakeholders, including intensive care physicians, critical care nurses, and pharmacists, from Johns Hopkins Hospital in our effort to design, engineer, and test a bidirectional infusion pump system that leverage open interfaces. Subjects were observed in clinical settings or convened during initial design sessions and early prototype review meetings. Further, Johns Hopkins Hospital critical care nurses were subjects of a voluntary simulation study to evaluate the safety and efficiency of the project's technology. All subjects participated voluntarily in the study activities.

Methods

Study Design - Principles of the Systems Engineering Development Lifecycle were tailored to conduct concept design, engineering, and laboratory validation phases of the proposed interoperable infusion pump system, as described below.

Design & Engineering - The study team conducted design thinking meetings to brainstorm and visualize potential integration solutions for bidirectional infusion pump systems. The solutions were developed to address the specific use cases of 1) managing patient serum glucose levels with insulin infusions and 2) managing patient postoperative pain using intravenous narcotics delivered with patient-controlled analgesia (PCA) systems. The purpose of these sessions was to identify design requirements and specifications for a candidate solution. Attendees included pharmacists, nurses, intensive care physicians, and engineers from various disciplines. Additionally, workflow analysis was performed through direct observation of critical care nurses who were performing insulin dose calculations and adjustments by a human-factors psychologist. Observations were tracked using a standardized data collection tool. We chose to pursue development of a bidirectional infusion system using the nurse-managed insulin workflow due to its simplistic, clinically validated algorithm, requirement for the nurse double check, and single source of clinical data in the EHR.

Verification of Insulin Infusion Prototype - A prototype system known as Smart Agent was developed based on user requirements for a modified nurse-managed Yale Insulin Infusion Protocol used at the Johns Hopkins Hospital. The pilot solutions applied components of the Medical Infusion Pump (MIP) software that was previously co-developed by APL and the Armstrong Institute under AHRQ funding (Grant No. HS20460). This full-stack program with user interface (Figure 4) was subsequently coupled to the MIP and open interface VistA medical record. The project team created a technology demonstration to simulate the bidirectional communication and control of insulin infusions by the Smart Agent algorithm using test scenarios to explore the full range of clinical situations. The project clinicians made system improvement recommendations with regard to design, appearance, and safety vulnerabilities. These recommendations were prioritized and implemented in the final model of the prototype for further testing.

System Finalization - User feedback was incorporated into a final set of requirements for the Smart Agent insulin infusion system. To optimize integration of the Smart Agent system into current workflows, the project team collaborated with the Epic Medical Record team at the Johns Hopkins Health System to embed the clinical decision support interface into the medical record. Additionally, Smart Agent was interfaced with a commercially available infusion pump, which permitted access to its electronic controller (Hospira PLUM 360).

Validation - To demonstrate safety and efficiency gains of the Smart Agent infusion system, clinician users were subjects of simulated scenarios comparing the standard of care workflow with the newly engineered system. Safety was measured as the rate of errors in calculating any necessary rate changes. A correct rate change was calculated using the infusion protocol. An observer logged the new rate calculation of the nurse for each scenario, and this was compared to the correct rate change. Efficiency was measured as the time to complete each scenario. Timing data were captured by an observer using a tablet timing application. Workload was measured by participant perceptions of effort. The NASA Task Load Index (TLX)[14] was administered after each block of scenarios. Completing the NASA-TLX involves two steps. User trust and perceptions of usability were measured using two surveys. The Systems Usability Scale (SUS) is an 11-item survey, widely used in research and commercial product development.[15] Trust in the system was measured using a previously validated measure of trust in automation.[16] For the repeated measures in this study (efficiency, workload, and perceived trust), we used a repeated measures, within-factors ANOVA, and the F statistic to test for differences between the two conditions (i.e., manual standard-of-care or Smart Agent system). For perceived usability, which was completed only at one time point, a paired samples t-test was conducted to determine differences in trust and perceived usability.

Participants' reactions to the two systems were recorded in session notes during the structured debrief and qualitative thematic analysis was performed. Potential barriers to use of the Smart Agent system or ideas for further refinement of the system were identified.

Results

Design and Engineering - The final version of the SA integrates intelligent clinical decision support with the EMR and the medication infusion pump. SA is launched from the EMR as a web service call and displayed within a window of the active record. Key design features are the following:

1. Easily accessible, within workflow, clinical decision support tool that automates the dosing calculation – The SA extracts patient's lab and insulin infusion data from the EMR, calculates a dose rate change using these and the imbedded Yale algorithm, and presents the dose rate change result to the user.

A clinician can accept and use this recommendation as determined by regulations and institutional policy. An important feature of the SA is the graphical mapping of the dose selection criteria onto the protocol. This gives the clinician the ability to see exactly how the insulin dose was determined. Importantly, these design features eliminate the need for an independent nurse double check.

2. Automated pump programming and EMR documentation with nurse verification - SA can be used to directly change the insulin dose by sending the automated rate calculation to the pump once a dose rate change is accepted by the nurse. By design, the SA is not fully automated. The SA requires a clinician to accept and approve the updates but reduces manual steps, including error-prone manual calculations and keying in values to the pump and medical record. By decoupling the functions of dose determination from changes made to the infusion pump and EMR, we have created a system that could be used in current workflow for CDS from the system that would require regulatory approval for use.

