



OPEN ACCESS

Smart agent system for insulin infusion protocol management: a simulation-based human factors evaluation study

Michael A Rosen ^{1,2}, Mark Romig,³ Zoe Demko,¹ Noah Barasch,¹ Cynthia Dwyer,¹ Peter J Pronovost ^{4,5}, Adam Sapirstein³

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjqs-2020-011420>).

For numbered affiliations see end of article.

Correspondence to

Dr Michael A Rosen, Armstrong Institute for Patient Safety and Quality, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; mrosen44@jhmi.edu

Received 27 April 2020

Revised 9 February 2021

Accepted 18 February 2021

Published Online First

10 March 2021

ABSTRACT

Objective To compare the insulin infusion management of critically ill patients by nurses using either a common standard (ie, human completion of insulin infusion protocol steps) or smart agent (SA) system that integrates the electronic health record and infusion pump and automates insulin dose selection.

Design A within subjects design where participants completed 12 simulation scenarios, in 4 blocks of 3 scenarios each. Each block was performed with either the manual standard or the SA system. The initial starting condition was randomised to manual standard or SA and alternated thereafter.

Setting A simulation-based human factors evaluation conducted at a large academic medical centre.

Subjects Twenty critical care nurses.

Interventions A systems engineering intervention, the SA, for insulin infusion management.

Measurements The primary study outcomes were error rates and task completion times. Secondary study outcomes were perceived workload, trust in automation and system usability, all measured with previously validated scales.

Main results The SA system produced significantly fewer dose errors compared with manual calculation (17% (n=20) vs 0, p<0.001). Participants were significantly faster, completing the protocol using the SA system (p<0.001). Overall ratings of workload for the SA system were significantly lower than with the manual system (p<0.001). For trust ratings, there was a significant interaction between time (first or second exposure) and the system used, such that after their second exposure to the two systems, participants had significantly more trust in the SA system. Participants rated the usability of the SA system significantly higher than the manual system (p<0.001).

Conclusions A systems engineering approach jointly optimised safety, efficiency and workload considerations.

INTRODUCTION

Glucose management improves outcomes for critically ill patients.¹ Current management practices entail bedside insulin infusion protocols,^{2 3} but protocols are work intensive for nurses⁴ and error prone. To complete a protocol, nurses must retrieve

multiple data points from one or multiple systems, make manual calculations to determine any dose rate change and manually enter that data into a different system or device. In the intensive care unit (ICU), insulin administration errors are common,⁵ reaching up to 185.9 events per 1000 patient days of insulin treatment.⁶ One safety recommendation is the independent human double check.⁷ However, this check does not guarantee errors will be caught and corrected^{8 9} and introduces further work for nurses.¹⁰ Compliance with the nurse double check is also low.¹⁰

To help manage blood glucose (BG), our hospital uses a modified version of the Yale protocol.¹¹ Once initiated, the process requires a nurse to retrieve the patient's current and previous BG levels from the electronic health record (EHR) and manually calculate an hourly rate change. Instructions are then found in a lookup table using the current BG level and rate of change. If an increase or decrease of infusion rate is recommended, the current infusion rate and a multiplier found in the first lookup table are used in a second lookup table to find the change in infusion rate. This change is then added or subtracted from the current infusion rate to determine new pump settings. A second nurse double checks this process before it is documented and changes made to the pump. This process requires manual retrieval and manipulation of discrete pieces of information and consequently is susceptible to errors arising from limitations of human memory and attention. The objective of this study was to compare insulin infusion management of critically ill patients by



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Rosen MA, Romig M, Demko Z, et al. *BMJ Qual Saf* 2021;**30**:893–900.

nurses using either the current manual protocol system (manual system) or a smart agent (SA) system. The SA integrates the EHR and infusion pump and automates data retrieval, calculation and insulin dose selection. The SA was developed using a systems engineering approach.¹² We sought to explore the degree to which the SA system could improve safety and efficiency of insulin infusion practices.

METHODS

Study design

This mixed methods study compared the manual and SA insulin management systems in a simulation-based, human factors evaluation. We used a within subjects design to evaluate the two systems across 12 simulation scenarios completed in blocks of 3. Participants were randomised to an initial system exposure and alternated the system used (manual or SA) in each block thereafter. Surveys quantitatively assessed workload and user trust after each block of scenarios, and usability once at the end of the session resulting in a 2 (manual vs SA system) by 4 (scenario block) design for workload and trust, and a 2 (manual vs SA system) group comparison for usability. Similarly, a two group (manual versus SA system) comparison was used for error rates. Efficiency was measured at the scenario level, resulting in a 2 (manual versus SA) by 12 (scenario) design. We chose this design to increase study power, by collecting repeated measures data, and to decrease the impact of sequence effects of first exposure to the same system. A qualitative debrief captured participant experiences and perceived barriers to use. The study design and flow of evaluation sessions is illustrated in the online supplemental material. The Johns Hopkins School of Medicine Institutional Review Board approved the study with informed consent required; data collection occurred May through July 2018.

Participants and recruitment

All critical care nurses working in an ICU or step down unit within the Department of Surgery at an urban academic institution who had used the manual insulin infusion protocol were eligible for inclusion in the study (approximately 200 nurses at the time of recruitment). We recruited nurses through email notifications and flyers posted on the inpatient units. The emails and flyers indicated that the project aimed to evaluate the potential safety and efficiency of continuous drug infusions using automated dose adjustments and integration of the medication infusion pump with the medical record. Participants had no prior exposure to the SA system. Nurses volunteered and were compensated based on the high needs pay rate for critical care nurses (\$50 per hour).

