

Research Paper ■

Impact of Clinical Reminder Redesign on Learnability, Efficiency, Usability, and Workload for Ambulatory Clinic Nurses

JASON J. SALEEM, PhD, EMILY S. PATTERSON, PhD, LAURA MILITELLO, MA, SHILO ANDERS, MA, MERCEDES FALCIGLIA, MD, JENNIFER A. WISSMAN, MS, EMILIE M. ROTH, PhD, STEVEN M. ASCH, MD, MPH

Abstract **Objective:** Computerized clinical reminders (CRs) were designed to reduce clinicians' reliance on their memory and to present evidence-based guidelines at point of care. However, the literature indicates that CR adoption and effectiveness has been variable. We examined the impact of four design modifications to CR software on learnability, efficiency, usability, and workload for intake nursing personnel in an outpatient clinic setting. These modifications were included in a redesign primarily to address barriers to effective CR use identified during a previous field study.

Design: In a simulation experiment, 16 nurses used prototypes of the current and redesigned system in a within-subject comparison for five simulated patient encounters. Prior to the experimental session, participants completed an exploration session, where "learnability" of the current and redesigned systems was assessed.

Measurements: Time, performance, and survey data were analyzed in conjunction with semi-structured debrief interview data.

Results: The redesign was found to significantly increase learnability for first-time users as measured by time to complete the first CR, efficiency as measured by task completion time for two of five patient scenarios, usability as determined by all three groupings of questions taken from a commonly used survey instrument, and two of six workload subscales of the NASA Task Load Index (TLX) survey: mental workload and frustration.

Conclusion: Modest design modifications to existing CR software positively impacted variables that likely would increase the willingness for first-time nursing personnel to adopt and consistently use CRs.

■ *J Am Med Inform Assoc.* 2007;14:632–640. DOI 10.1197/jamia.M2163.

Introduction

Medical care is not consistently delivered in the United States.^{1–3} Computerized clinical reminders (CRs),⁴ a form of automated decision support, are designed to improve quality of care by reducing reliance of healthcare providers on their memory and by presenting accepted clinical guidelines

at point of care. Computer-based decision support, including CR systems, increase adherence to guidelines.⁵ However, CR effectiveness is quite variable.^{6–10} For example, CRs have been demonstrated to increase adherence to some preventive care guidelines, but not others.¹⁰ Within the Veterans Health Administration (VHA), there also exists variation in adherence to CRs by clinic, individual clinician, and individual CR.⁶ Previous research in the VHA using ethnographic observation^{11, 12} and survey^{13, 14} helps explain this known variation in CR effectiveness by identifying several organizational, workflow, and interface design barriers and facilitators to clinician use of CRs.

Affiliations of the authors: Veterans Association Getting at Patient Safety (GAPS) Center, VAMC (JJS, ESP, MF, JAW), Cincinnati, OH; Veterans Association HSR&D Center on Implementing Evidence-based Practice, Roudebush VAMC (JJS), Indianapolis, IN; Department of Electrical and Computer Engineering, IUPUI (JJS), Indianapolis, IN; Institute for Ergonomics (ESP), Cognitive Systems Engineering Laboratory (SA), The Ohio State University, Columbus, OH; University of Dayton Research Institute (LM), Dayton, OH; Department of Internal Medicine, University of Cincinnati (MF), Cincinnati, OH; Roth Cognitive Engineering (EMR), Brookline, MA; Health Services Research, VA Greater Los Angeles Healthcare System (SMA), Los Angeles, CA; Associate Professor of Medicine at the David Geffen School of Medicine (SMA) UCLA, Los Angeles, CA.

Supported by the Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service (TRX 02-216). A VA HSR&D Advanced Career Development Award supported Dr. Asch and a VA HSR&D Merit Review Entry Program Award supported Dr. Patterson.

The authors thank Ms. Suzanne Brungs for assistance with developing the patient scenarios and Dr. Marta Render for use of the VA GAPS Center Usability Laboratory.

The views expressed in this article are those of the authors and do not necessarily represent the view of the Department of Veterans Affairs.

Correspondence and reprints: Jason J. Saleem, PhD, VA HSR&D Center of Excellence on Implementing Evidenced-Based Practice (CIEBP), Richard L. Roudebush VA Medical Center (11-H), 1481 West Tenth Street, Indianapolis, IN 46202; e-mail: <jsaleem@iupui.edu>.

Received for review: 5/28/2006; accepted for publication: 6/11/2007.