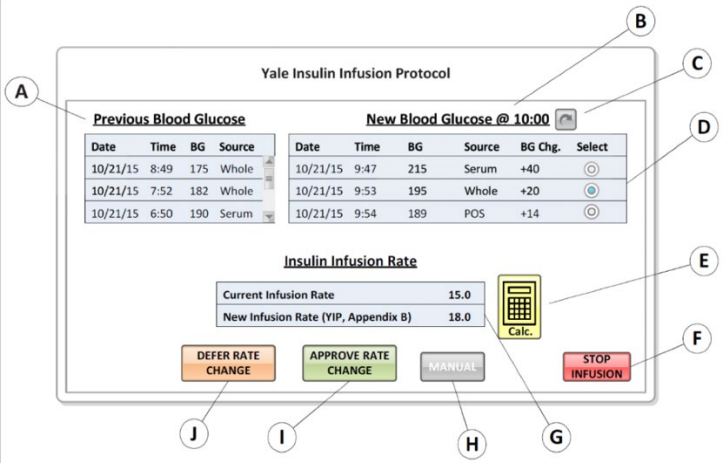


Figure 3: Smart Agent User Interface Prototype. This user interface mockup was presented to clinical users on the prototype system. A) Table of last three glucose values; B) most recent glucose values with C) refresh button and D) lab select button. E) Users can view the calculations embedded in the protocol or F) stop the infusion. G) The current and recommended infusion rates are displayed. The new insulin rate can be H) entered manually, I) accepted, or J) deferred.

Validation - Twenty critical care nurses completed the evaluation. Participants ranged from less than 1 to 37 years of nursing experience (M = 5.031, SD = 8.846) and from less than 1 to 35 years of critical care nursing experience (M = 4.481, SD = 8.602). They most commonly reported having patients on an insulin infusion protocol once per month (N=10, 50%), but seven participants (35%) reported having these patients once per week, and three participants (15%) reported having these patients every shift or more than once per week. Figure 4 illustrates main study findings for efficiency, workload, usability, and trust.

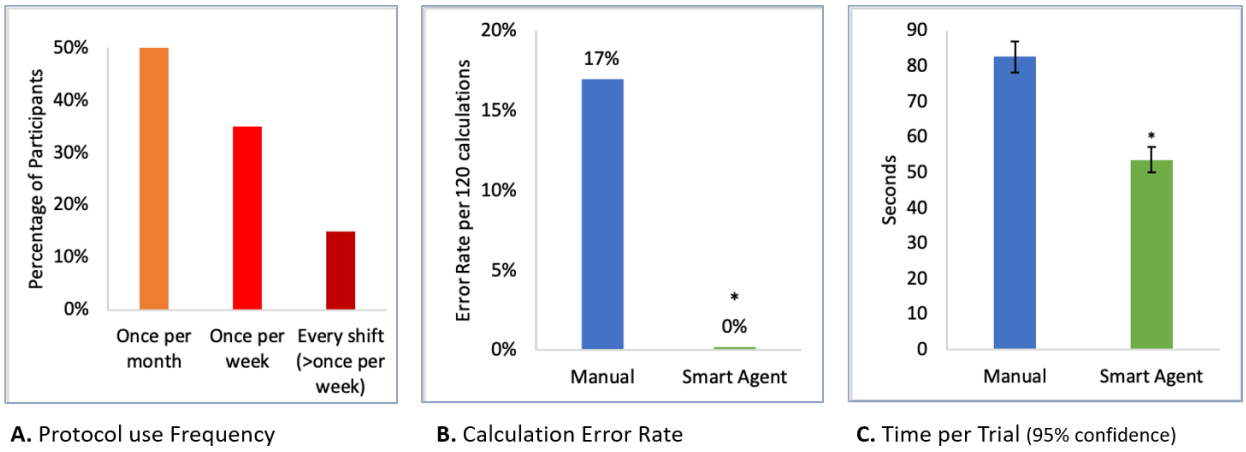
Safety - The rate of errors in calculating a new infusion rate was significantly different between the two systems, with the Smart Agent performing better than manual ($X^2 = 21.82$, df=1, $p<.001$). Out of 120 scenarios, no errors were made with the Smart Agent system, but 20 errors were made with the manual system (16.6% of the 120 scenarios completed using standard of care).

Efficiency - Participants completed the protocol significantly faster using the Smart Agent system (M=58.24 seconds per trial, SD=22.67) than with the manual standard-of-care system (M=91.32, SD=29.86; $F(1,244)=100.44$; $p < .001$). There was a mean difference of 33.08 seconds, favoring the Smart Agent condition.

Workload - Participants rated the overall workload of the Smart Agent system (M=6.33, SD=2.38) significantly lower than the manual system (M=10.23, SD=2.74; F(1,74)=44.39; p<.001).