Intervention procedure

Consented nurses completed a brief background survey requesting professional history, unit type and

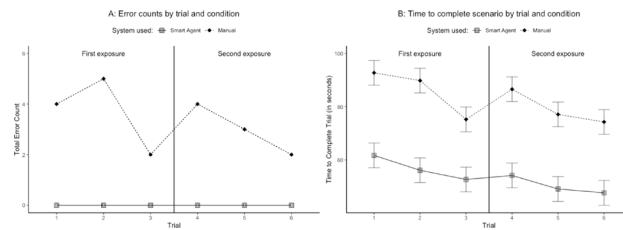


Figure 1 The primary study outcomes of safety and efficiency compared by manual system versus smart agent system. Panel A is the total error counts across nurses by exposure (safety measure). Panel B is the average time to complete a dose change by exposure. A trial is one of the six scenarios completed using one of the two systems (six trials of each of the two systems).

surgical specialty and experience managing patients on insulin infusion protocols. A coin toss randomised the nurse to their initial system exposure. A simulation session comprised four blocks (three scenario evaluations per block) and nurses alternated systems by block, completing six scenarios per system (figure 1). A session facilitator (a research coordinator, ZD or NB; or a research nurse, CD) oriented the nurse participant to the simulation session. This included the Epic EHR test environment with simulated patient records, the Hospira 360 medication infusion pump, paper protocol for the manual system and the SA system. The pump used in this study was not in clinical use and not linked to drug libraries or dose error reduction software. The facilitator demonstrated completion of a training scenario with the manual protocol system and the same scenario with the SA system. Nurses were encouraged to ask questions throughout the orientation and training.

Simulation scenarios

Each scenario contained the required EHR values (current BG level, past BG level and current infusion rate) for the insulin infusion protocol. Nurses were given a name associated with a patient record in the EHR test environment for each scenario. For the manual protocol system, nurses searched the EHR for the three values, used the protocol to calculate the correct insulin dose change and adjusted the pump settings as needed. For the SA workflow, nurses started the SA in the EHR, selected the glucose value for calculations, viewed the dose calculation result, had the ability to accept or modify the dose and confirmed the rate at the infusion pump. Technical and clinical process details of the SA system have been previously published.^{12 13} The online supplemental appendix includes full information for each scenario. The order of scenarios remained constant across participants. We only included scenarios within the range of the protocol that would not require pausing or stopping the infusion or require going off the protocol (ie, the hyperglycaemic and hypoglycaemic ranges were avoided). We excluded the nurse double check from this study because of the added logistic complexity (eg,

recruiting additional participants or a confederate) and our belief that this would not vary by system. An observer (ZD) used a custom android tablet application to time completion of each scenario and logged the final rate change entered by the nurse.

Surveys and debriefing

We administered three surveys during the simulation session to elicit nurse perceptions of each system. The workload and trust surveys were repeated after each block of scenarios and the usability scale was done after the session for each system. We randomised the order of system usability ratings to minimise any sequence effects in ratings. The facilitator engaged each nurse in a structured debriefing after the session. Surveys and the debriefing protocol are included in the online supplemental appendix.

Study outcomes

The primary study outcomes were error rates (safety) and task completion times (efficiency). Secondary outcomes were perceived workload,¹⁴ trust in automation¹⁵ and system usability,¹⁶ all measured with previously validated scales.

Measures and data sources

Safety was measured as the rate of errors. Specifically, an error was defined as an infusion rate entered into and confirmed at the pump that differed from the infusion rate recommended by the protocol, regardless of the magnitude of deviation. Errors were captured dichotomously for each scenario (ie, an error did or did not occur) by comparing the insulin dose entered into the infusion pump by the nurse to the correct dose rate. This approach did not allow differentiation between errors made during the calculation of changes in glucose, calculations of infusion rates or slips (eg, keystroke entry error) or mistakes (eg, memory failures after calculation and before pump entry). Much of the calculation work is done internally and not directly observable. We did not prompt for intermediate calculation results to avoid altering the participants' work processes and efficiency.

Efficiency was measured as time (seconds) to complete each scenario, starting when the nurse was given the patient name and ending when pump programming was complete. We also timed specific tasks: information retrieval in EHR, performing manual dose calculation (eg, on paper, using calculator), performing dose calculation using EHR and changing infusion pump settings.

Workload was measured using the NASA Task Load Index (TLX).¹⁴ The index involves two steps: (1) participants rate six distinct dimensions of workload on a 21-point scale (higher numbers indicate more workload) and (2) participants complete pairwise comparisons of the six dimensions, indicating which workload dimension was more important to their

overall performance. Nurses completed step 1 after each block of scenarios and step 2 at the end of the session. The pairwise comparisons were used to create a total weighted workload score, ranging from 0 to 100 (higher numbers indicated higher levels of workload).

User trust and usability. Trust in the system was measured using a previously validated 12-item instrument on human trust in automation.¹⁵ An 8-point response scale (0=not at all, to 7=extremely) measures level of agreement with statements. Five negatively worded items are reverse scored. A sum of the item scores (potential range, 0–96, higher scores indicated higher levels of trust) was used in analysis. The Systems Usability Scale is a 10-item, widely used survey in research and commercial product development.¹⁶ A 5-point response scale (0=strongly disagree, to 4=strongly agree) measures agreement with statements. Five negatively worded items are reverse scored. We counted item scores of 3 or 4 as a *positive response* indicating high levels of perceived usability. A composite score was calculated as the percentage of positive responses across the ten items (potential range, 1–100) and was used in analysis.

Qualitative debriefing comments. Data were collected during a postsession debriefing with each nurse. The debriefing was semistructured and asked about the utility of the SA system, and perceptions of ease of use, efficiency and safety relative to the manual system. Open-ended questions elicited nurse opinions of the SA system, barriers to use and uptake and suggestions for improvement. Debrief sessions were not audio recorded, but notes were taken using a structured data collection form. Data were analysed and presented by first categorising specific statements by the focus of the comment (either technology-focused or workflow-focused) and subsequently by their valence (positive, neutral or negative sentiments), perceived barriers to use or suggestions for improvements. This structure for reporting comments aligns closely with specific questions asked during debriefing.