Patterson and colleagues identified barriers to the effective use of ten CRs used for HIV care.¹¹ These barriers included workload, time to remove inapplicable reminders, false alarms, training, reduced eye contact, and the use of paper forms rather than software. A survey of VA providers in 2003 supported these barriers and also found that poor usability, accessibility (i.e., slow computer speed, not enough workstations), and a perception that administration benefits more than providers were additional barriers to use.^{13,14} A recent observational study that focused on general medicine clinics and all clinicians confirmed many of these previous results and discovered a key barrier not found in the previous studies: coordination issues between nurses and providers, including confusion of responsibilities for satisfying particular CRs.¹²

Usability of Clinical Information Systems

Many of these barriers to the use of CRs relate to usability and human-computer interaction (HCI). Usability and HCI have a rich literature base outside of healthcare. However, empirical work on usability of clinical systems is not as large. Kushniruk and Patel provided a comprehensive methodological review of cognitive and HCI methods¹⁵ such as usability testing, cognitive task analysis,¹⁶ and heuristic evaluation¹⁷ for evaluation of clinical information systems, including the importance of formative usability testing throughout the development process of a clinical information system in order to provide prompt and actionable feedback to developers prior to final implementation. Usability testing has been performed to evaluate clinical information systems^{16, 18-21} and can range in formality from informal qualitative studies with a few participants for exploratory purposes to a controlled laboratory simulation with a large number of participants for statistical comparisons across usability design metrics. The simulation study reported in this paper is reflective of the latter type of evaluation.

Approach and Hypotheses

Based on the findings from previous research,¹¹⁻¹⁴ we prototyped a redesign of the VHA CR system to reduce some of the barriers to CR use and conducted a laboratory simulation experiment to compare the redesign against the current CR system. We specifically based our redesign on two findings from the previous research that could be explicitly addressed at the computer interface level: nurse-physician coordination and poor usability. Barriers that relate more to clinic workflow and/or organizational issues were not in the scope of the present study and may be addressed by means other than interface redesign.

We hypothesized that the redesigned CR system would be (1) more "learnable" for first time users, (2) more efficient to use, (3) perceived as easier to use, and (4) perceived to have lower workload during its use than with the current CR system. We expect our findings to be relevant to healthcare organizations that currently use or are planning to implement computerized decision support systems. With the United States Congress advocating a national transition from paper-based to electronic medical record systems in the next ten years, this study may provide useful insight for incorporating automated decision support capability from the beginning, reducing the need for costly changes at a later date.

Methods

Participants

We recruited 16 non-VHA nurses to participate in a comparison study of the current VHA CR system and a redesigned prototype. A complete description of VHA Computerized Patient Record System (CPRS) and CRs, with supporting illustrations, is available in other articles^{11, 12, 22} and not included here to conserve space. We decided to plan a separate, future study for physicians because the workflow for the two users groups is quite different, as well as their interaction with the CR system. The nursing participants were all experienced with patient check-in/intake in outpatient clinic settings, but had no experience with the VHA's CPRS software so as not to bias one design over the other. Computer experience ranged from novice to expert. All participants were Registered Nurses (RNs), and one participant held the additional status of Nurse Practitioner (NP). Each nurse was given a \$50 gift certificate for his/her participation.

Apparatus

Physical set-up

The experiment, approved by the University of Cincinnati's Institutional Review Board, was conducted at the Veterans Administration (VA) Getting at Patient Safety (GAPS) Center's Usability Laboratory, located at the Cincinnati VA Medical Center. The usability laboratory was a closed setting, designed to simulate a clinic room in a typical VHA ambulatory clinic. Details of an actual clinic room, not directly related to CPRS-related tasks, were not simulated (e.g., physical exam instruments). Participants used a clinical computer workstation, and a researcher who played the part of the patient for the experimental session was seated next to and facing the nurse participant, as many of the actual VHA clinics are arranged. The experimenter's station consisted of media recording devices and a slave LCD TV monitor to observe the participant's screen. We recorded audio and two video sources, one of the computer screen interaction and one of the participant's face.

Prototypes

The redesigned prototype was programmed in Hyper Text Markup Language (HTML) as a low-fidelity mock-up. In other words, screen captures of the current design were used as a visual base and then redesigned. Links and buttons were made interactive to mimic the actual function of how the CRs would work if fully programmed. To enable us to compare the redesign with the way the current system functions, we also "prototyped" the current system in the same fashion so that both designs were at the same simulation fidelity level.

The redesigned prototype (B) differed from the current design (A) in four ways: (1) In design B, CRs were prefaced with a 'P' or 'N' for primary care provider or nurse, (2) CR dialog boxes were accessible directly from the cover sheet via a single mouse click rather than through the new progress note submenu, (3) CR dialog box formats were standardized; information was displayed as What, When, Who, and More Information, and (4) an electronic visit checklist feature was added which summarized tasks the clinician should complete during the patient encounter as well as other relevant clinical information (Figure 1). The