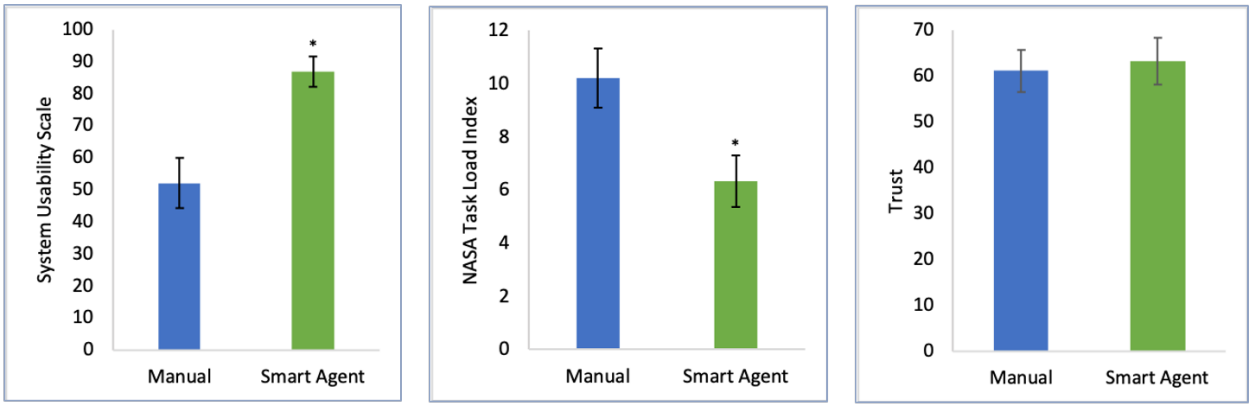
Due to technical issues with the eye tracking system (e.g., problems with system calibration or general functioning), data were usable for only 10 scenarios (59%) across 17 participants. Three participants were excluded from using the eye tracking system due to wearing eyeglasses.

Perceptions of trust and usability - Participants rated the usability of the Smart Agent system (M=86.88, SD=10.73) significantly higher than that of the manual system (M=52.00, SD=17.97; t (19)=-6.23; p<.001). Participants did not rate their level of trust as significantly different between the Smart Agent system (M=63.33, SD=12.17) and the manual system (M=61.13, SD=11.31).



A. Protocol use Frequency **B.** Calculation Error Rate **C.** Time per Trial (95% confidence)

Figure 4. Technical Performance of Smart Agent-Nurse System. A. Percentage of Participants by frequency of performance of nurse-managed insulin infusion protocol in clinical practice. B. Comparison of calculation error rates in simulation activities between manual nurse-managed insulin infusion protocol and Smart Agent system. C. Comparison of time to perform protocol between manual and Smart Agent system.



A. Usability (95% confidence) **B.** Workload (95% confidence) **C.** Trust (95% confidence)

Figure 5. Smart Agent Clinical Users Survey Results. A. Usability of manual protocol vs. Smart Agent on the System Usability Scale out of 100. B. Comparison of workload between manual and Smart Agent using NASA-TLX survey. C. Comparison of user reported trust between the manual and Smart Agent system.

Discussion - We believe that there are a lot of opportunities to improve care and decrease clinical workloads using medical infusion pump integration. Our team selected a high-risk, nurse-managed infusion in which serial dose calculations, pump programming, and medical record documentation are, in-whole or in-part, manual practices. Smart Agent addresses error prone steps in this workflow while reducing workload demands. Automated rate calculations eliminate the potential for manual calculation and pump programming errors, reduce the number of overall steps, and should eliminate the need for an independent nurse double check.
As evidenced by survey respondents, our user-centered design resulted in a system that clinicians find safe and usable. By localizing relevant data and actionable steps to a given screen, we have reduced overall cognitive workload and increased compliance with protocol steps that may be omitted in manual workflows. Importantly, a demonstration of the user interface to the Johns Hopkins Hospital Glucose Steering Committee was met with support to implement as a smart calculator with the belief that it would reduce calculation errors associated with poor usage of the current paper nomogram.

Limitations of the technical workflow include its specificity of the Johns Hopkins Nurse-Managed Insulin Infusion Protocol; generalizability to protocols not driven by lookup tables, including physician-managed protocols or those that rely on clinical judgment. Additionally, only a single pump and medical record were tested, and our ability to apply pump controls even in a laboratory setting was limited by the manufacturer's willingness to develop with us.

Conclusion - Using systems engineering methods and user-centered design principles, we automated the nurse-managed Johns Hopkins insulin infusion protocol in a test system and described the functional requirements that would be necessary for the system to function clinically. The Smart Agent system dynamically controls insulin infusions to reduce the time and workload burden on clinical staff, improve accuracy and efficiency of the protocol, and prevent calculation errors that can lead to patient harm. When compared to the current standard-of-care insulin infusion process, the Smart Agent system was more efficient, safer, less workload intensive, and perceived by nurses as more usable and similarly trustworthy. This study clearly illustrates the potential of this approach to jointly optimize safety, efficiency, and workload considerations in a way not possible without addressing system interoperability. Although this is a particular example we believe this approach can be replicated for a number of medications.