Analysis

All statistical tests for previously described study designs were conducted in R (V3.5.1; R Foundation for Statistical Computing, Vienna, Austria). Using the G*Power (V3.1.9.4) programme,¹⁷ we performed an a priori power analysis using a large (Cohen's *f* of 0.40) effect size heuristic method. There were no prior studies available to provide an effect size estimate for this type of intervention. Based on informal piloting of the SA and practical constraints for recruitment, a large effect size was chosen for sample size calculations. The χ^2 test for independence tested for differences in error rates between the SA and manual system conditions. Power analysis for the two by four analysis of variance for workload and trust included study design parameters (two levels of insulin infusion systems and four measurement periods) and a Bonferroni correction

for four main tests to maintain a familywise alpha of 0.05 (an alpha of 0.0125 for each test). Sample size requirements were 20 nurses (for an actual power of 0.96) for a large effect size. A paired samples t-test was conducted to determine differences in perceived usability between the manual and SA systems. An a priori power analysis using a large (Cohen's d_z of 0.8) effect size estimate yielded a required sample size of 19 for an actual power of 0.96.

RESULTS

Twenty critical care nurses completed the study. Professional experience ranged from <1 year to 37 years of nursing (mean (SD) of 5.03 (8.85)) and from <1 year to 35 years of critical care nursing (mean 4.48 (8.60)). Fifty per cent (n=10) of nurses reported once per month use of the manual insulin infusion protocol, 35% (n=7) reported once per week and 15% (n=3) more than once per week.

Error rates

The insulin dose accuracy calculation of the SA was significantly better than a single manual calculation ($\chi^2=19.70$, $df=1$, $p<0.001$). In 120 scenarios, nurses made zero calculation errors using the SA system and made 20 calculation errors (16.6%) using the manual system. [Figure 1A](#) compares the total errors for SA and manual systems by exposure.

Description of error types

Observed errors were classified in terms of how they differed from recommended pump settings (ie, changes that differ in magnitude or direction, unnecessary changes or omissions of recommended changes). [Table 1](#) describes the wide range of error types observed with the manual system. Of the 20 total errors, 30%⁶ were an incorrect change in magnitude, 30%⁶ were an incorrect change in direction, 25%⁵ were failures to make a change when one was required, 10%² were inappropriate stopping or pauses of the protocol and 5%¹ were unnecessary changes (ie, pump adjustments made when the protocol recommended none). In eight scenarios, infusions were set too high (with a range of discrepancies from 0.5 to 2 units of insulin/hour) and too low in 12 scenarios (ranging from -4 to -0.5 units of insulin/hour).

Efficiency

Nurses were significantly faster, completing the protocol using the SA system (mean 53.57 (20.30) s per scenario) than when using the manual system (mean 82.62 (24.71)), $F(1, 19) = 68.45$, $p<0.001$ ([figure 1B](#)). The SA system achieved protocol finalisation, on average, 29.0s faster than the manual system. Tasks consuming the most time were retrieving information in the EHR and performing manual calculations ([figure 2](#)). Compared with the manual system,

Table 1 Descriptive information for each observed error (units of insulin per hour)

Error type	Scenario	Rate entered*	Rate change made†	Correct rate‡	Correct rate change§	Rate discrepancy¶
Incorrect magnitude of change	4	2	-1.5	2.5	-1	-0.5
	5	3	0.5	3.5	1	-0.5
	5	3	0.5	3.5	1	-0.5
	6	4	1	5	2	-1
	8	3	1	2.5	0.5	0.5
	9	6	2	5	1	1
Incorrect direction of change	2	2	-2	6	2	-4
	8	1.5	-0.5	2.5	0.5	-1
	9	3	-1	5	1	-2
	10	4.5	1	2.5	-1	2
	10	4.5	1	2.5	-1	2
	11	0.5	-0.5	1.5	0.5	-1
Omission of required change	4	3.5	0	2.5	-1	1
	4	3.5	0	2.5	-1	1
	6	3	0	5	2	-2
	7	4.5	0	3.5	-1	1
	10	3.5	0	2.5	-1	1
Inappropriate stop or pause	5	0	-2.5	3.5	1	-3.5
	5	0	-2.5	3.5	1	-3.5
Unnecessary change	1	0.5	-0.5	1	0	-0.5

*The final rate accepted at the pump by participant.

†The increase (positive numbers) or decrease (negative numbers) in rate from the rate settings at the start of the scenario.

‡The infusion rate recommended by the protocol.

§The correct increase (positive numbers) or decrease (negative numbers) in rate recommended by the protocol.

¶The difference in rate entered and rate recommended by the protocol.

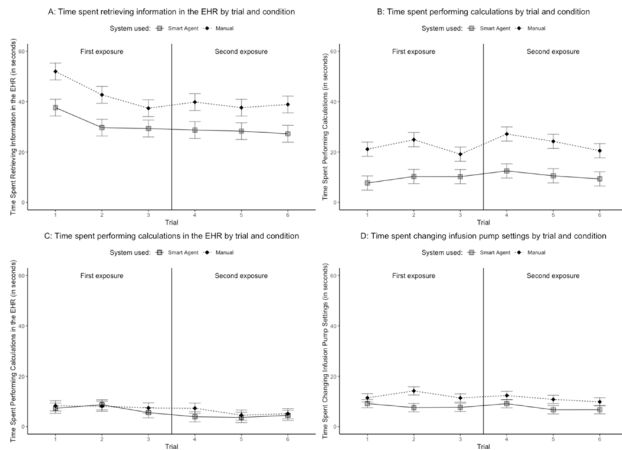


Figure 2 The time spent completing four (panels A through D) specific tasks involved in an insulin infusion rate change by trial and manual system versus smart agent system. A trial is one of the six scenarios completed using one of the two systems (six trials of each of the two systems). EHR, electronic health record.

the SA system performed significantly better for both of these tasks. Additionally, the SA system outperformed the manual system for changing infusion pump settings, but not for calculations performed within the EHR.

Workload

Nurses rated overall workload of the SA system (mean 6.33 (2.38)) significantly lower than the manual system (mean 10.23 (2.74)); $F(1, 19) = 37.32, p < 0.001$ (figure 3A).

User trust and usability

Trust ratings were similar between the systems after the first exposure. However, trust in the SA system increased significantly when compared with the standard system for the second exposure $F(1, 19) = 8.41, p = 0.009$ (figure 3B). Nurses rated usability of the SA system (mean 86.88 (10.73)) significantly higher than the manual system (mean 52.00 (17.97)); $t(19) = -6.23, p < 0.001$ (figure 3C).

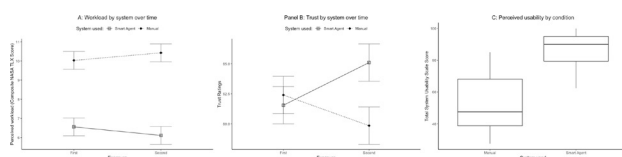


Figure 3 Nurses' self-reported ratings of workload, trust and usability compared by manual system and smart agent system. (A) Workload was measured after each block of three scenarios using the NASA-TLX scale composite score and compared by exposure (first or second) to each system. (B) Trust was measured after each block using the trust in automation scale total score and also compared by exposure. (C) Perceived usability was measured once at the end of the session using the System Usability Scale total score.

Qualitative debrief comments

In general, study participants had positive remarks about the SA system. All 20 nurses thought the SA would be helpful and found it more efficient than the manual system, and 18 (90%) found it easier to use. Fifteen (75%) nurses believed the SA system was safer than the manual system, while 5 (25%) were unsure or believed both systems were about as safe. These results are consistent with those obtained from the workload, usability and trust surveys.

Table 2 summarises the qualitative comments on the SA system. Perceived barriers to use included interruption or connectivity failures to the SA and other electronic tools, and differences in some nurse's level of comfort with and willingness to use the system. Suggestions for improvement mainly focused on the SA interface (improving access and reorganising the display) and optimising its connectivity with other components of the insulin infusion workflow.

Differences in outcomes by protocol use frequency

Descriptive data for study outcomes by self-reported frequency of protocol use are provided in the online supplemental appendix. Outcomes are generally consistent across groups of more and less frequent protocol users with three potential exceptions. First, there appears to be a trend for higher error rates for participants with more frequent protocol use (0.22 for participants reporting protocol use more than once per week; 0.15 for those reporting use once per month). Second, the largest differences in perceptions of usability between the SA and manual systems occur for very high (more than once per week) and very low (once per month) protocol users. Third, frequent protocol users (more than once per week) report the highest levels of trust in the SA system and the largest difference in trust between SA and manual systems.

DISCUSSION

This study demonstrated the potential of automated tools enabled by systems engineering and integration to reduce human error in the administration of insulin by protocol, in a critical care setting. When compared in a simulated work environment to the current manual system, the SA system was safer, more efficient, less workload intensive and perceived by nurses as more usable and similarly trustworthy. Most striking, the SA system eliminated errors in insulin dosing changes while concurrently reducing protocol completion time by 35.2%, or an average of 29 s per scenario.

We excluded a double check in this study, which would presumably reduce human errors and also increase task completion time. However, nurse double checks in a busy ICU are challenging to coordinate, time consuming and may have poor integrity.^{18 19} Our finding of zero errors with the SA system suggests that clinical implementation of such a system could obviate the need for a human double check. Thus,

Table 2 SA debrief comments

Comment category	Technology	Workflow
Positive comments	<ul style="list-style-type: none"> ▶ Improved speed and efficiency of SA compared with manual system ▶ User-friendly and convenient ▶ Safe, trustworthy and accurate ▶ Appreciated ability to double check SA or use it to double check nurses' work 	<ul style="list-style-type: none"> ▶ Belief that trust and comfort with SA would increase with more time and use ▶ SA helped to avoid the major barrier of finding a second nurse for a double check ▶ SA would be useful given ICU business and high workload
Neutral comments	<ul style="list-style-type: none"> ▶ SA similarly easy to use as manual system ▶ SA similarly safe as manual system 	<ul style="list-style-type: none"> ▶ Safety, efficiency and utility of SA would depend on interuser variability in trust and workload
Perceived barriers to use	<ul style="list-style-type: none"> ▶ Concern about SA's ability to integrate proper data and correctly run algorithm ▶ Concern about pump-SA interoperability ▶ EMR downtime, Wi-Fi connectivity and glucometer syncing issues ▶ Confusing display 	<ul style="list-style-type: none"> ▶ Discomfort with lack of a second nurse ▶ Uneasiness with new technology and electronic protocols in general ▶ Concern that SA would induce laziness among clinicians, loss of skills or lack of awareness of patient conditions and trends ▶ Necessary for nurses to double check SA
Suggestions for improvement	<ul style="list-style-type: none"> ▶ Display of blood glucose values linearly from oldest to newest ▶ Display data and protocol side-by-side ▶ Occupy larger portion of the screen ▶ Decrease lag time from pump to EMR ▶ Create an SA tab in EMR 	<ul style="list-style-type: none"> ▶ Link to SA in the MAR and redirection to the MAR after dose confirmation in SA ▶ Link to SA in the flowsheets ▶ Generate alert when dose check is required

EMR, electronic medical record; ICU, intensive care unit; MAR, medical administration record; SA, smart agent.

SA implementation in a new workflow could provide workload relief and efficiency gains without compromising safety. We also found that nursing staff had greater trust after using the SA system and more positive perceptions of system usability. Trust and ease of use are important considerations in the adoption and use of technology.^{20 21}

The current SA system is limited to insulin infusion management, but the approach employed here could be expanded to other medication administration tasks where a structured protocol exists. Heparin infusion is another example in which dose adjustments are often made by a protocol that is based on laboratory results.²² As more automation tools are developed, further design and evaluation work will be required to understand how to integrate these agents effectively to support nurse monitoring and management of the automated agents. For complex patients on multiple infusions, the design of interfaces for nurses will become more challenging, but also the interactions between multiple algorithms driving the administration of medications may become more challenging to predict. Complex (and potentially risky) behaviour emerges from systems whose component agents follow simple rules.²³ Therefore, caution is warranted and rigorous software, system and human performance evaluation studies will be critical.