Project 3: Develop and implement an indicator of unit-level stress in an engineered care system to predict and mitigate risk

Abstract

Purpose: We postulate that unit, patient, and disease factors may contribute to the susceptibility of patients to preventable harms. We further hypothesize that this risk can be empirically derived and calculated as a Susceptibility Index. This index could then be used to minimize the risk of patient harms and achieve High-Reliability ICU status. The purpose of this study was to determine if data on patient conditions, unit environment, and best care practices can be aggregated and analyzed to create a prediction model for adverse outcomes in the ICU.
Scope: Involved were 510 surgical ICU patients admitted to a 12-bed mid-Atlantic academic hospital from 1 April to 30 September 2015.
Methods: This was a retrospective study in which data were drawn from EMR, research data, manpower data, and the error reporting system. The binary outcome measure was whether each patient experienced an adverse outcome (1=yes, 0=no). Hierarchical stepwise logistic regression was applied sequentially to analyze the influences of patient conditions, unit environment, and therapies received on ICU adverse outcome. This predictive model was validated using machine learning algorithms.

Results: Patients who were older and more ill faced increased susceptibility to ICU adverse events. Therapies that contradicted delirium best practices or mobility practices, or that used multiple or synchronous mode ventilators, also increased patient susceptibility to ICU adverse events. However, having more clinical technicians in the ward and the use of a proton pump inhibitor or tracheostomy placement reduced ICU patients' susceptibility to adverse events. Machine learning algorithms validated the predictive model given a limited set of data.
Key Words: age, DRG, staffing, mobility, delirium, ventilator, central line

Purpose

The susceptibility of ICU patients to harm may be a function of the dynamic relationships between several unit-level factors. For example, staffing levels and 'capacity strain' may erode safety culture and weaken compliance with best care practices. Leveraging data sources may unearth unit-level interactions that are not perceptible to a single ICU team member. We propose that integrating ICU patient, team, and environmental data from several disparate sources can predict a unit's susceptibility to propagate patient harm. Here, we focus on the methods and challenges of integrating data for this purpose.

Objectives:

1. Collect the patient, unit, and therapy-based factors that contribute to adverse patient outcomes in the ICU.
2. Curate the factors data for use in analysis.
3. Perform data analysis using regression and machine learning methods to demonstrate that unit-based factors can be used to create a meaningful index that predicts patients' susceptibility to preventable harm.

Scope

Medical errors continue to be a major cause of death and disability in the US. Estimates of mortality from preventable harm are in the range of 215,000 deaths/year.[17] Preventable harms have been largely addressed at a systems level by the creation of "care bundles" that incorporate best practices for prophylaxis of specific harms. Critical care units are a system of systems at the nexus of the sickest patients, the most complex equipment, and a highly trained workforce. In this system, patients who have critical, life-threatening conditions receive constant monitoring and therapy to prevent significant injuries and death. In the ICU, multidisciplinary teams have been established to improve and standardize the care of patients. However, it is estimated that there are still over 3.6 preventable harms for every 100 ICU patient days.[18] Many factors contribute to harms propagation. These factors include patient conditions, unit-level factors, and provider team factors. The standard approach to prophylactic care often fails because the system model assumes that all patients and situations are approximately "equal." In reality, we know that this is not true. For example, prescribing ambulation therapy to a patient who is physically unable to walk is destined to fail. To determine if there is a more comprehensive indicator of potential for patients to suffer harm, we hypothesized that unit-level factors and patient conditions may accurately predict risk of patient harm.

The Surgical ICU at Johns Hopkins Hospital was selected as the source for testing this hypothesis due to the existence of the Emerge dataset, a database comprising hundreds of patient-days' worth of process and outcomes data for five preventable ICU harms.

Methods

Unit Susceptibility Index Framework – We created a conceptual model using a causal loops framework to test the hypothesis that numerous, disparate events and conditions contribute to unit-level susceptibility to harm (Figure 6). This framework was based upon input from the medical literature, clinical experts, and systems engineers. Conceptual nodes within the causal loops model were linked to data sets from the Surgical ICU at Johns Hopkins Hospital.

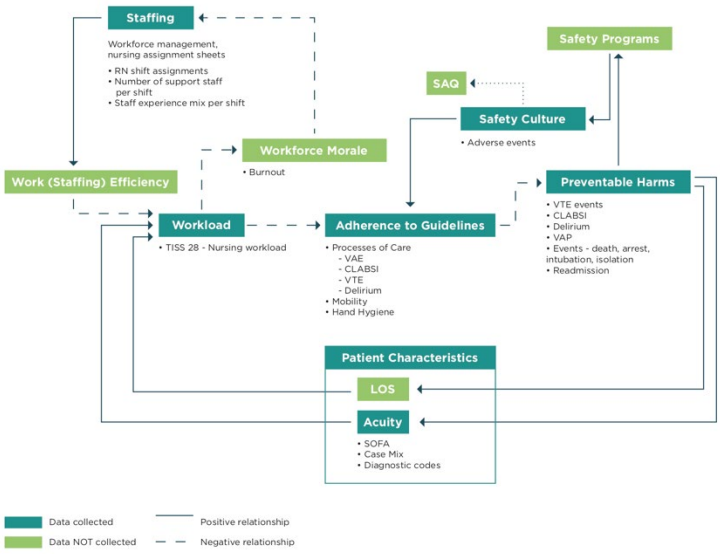


Figure 6. The Causal Loop Model for Susceptibility to Preventable Harm in the ICU. Postulated contributory data categories are illustrated in boxes. The single outcome variable is labeled "preventable harms" and includes the harms data that had been collected from the Emerge project as well as some standard ICU performance metrics. We postulated that elements that had a solid line would contribute positively to susceptibility while dotted lines would decrease susceptibility. This *a priori* assignment had no bearing on the analysis. Dark green elements are those that were collected while those in light green were not collected either because of availability or because they could not be incorporated with the time scale of this analysis.