While reactions to the SA system were generally positive, participants did surface issues to address before this or similar systems integration based automation agents could be fielded effectively. Some potential barriers were common dispositional factors influencing technology adoption (ie, a general dislike or distrust of new technology), but many focused on the interplay of technology and workflow. Specifically, there was discomfort among some nurses with fully removing the second nurse from the double checking

process (ie, a workflow change) particularly as there were a set of related concerns about the reliability of the ecosystem of devices (ie, smart pump, glucometer) and systems (ie, EHR, hospital network) required to function properly for the SA system to work. These are valid concerns that would require risk mitigation (eg, redundant systems) and management strategies similar to contingency plans developed for other forms of infrastructure failure (eg, reversion to paper charting when EHRs go offline).

The error summary provided in table 1 provides valuable insights into the clinical significance of errors that may be made using the manual system. In general, the protocol attempts to minimise the risk of intravenous insulin administration by requiring hourly glucose checks. This frequent safety check protects against both administration errors and variability in patients' responses to insulin. Thus, errors of between 0.5 and 1 unit of insulin/hour are unlikely to cause severe hypoglycaemic or hyperglycaemic over the course of 1 hour. However, errors from incorrect direction of change and inappropriate stopping of the infusion lead to much higher differences from the correct infusion rates. These errors accounted for 40% of the errors sampled and it is very likely that they would significantly alter the serum glucose values. The magnitude of such glucose changes and their impact on patients cannot be generalised but are subjects for subsequent study.

Complacency with or over-reliance on automation is a well-documented phenomenon in the human performance literature,²⁴ and a concern raised by participants as well. Over-reliance on automation occurs when novice or expert users manage multiple concurrent tasks, and manual tasks compete with automation for attentional resources.²⁵ However, the tasks automated by the SA system are information management and

calculation, two types of tasks known to be error prone for humans even outside of multitasking demands.²⁶ The ultimate decision on whether or not to accept protocol recommendation remains with the end user as some action is required to accept and implement any changes. This study was not designed to assess the impact of automation on overall situational awareness of a patient's trajectory. Additionally, complacency is not a problem unique to automation, but general to any work system. Experience can lead to overconfidence in performance (eg, in this study, more frequent protocol users tended to have higher error rates) or in the ability of other system components (eg, a double check) to catch and correct an error.

There are a number of limitations with this study. First, the study design isolated components of insulin infusion management that differed between the SA and manual systems and ignored aspects that did not change (nurse double check). We were surprised by the nearly 17% rate of errors, given nurses knew we were measuring their performance (which likely prompted maximum vs typical levels of performance²⁷) and the simulation environment was free from distractions and competing work demands. This and other data indicating an overall insulin administration error rate over 18% suggest that evaluation in a real clinical setting may demonstrate an even higher primary error rate before double check. Second, we did not evaluate the effect of the nurse double check on error elimination. Therefore, we cannot say with certainty that the SA system is safer than the manual nurse-managed system with human double check. However, recent reviews and primary studies continue to raise doubts about the efficacy of the human double check for medication administrations.^{28 29} Third, the SA was explicitly designed to work with our hospital's insulin management protocol and was evaluated using study participants who worked in this system and with this protocol. Thus, the SA system needs further testing in other hospital settings. Fourth, this simulation evaluation focused on insulin infusion administration. It did not represent the complexity of task management in a real clinical setting. A better estimate of the value of SA could be determined by performing in situ field studies or higher fidelity simulation-based evaluations. Finally, this study used one type of EHR and one smart pump. The functionality and interoperability of different EHRs and smart pumps varies widely, and the ability to replicate this type of system with different technologies is technically challenging.

CONCLUSION

Systems engineering approaches that integrate EHRs, medical devices and work systems can enable the types of automated systems evaluated in this study and benefit care delivery. This study illustrated the potential of this approach to jointly improve safety and efficiency and decrease workload by automating

information management and calculation tasks. The SA eliminated errors and reduced task completion time, providing safety alongside efficiency in the ICU. Integrated management and automation systems such as the SA should be pursued to further enhance systems-level improvements in safety and clinical workflow.

Author affiliations

¹Armstrong Institute for Patient Safety and Quality, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

²Department of Health Policy and Management, Bloomberg School of Public Health, School of Nursing; Institute for Clinical and Translational Research, Baltimore, Maryland, USA

³Anesthesiology and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

⁴University Hospitals of Cleveland, Shaker Heights, Ohio, USA

⁵Anesthesiology and Critical Care Medicine, Case Western Reserve University School of Medicine, Cleveland, Ohio, USA

Acknowledgements We thank Christine Holzmüller for edits to an early version of this manuscript.

Contributors MAR, PJP, MR and AS conceived of the study. MAR designed the study. ZD, NB and CD collected data. MAR and ZD analysed the data and drafted the manuscript. All authors revised the manuscript.

Funding This study was funded by Agency for Healthcare Research and Quality (Grant number: P30HS023553).

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Johns Hopkins University School of Medicine Institutional Review Board: IRB00158348.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available by request from the corresponding author.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Michael A Rosen <http://orcid.org/0000-0002-6621-906X>

Peter J Pronovost <http://orcid.org/0000-0002-9740-3775>

REFERENCES

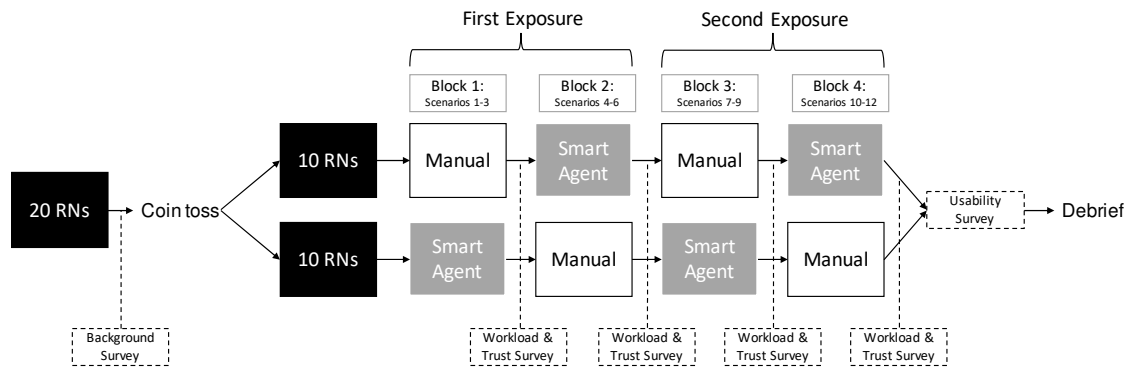
- 1 Mesotten D, Preiser J-C, Kosiborod M. Glucose management in critically ill adults and children. *Lancet Diabetes Endocrinol* 2015;3:723–33.

- 2 Jacobi J, Bircher N, Krinsley J, *et al.* Guidelines for the use of an insulin infusion for the management of hyperglycemia in critically ill patients. *Crit Care Med* 2012;40:3251–76.
- 3 Kanji S, Singh A, Tierney M, *et al.* Standardization of intravenous insulin therapy improves the efficiency and safety of blood glucose control in critically ill adults. *Intensive Care Med* 2004;30:804–10.
- 4 Aragon D. Evaluation of nursing work effort and perceptions about blood glucose testing in tight glycemic control. *Am J Crit Care* 2006;15:370–7.
- 5 Sowan AK, Vera A, Malshe A, *et al.* Transcription errors of blood glucose values and insulin errors in an intensive care unit: secondary data analysis toward electronic medical Record-Glucometer Interoperability. *JMIR Med Inform* 2019;7:e11873.
- 6 Garrouste-Orgeas M, Timsit JF, Vesin A, *et al.* Selected medical errors in the intensive care unit: results of the IATROREF study: parts I and II. *Am J Respir Crit Care Med* 2010;181:134–42.
- 7 Independent double checks: worth the effort if used judiciously and properly. Acute Care Newsletter Issue [Internet]. 2019 Mar 4, 2020. Available: <https://www.ismp.org/resources/independent-double-checks-worth-effort-if-used-judiciously-and-properly>
- 8 Douglass AM, Elder J, Watson R, *et al.* A Randomized Controlled Trial on the Effect of a Double Check on the Detection of Medication Errors. *Ann Emerg Med* 2018;71:74–82.
- 9 Armitage G. Double checking medicines: defence against error or contributory factor? *J Eval Clin Pract* 2008;14:513–9.
- 10 Hewitt T, Chreim S, Forster A. Double checking: a second look. *J Eval Clin Pract* 2016;22:267–74.
- 11 Shetty S, Inzucchi SE, Goldberg PA, *et al.* Adapting to the new consensus guidelines for managing hyperglycemia during critical illness: the updated Yale insulin infusion protocol. *Endocr Pract* 2012;18:363–70.
- 12 Griffiths SM, Sapirstein A, Guzman JC, *et al.* Automated, web-based solution for bidirectional EHR-Infusion pump communication. *Biomed Instrum Technol* 2019;53:30–7.
- 13 Barasch N, Romig MC, Demko ZO. Automation and interoperability of a nurse-managed insulin infusion protocol as a model to improve safety and efficiency in the delivery of high-alert medications. *Journal of Patient Safety and Risk Management* 2019;2516043519893228.
- 14 Hart SG, Staveland LE. Development of NASA-TLX (task load index): results of empirical and theoretical research. *Advances in psychology* 1988;52:139–83.
- 15 Jian J-Y, Bisantz AM, Drury CG. Foundations for an empirically determined scale of trust in automated systems. *Int J Cogn Ergon* 2000;4:53–71.
- 16 Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *Int J Hum Comput Interact* 2008;24:574–94.
- 17 Faul F, Erdfelder E, Buchner A, *et al.* Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods* 2009;41:1149–60.
- 18 Rosen MA, Dietz AS, Lee N, *et al.* Sensor-based measurement of critical care nursing workload: Unobtrusive measures of nursing activity complement traditional task and patient level indicators of workload to predict perceived exertion. *PLoS One* 2018;13:e0204819.
- 19 Furniss D, Dean Franklin B, Blandford A. The devil is in the detail: how a closed-loop documentation system for IV infusion administration contributes to and compromises patient safety. *Health Informatics J* 2020;26:576–91.
- 20 Lee JD, See KA. Trust in automation: designing for appropriate reliance. *Hum Factors* 2004;46:50–80.
- 21 Holden RJ, Karsh B-T. The technology acceptance model: its past and its future in health care. *J Biomed Inform* 2010;43:159–72.
- 22 Schurr JW, Stevens CA, Bane A, *et al.* Description and evaluation of the implementation of a Weight-Based, Nurse-Driven heparin nomogram in a tertiary academic medical center. *Clin Appl Thromb Hemost* 2018;24:248–53.
- 23 Epstein JM. Agent-based computational models and generative social science. *Complexity* 1999;4:41–60.
- 24 Parasuraman R, Riley V. Humans and automation: use, misuse, disuse, abuse. *Hum Factors* 1997;39:230–53.
- 25 Parasuraman R, Manzey DH. Complacency and bias in human use of automation: an attentional integration. *Hum Factors* 2010;52:381–410.
- 26 Boff KR, Lincoln JE. Engineering data compendium: Human perception and performance. In: Harry G, ed. *Armstrong aerospace medical research laboratory*, 1988.
- 27 Sackett PR, Zedeck S, Fogli L. Relations between measures of typical and maximum job performance. *Journal of Applied Psychology* 1988;73:482–6.
- 28 Koyama AK, Maddox C-SS, Li L, *et al.* Effectiveness of double checking to reduce medication administration errors: a systematic review. *BMJ Qual Saf* 2020;29:595–603.
- 29 Westbrook JI, Li L, Raban MZ, *et al.* Associations between double-checking and medication administration errors: a direct observational study of paediatric inpatients. *BMJ Qual Saf* 2021;30:320–30.