The model demonstrates unavoidable levels of endogeneity, because we hypothesized that the occurrence of a preventable harm actually increased the susceptibility to further harmful events.

Data Aggregation - Data from 2434 patient days spanning 6 months were compiled, including 60 manually collected elements to calculate SOFA and TISS scores for acuity and workload, respectively. Unplanned events, such as cardiopulmonary arrests and intubations, were documented. Compliance with best care practices was captured via an EMR-integrated application. Nursing schedule and assignment data were integrated to assess nurse experience per shift, support staff available, sick calls, orientees, nurse-patient ratio, and planned and unplanned admissions. VTE events, CLABSI, VAP, and hand hygiene rates were extracted from the hospital's epidemiology and safety reporting systems.

Data Analytics - The outcome measure on ICU All-Cause Adverse Outcome was a binary variable, which was coded as '1' if a patient experienced at least one event pertaining to central line-associated blood stream infection (CLABSI), *Clostridium difficile (C.diff)*, ventilator-associated pneumonia (VAP), intubation, a delirium episode, an arrest, or a readmission to the ICU and as '0' otherwise.

There were three groups of predictor variables. The first group was related to the patient's presenting conditions, age, gender, race, and severity of illness (measured by diagnosis-related group (DRG) and relative weight).

The second group was related to the unit environment measured by a binary variable on whether the patient was admitted after 5 pm; seniority of, number of hours worked, and number of overtime hours worked by the primary, preceptor, and orientee nurses; and the number of clinical and non-clinical technicians who were working during the shift. The third group included best practice therapies applied to the patient. These were each coded as binary variables of '1' if the patient received a central line bundle, mobility therapy, mechanical and pharmacological venous thromboembolism (VTE) therapy, ventilation therapy, tilting the head of the bed more than 30 degrees, using tracheostomy, subglottic suction, oral care, in-line suction, spontaneous breathing trial, proton pump inhibitor, delirium therapy, or needed isolation for infection.

Hierarchical stepwise logistic regression was applied sequentially as follows. First, the patient's presenting conditions were regressed against the outcome variable, followed by the conditions in the unit environment. Next, the therapy choices were regressed against the outcome variable to determine the incremental contributions of each group of factors on ICU adverse outcomes. Finally, this predictive model was validated by machine learning algorithms.

Limitations - Six months of data from one hospital in one ICU limited the generalizability of the results to other hospitals and other types of ICUs. These limitations present opportunities to expand the study using more longitudinal data from other hospitals and in other types of ICUs.

Results

Principal Findings - The hierarchical stepwise logistic regression indicated that older patients with higher DRG relative weights were more susceptible to adverse outcomes in the surgical ICU. Six of the 18 specific therapies that were compiled in our database (Table 1) were associated either positively or negatively with the composite adverse outcomes. When patients did not receive care according to the delirium care bundle or, when multiple ventilator modes were employed, there was also increased susceptibility to adverse events in the ICU. In contrast, susceptibility to an adverse ICU event appeared to be reduced by an increased number of clinical technicians or the appropriate use of a proton pump inhibitor or a tracheostomy. Patient gender and race were not significantly associated with ICU adverse outcomes after unit conditions and therapy choices were considered. Perhaps surprisingly, we found that the application of multiple different modes of mechanical ventilation, the use of a synchronous mechanical ventilation mode, and the use of standard mobility protocols were associated with worse outcomes. With respect to mechanical ventilation, the use of multiple modes may simply be a marker of significant respiratory failure, or it may suggest that common, but unproved, therapies may be contraindicated more frequently or even injurious. As expected, there was a positive association between adverse outcomes and when delirium best practices were not applied.

More important for the objective of this study is the possibility that organizational factors may play a role in the risk propensity of an ICU environment. We included 11 measures of organizational design to ascertain erosion of the capacity of providers to monitor patient well-being and deliver timely interventions. We found that the number of clinical technicians is protective (reduces the probability) of all-cause adverse outcomes. Finally, our logistic model returned an area under the curve (AUC) statistic of 72.2%. Given the limited number of factors found to be statistically significant, this result indicates an excellent fit of the data to the theoretical model.

	Odds Ratio	Sig.	95% C.I. for Odds Ratio	
Variable			Lower	Upper
Hierarchy 1: Patient's Presenting Conditions				
Age	1.006	0.043	1.000	1.011
Male	1.096	0.303	0.920	1.306
White	0.908	0.285	0.761	1.084
DRG Relative Weight	1.083	0.000	1.065	1.100
Hierarchy 2: Unit Environment				
Number of Clinical Technicians	0.797	0.007	0.676	0.939
Hierarchy 3: Therapy Choices				
Exceeded Mobility Target	1.559	0.000	1.287	1.888
Did Not Apply Delirium Best Practice	2.704	0.000	2.113	3.461
Applied Multiple Ventilators	1.748	0.000	1.383	2.209
Applied Synchronous Ventilators	1.424	0.000	1.175	1.725
Applied Proton Pump Inhibitor	0.674	0.000	0.549	0.827
Applied Tracheostomy	0.115	0.000	0.064	0.205
Constant	0.124	0.000		

Table 1: Hierarchical Stepwise Logistic Regression for Risk of All-cause Adverse Outcome. For clarity non-significant variables were excluded from the table (but were included in the logistic regression model).

Having performed a proof of concept for the causal loops model using the regression, we applied machine learning (ML) techniques to analyze the data in a non-biased manner. Because our data sets were relatively sparse for ML methods, we used four ML-based prediction models that use algorithms that are relatively less data dependent. These were the Decision Tree Classifier, Random Forest Classifier (RF), Support Vector Machine (SVM), and K Nearest Neighbor (KNN) classifier. After optimization with validation curves and grid-search, the algorithms with the optimal parameters were trained with all observations in the training set (approximately 80% of the entire data set). Subsequently, the optimized-and-trained model was used to predict the independent variables in the test set. The RF, SVM, and KNN all had very good area under the receiver operator curves (AUROCs) (Fig. 7). In addition, these ML predictive models serve to indirectly validate the explanatory regression model.

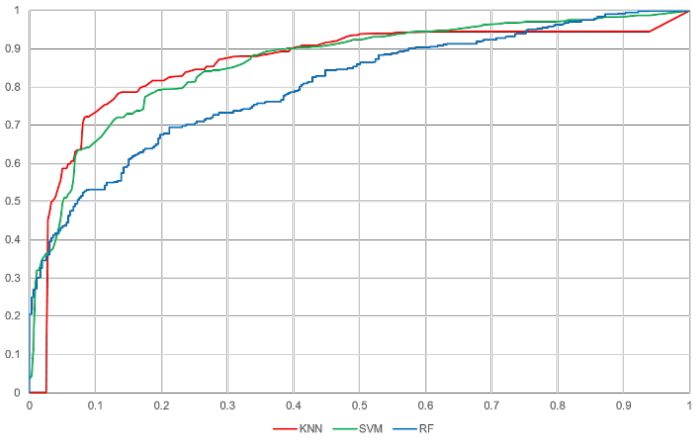


Figure 7. Receiver Operator Curves for Machine Learning prediction of all-cause harms. The SICU databases were evaluated using optimized K-Nearest Neighbor (KNN, in red), Support Vector Machine (SVM, in green), or Random Forest (RF, in blue). The AUC for each algorithm is given in the text.

Table 2 is a comparison of the logistic regression with all of the ML analyses. The results demonstrate the superiority of the ML techniques.

Model	AUROC	Precision	Recall	F1 Score	BACC	MCC	10x CV AUROC Mean	10x CV AUROC StDev
Logistic Regression	0.7578	0.8393	0.2327	0.3643	0.5913	0.2574	0.7864	0.064
Decision Tree Classifier	0.7062	0.8861	0.4431	0.5908	0.6895	0.4288	0.7939	0.0623
Random Forest Classifier	0.8157	0.8375	0.5743	0.6814	0.7245	0.4659	0.8544	0.0399
Support Vector Machine (RBF Kernel)	0.8619	0.8381	0.7302	0.7804	0.7857	0.5721	0.8078	0.0392
K-Nearest Neighbors Classifier (K=5)	0.8763	0.8435	0.7871	0.8143	0.8114	0.6217	0.8556	0.0273

Table 2: Comparison of Models to Predict Susceptibility of ICU Patients to Preventable Harm

In particular, two algorithms, the KNN and SVM, showed relatively high F1 scores, which indicated high levels of precision and recall in predictions. The results of the prediction models strongly validate the hypothesized model of effects (as represented in the regression models).

Discussion - This study is the first application of statistical and machine learning techniques to develop a prediction model for the susceptibility of patients in an ICU to preventable harms. It is also unique because it simultaneously considers both organizational and clinical factors in determining the reasons for all-cause adverse outcomes in the SICU. ICU clinicians, administrators, and even patients believe that they know the factors that compromise patient care and contribute to avoidable harm. For example, the perception that nurse staffing and patient acuity contribute to harm have motivated mandatory staffing ratios.[19] Though perceptions are strong in many areas, the applicability of the evidence to critically ill patients may be questioned.[20] In this study, we did not find evidence that organizational design or nurse experience and staffing was associated with adverse outcomes. This may be an example of a case where misperceptions have long misguided unit functions. However, we did find that the number of clinical technicians provided a protective effect. It is not surprising that clinical technicians may improve care processes because, in our system, their major role is precisely for care delivery.