Appendix

1. Study overview visualization
2. Equipment and Test System
3. Scenario details
4. Participant debriefing form and questionnaires
5. Participant characteristics
6. Supplementary post hoc analysis of results by Participant self-reported frequency of protocol use

1. Study overview visualization



This figure illustrates the study design and flow of a simulation session with each critical care registered nurse (RN). Manual is the existing insulin infusion management system in which the nurse used the given patient name and searched the Electronic Health Record for the three values, used the protocol to calculate the correct insulin dose change, and adjusted the pump settings as needed. Smart Agent automates key steps of information retrieval from the Electronic Health Record, performs calculations, and communicates with the infusion pump.

2. Equipment and Test System

Technical and clinical process details of the SA system have been previously published.^(1, 2) At a high level, the SA system automates calculation, delivery, and documentation of the correct dose of insulin as determined by the institutionally established nurse-managed protocol. The system includes a user interface accessed as a web service from the Epic EHR. This interface is presented as a sidebar window of the active patient record. SA is integrated with the infusion pump (i.e., detects current settings, sends new settings) and the EHR (i.e., extracts patient lab and insulin infusion data, documents changes), and graphically maps the dose selection criteria to the steps of the protocol so that the clinician can see how the dose was determined. The clinician is given the option to accept or override SA dose calculations. If the clinician accepts the SA dose calculation, it is automatically sent to the infusion pump. The clinician then simply accepts changes on the pump, and avoids manual entry of changes.

References

- 1 Griffiths SM, Sapirstein A, Guzman JC, et al. Automated, Web-Based Solution for Bidirectional EHR-Infusion Pump Communication. *Biomed Instrum Technol* 2019;53(1):30-7.
- 2 Barasch N, Romig MC, Demko ZO, et al. Automation and interoperability of a nurse-managed insulin infusion protocol as a model to improve safety and efficiency in the delivery of high-alert medications. *J Patient Saf Risk Manag*. 2019:2516043519893228.

3. Scenario details

Information used in each of the 2 training and 12 experimental scenarios is provided below. For each scenario, participants looked up a patient record in the test Electronic Health Record (EHR) environment, and used the prior infusion rate, and prior and current glucose readings to calculate a change in glucose and new infusion rate. The infusion rate recommended by the protocol for each patient is provided below in the ‘new infusion rate’ column. Information in the last three columns of the table below were not provided to participants, but are provided here as the ‘correct’ solution (i.e., the changes recommended by the protocol) for each case.

Block	Patient	Prior Infusion Rate Entry Time in MAR	Prior Infusion rate in MAR & on Pump (insulin units)	Prior Glucose Reading Time	Prior Glucose Reading (mg/dL)	Current Glucose Reading Time	Current Glucose Reading (mg/dL)	Change in Glucose Reading (mg/dL)	New Infusion Rate (insulin units)	Change in Infusion Rate (insulin units)
Orientation	Insulin, Testa	0600 am	3	0545 am	251	650 am	242	-9	4	1
	Insulin, Testb	0600 am	2	0545 am	193	650 am	149	-44	1.5	-0.5
1	Insulin, Testc	0600 am	1	0545 am	168	650 am	154	-14	1	0
	Insulin, Testd	0600 am	4	0545 am	325	650 am	317	-8	6	2
	Insulin, Teste	0600 am	1.5	0545 am	128	650 am	128	0	1	-0.5
2	Insulin, Testf	0600 am	3.5	0545 am	285	650 am	215	-70	2.5	-1
	Insulin, Testg	0600 am	2.5	0545 am	107	650 am	158	51	3.5	1
	Insulin, Testh	0600 am	3	0545 am	125	650 am	179	54	5	2
3	Insulin, Testi	0600 am	4.5	0545 am	288	650 am	232	-56	3.5	-1
	Insulin, Testj	0600 am	2	0545 am	134	650 am	178	48	2.5	0.5
	Insulin, Testk	0600 am	4	0545 am	375	650 am	339	-36	5	1
4	Insulin, Testl	0600 am	3.5	0545 am	215	650 am	168	-47	2.5	-1
	Insulin, Testm	0600 am	1	0545 am	205	650 am	185	-20	1.5	0.5
	Insulin, Testn	0600 am	3.5	0545 am	242	650 am	249	7	5.5	2
Extra	Insulin, Testo	0600 am	2	0545 am	135	650 am	139	4	2	0

EHR = Electronic Health Record

MAR = Medication Administration Record

4. Participant debriefing form and questionnaires

BACKGROUND SURVEY ITEMS

How many years have you been a nurse?

How many years have you been a critical care nurse?

How often do you care for a patient using the Yale Infusion Protocol?

Every shift (more than once a week)

Once a week

Once a month

Less than once a month

POST SCENARIO / BLOCK SURVEY ITEMS

NASA Task Load Index Scale

Using the scale below, please rate the degree of demand you felt while completing the scenarios. **Focus your responses on your experience in the last set of scenarios.**

- How mentally demanding was the task?
- How physically demanding was the task?
- How hurried or rushed was the pace of the task?
- How successful were you in accomplishing what you were asked to do?
- How hard did you have to work to accomplish your level of performance?
- How insecure, discouraged, irritated, stressed, and annoyed were you?

Trust in automation scale

Instructions:

Below is a list of statements for evaluating trust between people and automation. There are several scales for you to rate intensity of your feeling of trust, or your impression of the system while operating it. Please indicate your level of agreement with the following statements on the scale provided. **Provide your responses while thinking about the system you just used, in the last block of scenarios.**

Note: Responses are recorded on a seven-point scale (1 = not at all; 7 = extremely)

- The system is deceptive.
- The system behaves in an underhanded manner.
- I am suspicious of the system's intent, action, or outputs.
- I am wary of the system.
- The system's actions will have a harmful or injurious outcome.
- I am confident in the system.
- The system provides security.
- The system has integrity.
- The system is dependable.
- The system is reliable.
- I can trust the system.
- I am familiar with the system.

POST SESSION SURVEY

Participants will complete the System Usability Scale twice, once for each of the conditions. The order of completing surveys will be counterbalanced across participants.

System Usability Survey

I think that I would like to use this system frequently.

I found the system unnecessarily complex.

I thought the system was easy to use.

I think that I would need the support of a technical person to be able to use this system.