Although this pilot study is limited in scope to the surgical ICU and over time to a 6-month period, the findings suggest intriguing directions for future research. The findings of the logistical regression support some activities while calling others into question. Although this may be surprising at first, it should be remembered that many elements of best care practice bundles have been established from studies using very small numbers of patients and low effect sizes. Our techniques may become an important tool in improving the effectiveness of best care practice bundles over time and creating a learning healthcare system. Despite the limited data set, with relatively lower levels of observed variance and the restricted access to patient data, the results of the ML prediction models are very promising. These models show that there is considerable accuracy in predicting the Target Class based on the feature set. The hypothesized variables when transformed into a feature used for prediction deliver significant levels of accuracy in predicting adverse outcomes. This indicates that out project goal of creating a real-time susceptibility index can be achieved with improved automation of data collection and cloud-based computing. Moreover, such a susceptibility index could be used to prompt meaningful interventions to lower the susceptibility of patients in the ICU to harm.

As with all retrospective studies that attempt to use multiple sources of data to test a multifactorial model, ours encountered challenges in data collection and curation. The collection of EMR data was technically straightforward, but we found that useful data were also contained in provider notes and are not structured for easy extraction.

Clinical and operational systems remain disjointed, making comparison of unit- and hospital-level variables with patient outcomes challenging. Because we collected data from two sources (nurse shift rotation schedules and profiles and patient EMR), matching the sets using manual matching and programming required significant project resources. We had to interpret, classify, and code notes and conduct-associated reliability checks to minimize error and to ensure that the data were comparable across patients, time periods, and providers. This kind of data management is not compatible with an automated susceptibility index. Our experience suggests that an idealized system architecture would require integrated data systems in order to operationalize a real-time predictive analytics system like the Susceptibility Index.

We do not believe that the findings were severely biased by the structural limitations of the study, but they clearly limit the power of the test. That is, we may have found more determinants of unit-level safety with a larger sample consisting of a longer time period and across more ICUs. In contrast, we may have discovered some factors that will not be significant when exposed to a broader patient population. To address these limitations, in future studies, we plan to collect data over a longer time period and across a larger set of ICU types. In addition, we believe that physician-related rotation and demographic variables may play a role, because many of the therapies were ordered by physicians, often based on judgment.

Conclusions - In this first-of-its-kind pilot study, we attempted to determine the organizational and clinical factors associated with all-cause adverse outcomes in a surgical ICU. We found that both types of factors were important. Limitations in sample size and time period prevent us from making statements of causality, but our results suggest that by expanding the sample size and lengthening the time period, we are likely to achieve more robust findings that could guide staffing policies as well as personalize protocol-driven interventions in the ICU. The stepwise hierarchical logistic regression analysis identified therapies and conditions that appear to either prevent, contribute to, or have no apparent effect on the harms outcomes measured. This analysis begins to shed new light on the relative importance of elements within and outside of care bundles. The calibration of this analysis is only fair and is limited by a number of factors, such as a missing data and the inclusion of only a single ICU. We next applied machine learning classification algorithms to overcome limitations of regression analysis and of our data sets. The machine learning analysis of the data predicted the likelihood of an adverse event with greater than 80% accuracy in each of three algorithms. Together, these results indicate that data visualization can improve performance while more expansive data collection and advanced analysis can improve risk prediction and better define contributors to harm.

List of Publications and Products

Publications:

1. Griffiths SM, Sapirstein A, Guzman JC, Soriano Z, Ravitz AD. Automated, Web-Based Solution for Bidirectional EHR-Infusion Pump Communication. *Biomed Instrum Technol*. 2019 Jan./Feb.;53 (1):30-37.
2. Barasch N, Romig M, Demko Z, Dwyer C, Dietz A, Rosen M, Griffiths S, Ravitz A, Pronovost P, and Sapirstein A. Interoperable integration of a nurse-managed insulin infusion protocol as a model to improve safety and efficiency in the delivery of high risk medications (submitted). *Journal of Patient Safety & Risk Management*. 2019 Jun.

3. Mathews S, Stoll R, Sternberger W, Cox P, Tober T, Mattina J, Dwyer C, Barasch, N Carolan H, Romig M, Pronovost P, Barnes J, Ravitz A, Sapirstein A. Prioritizing Healthcare Solutions using the Quality Function Deployment Process. *American Journal of Medical Quality* (submitted). 2019 Jan.

4. Rosen MA, Demko Z, Barasch N, Dwyer C, Ravitz A, Pronovost P, Sapirstein A. A Smart Agent System for Insulin Infusion Protocol Management Increases Safety & Efficiency while Reducing Workload: A Simulation-based Human Factors Evaluation Study (in preparation). 2019 Jun.

Presentations:

1. Mathews S, Stoll R, Sternberger W, Cox P, Tober T, Mattina J, Dwyer C, Barasch, N Carolan H, Romig M, Pronovost P, Barnes J, Ravitz A, Sapirstein A. Prioritizing Healthcare Solutions using the Quality Function Deployment Process. *Annual Congress of the Society for Critical Care Medicine*, San Diego, CA. 2019 Feb 21.
2. Demko Z, Barasch N, Dwyer C, Barnes J, Griffiths S, Ravitz A, Romig M, Pronovost P, Sapirstein A, Rosen MA. Engineering Interoperability Between EHRs and Infusion Pumps: A Human Factors Evaluation Study. *Annual Congress of the Society for Critical Care Medicine*, San Diego, CA. 2019 Feb 21.
3. Dwyer C, Carolan H, Demko Z, Ravitz A, Pronovost P, Sapirstein A. Managing Data Requirements to Measure ICU Susceptibility to Patient Harm. *Annual Congress of the Society for Critical Care Medicine*, San Diego, CA. 2019 Feb 21.
4. Lee S, Dwyer C, Carolan H, Demko Z, Aron R, Phan P, Ravitz A, Pronovost P, Sapirstein A. The Impact of Patient, Ward, and Best Practice Therapies on Adverse Outcomes in the ICU. *Annual Congress of the Society for Critical Care Medicine*, San Diego, CA. 2019 Feb 21.
5. Dwyer C. Managing data requirements to measure ICU susceptibility to patient harm. *Johns Hopkins Medicine Patient Safety Summit*. 2019 Mar 8.
6. Aron R. Person, organization, and therapeutic causes of ICU adverse outcomes. *Johns Hopkins Medicine Patient Safety Summit*. 2019 Mar 8.
7. Demko Z. Engineering interoperability between EHRs and infusion pumps: A human factors evaluation study. Johns Hopkins Medicine Patient Safety Summit. 2019 Mar 8.
8. Barasch N. Patient safety: Engineering from the top and the bottom. Johns Hopkins Medicine Patient Safety Summit. 2019 Mar 8.

References

1. https://www.healthit.gov/topic/laws-regulation-and-policy/health-it-legislation, last accessed June 20, 2019.
2. Stevens J. The how and why of open architecture. Undersea warfare 2008 Spring (Available: http://www.navy.mil/navydata/cno/n87/usw/spring08/HowAndWhy.html).

3. https://www.ahrq.gov/professionals/systems/hospital/pressureulcertoolkit/putool1.html, last accessed May 11, 2018.
4. https://www.ahrq.gov/professionals/systems/hospital/pressureulcertoolkit/putool1.html, last accessed May 11, 2018.
5. Leaf Healthcare, Inc. (2014). The Financial Impact of Pressure Ulcers (White paper). Retrieved from http://leafhealthcare.com/pdfs/LH_WP_FinancialOverview_1563AB_101316.pdf. Link last accessed on May 11, 2018.
6. Stevens J. The how and why of open architecture. Undersea warfare 2008 Spring (available: http://www.navy.mil/navydata/cno/n87/usw/spring08/HowAndWhy.html).
7. Design Thinking for Educators, 2nd edition. IDEO. Retrieved from https://designthinkingforeducators.com/. Link last accessed on June 1, 2018.
8. Georgiadis, D.G., et al. (2012). "Using multi criteria decision making in analysis of alternatives for selection of enabling technology". Systems Engineering, volume16, issue3, pp. 287-303, https://doi.org/10.1002/sys.21233
9. Leape LL, Bates DW, Cullen DJ, et al. Systems analysis of adverse drug events. ADE Prevention Study Group. *JAMA*. 1995; 274: 35-43.
10. Medicine Io. *To Err Is Human: Building a Safer Health System*. Washington, DC: The National Academies Press, 2000, p.312.
11. Bally L, Thabit H and Hovorka R. Glucose-responsive insulin delivery for type 1 diabetes: The artificial pancreas story. *Int J Pharm*. 2018; 544: 309-18.
12. Latif A, Rawat N, Pustavoitau A, Pronovost PJ and Pham JC. National study on the distribution, causes, and consequences of voluntarily reported medication errors between the ICU and non-ICU settings. *Crit Care Med*. 2013; 41: 389-98.
13. Hewitt T, Chreim S and Forster A. Double checking: a second look. *J Eval Clin Pract*. 2016; 22: 267-74.
14. Hart SG, Staveland LE. Development of NASA-TLX (task load index): Results of empirical and theoretical research. In: Hancock PA, Meshkati N, eds. *Advances in psychology.* Vol 52. North-Holland; 1988:139-183.
15. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*. 2008;24(6):574-594. http://www.tandfonline.com/doi/abs/10.1080/10447310802205776. doi: 10.1080/10447310802205776.
16. Jian J, Bisantz AM, Drury CG. Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*. 2000;4(1):53-71.
17. Makary MA and Daniel M. Medical error-the third leading cause of death in the US. *BMJ*. 2016; 353: i2139.
18. Rothschild JM, Landrigan CP, Cronin JW, et al. The Critical Care Safety Study: The incidence and nature of adverse events and serious medical errors in intensive care. *Crit Care Med*. 2005; 33: 1694-700.
19. Needleman J, Buerhaus P, Pankratz VS, Leibson CL, Stevens SR and Harris M. Nurse staffing and inpatient hospital mortality. *N Engl J Med*. 2011; 364: 1037-45.
20. Law AC, Stevens JP, Hohmann S and Walkey AJ. Patient Outcomes After the Introduction of Statewide ICU Nurse Staffing Regulations. *Crit Care Med*. 2018; 46: 1563-9.