I found the various functions in this system were well integrated.

I thought there was too much inconsistency in this system.

I would imagine that most people would learn to use this system very quickly.

I found the system very cumbersome to use.

I felt very confident using the system.

I needed to learn a lot of things before I could get going with this system.

Smart Agent Usability Testing Debrief form

- What do you think about the SmartAgent system?
 -
- Do you think it would be helpful to have a system like this?
 -
 - Why or why not?
 -
 - Would it make managing insulin infusions...
 - Easier?
 - More efficient?
 - Safer?
- Can you think of anything that would make SmartAgent more useful or usable?
 -
 - Can we improve clarity or simplicity of its use?
 -
 - Can we do anything to make it more efficient / fit your workflow better?
 -
 - Can we provide better feedback about what it is doing?
 -
- Do you foresee any problems with this system if it is eventually put in use?
 -
 - Are there any risks or hazards to using it?
 -
 - Do you think other nurses would use it?
 -
- Any other comments or feedback about SmartAgent?
 -

5. Participant characteristics

Nursing Experience	Mean (SD)	Median (IQR)	Range
Years of nursing experience	5.0 (8.9)	2 (3.8)	0.6 - 37
Years of critical care nursing experience	4.5 (8.6)	2 (1.4)	0 - 35

Insulin Protocol Use	Frequency	N	%
Frequency of use for the Yale Insulin Infusion Protocol	> Once/week	3	15
	Once/week	7	35
	Once/month	10	50

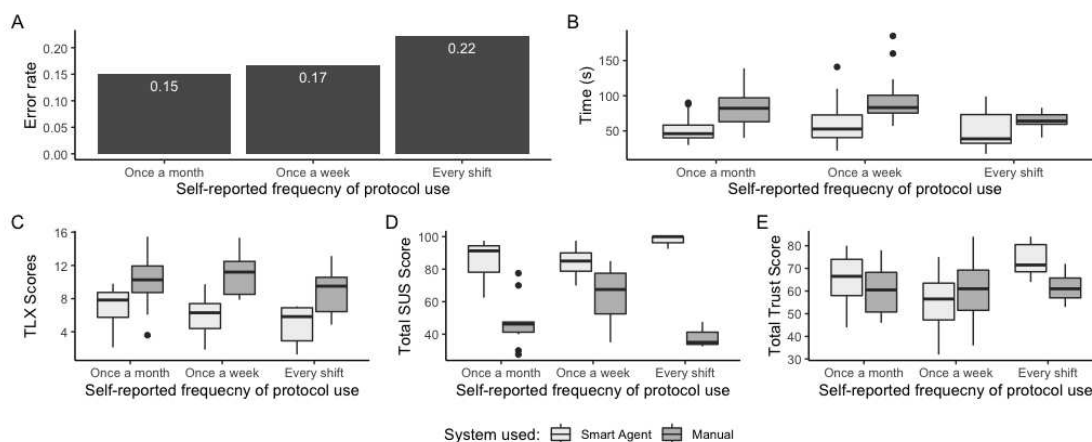
Unit Representation	N	%	
Type	ICU	14	70
	Stepdown	6	30
Surgery Specialty	Cardiac	9	45
	General	11	55

SD, standard deviation; IQR, interquartile range; ICU, intensive care unit

6. Supplementary post hoc analysis of results by Participant self-reported frequency of protocol use

In this section we provide a descriptive analysis of study outcomes by self-reported protocol use in their work. Of the 20 participants, 15% (3) reported using an insulin infusion protocol every shift (or more than once per week), 35% (7) at least once per week, and 50% (10) at least once a month. Due to a relatively low sample size we provide only descriptive data.

Study outcomes by condition and self-reported frequency of protocol use



A: Error rates by self-reported protocol use for Manual condition only
 B: Time (in seconds) to complete scenario by condition and self-reported protocol use
 C: Perceived workload (weighted NASA TLX) by condition and self-reported protocol use
 D: Perceived usability (Total Systems Usability Scale score) by condition and self-reported protocol use
 E: Trust by condition and self-reported protocol use

Error rates

We report error rates for the manual condition only (as no errors occurred in the smart agent condition). As shown below, there appears to be a trend for HIGHER overall error rates with increasing self-reported frequency of protocol use.

Self-reported frequency of use	Number of participants (Number of trials)	Total Error Count for Manual Condition	Overall Error Rate for Group
Every shift (more than once a week)	3 (18)	4	22%
At least once a week	7 (42)	7	17%
At least once a month	10 (60)	9	15%

Efficiency

	Efficiency: Mean (Standard Deviation) in seconds		
	Total	Smart Agent	Manual
Every shift (more than once a week)	56.5 (21.0)	51.6 (15.1)	65.2 (11.6)
At least once a week	74.2 (28.8)	58.9 (23.8)	89.6 (25.0)
At least once a month	67.3 (26.0)	51.6 (15.1)	83.0 (25.1)

Workload

	NASA TLX Scores: Mean (Standard Deviation)		
	Total	Smart Agent	Manual
Every shift (more than once a week)	6.8 (3.5)	7.1 (2.6)	8.9 (3.2)
At least once a week	8.4 (3.5)	5.9 (2.3)	10.9 (2.6)
At least once a month	8.6 (2.9)	4.9 (2.2)	10.1 (2.7)

Usability

	Perceived Usability: Mean (Standard Deviation)		
	Total	Smart Agent	Manual
Every shift (more than once a week)	67.9 (32.9)	97.5 (4.3)	38.3 (8.0)
At least once a week	74.1 (17.6)	84.3 (9.2)	63.9 (18.6)
At least once a month	66.6 (32.9)	85.5 (11.7)	47.8 (15.6)

Trust

	Trust Scores: Mean (Standard Deviation)		
	Total	Smart Agent	Manual

Every shift (more than once a week)	67.7 (9.6)	73.7 (9.8)	61.7 (7.0)
At least once a week	58.4 (13.3)	55.6 (12.5)	61.3 (14.0)
At least once a month	63.2 (10.4)	65.6 (8.2)	61.7 (10.7)