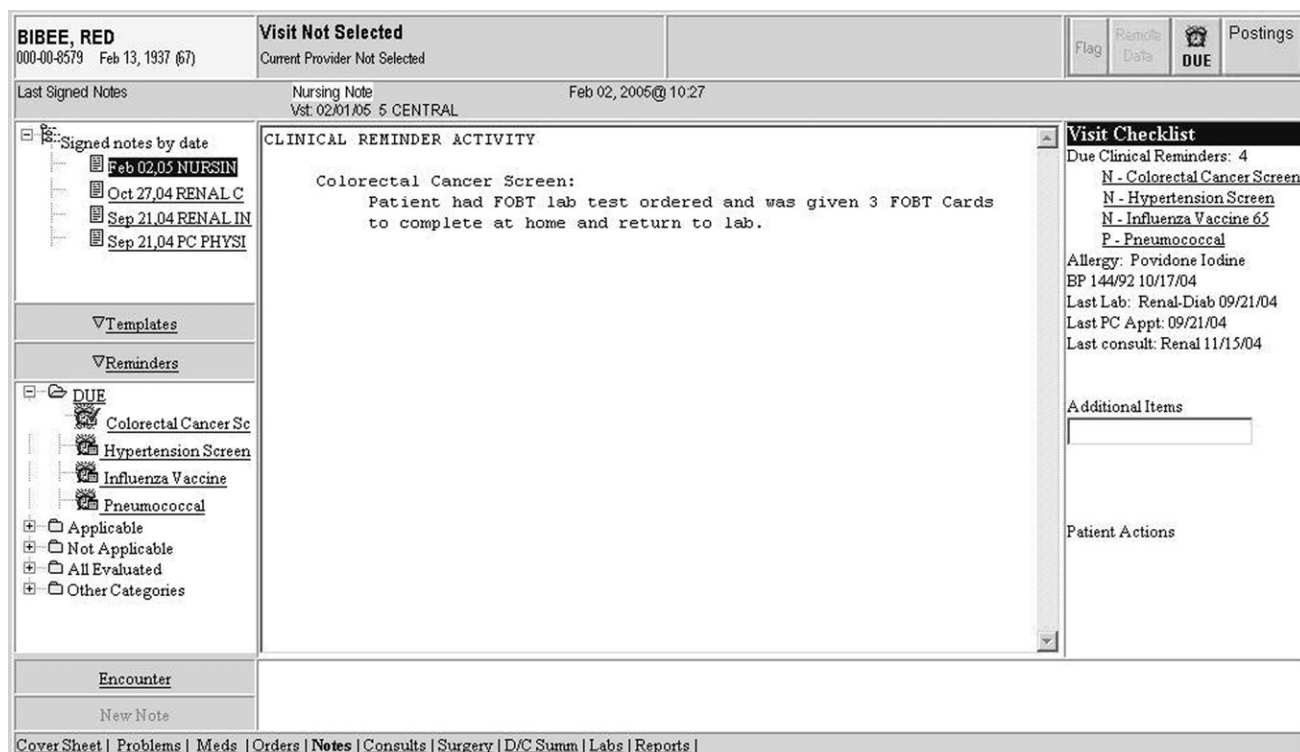


Figure 1. Design B includes an electronic visit checklist (right side). Note the CRs are prefaced with 'N's and 'P's for nurse/provider role assignment.

design changes we implemented were guided by the HCI literature. Nielson derived a set of nine general interface design principles from the HCI literature that are widely used today.²³ Table 1 shows how each of these design changes are mapped to HCI design principles from the literature.

Four design changes were implemented, primarily to address barriers to effective CR use identified during a field study observing both intake nurses and primary care providers using CRs during patient encounters in four VHA outpatient clinics.¹² Two barriers that we focus on from this previous study are nurse-physician coordination and poor usability, as these barriers could be explicitly addressed at the computer-interface level. By poor usability, we mean the seemingly inadequate visibility and accessibility of the CRs, as well as the observed difficulty in navigating the CR dialog boxes to satisfy them efficiently.¹² Two design changes targeted improving the learnability and usability of the system for first-time users, such as the resident physicians who were found to use CRs less than others: 1) accessing CR dialog boxes by clicking directly on the names of CRs on the interface prior to opening a progress note, and 2) prefacing the CR names with N's and P's to clarify which CRs were

assigned to intake nurses and physicians, respectively. Note that, although these changes targeted first-time users, experienced users could also benefit. The first design change requires fewer "clicks" to remove a CR, thereby improving efficiency, and the second design change could reduce confusion due to "floating" personnel between clinics with different assignments and changes in responsibility over time, which was found to occur both from physicians to nurses and nurses to physicians in the prior study for a variety of reasons. We designed simple access to the CR dialog boxes to help address the barrier of poor usability found in previous research.¹²⁻¹⁴ Prefacing the CRs with role assignments to reduce confusion was a specific attempt to address the barrier of coordination issues between nurses and providers.¹²

Like the first design change, the third and fourth design changes also targeted the poor usability barrier to CR use.¹²⁻¹⁴ Standardizing the formatting of text on dialog boxes that were used to remove "due" CRs was the third design change. Grouping and standardizing information presentation was expected to increase efficiency and usability. In addition, in several cases, observed care providers erroneously judged that the CR software had a "glitch"

Table 1 ■ Mapping Design Changes to Interface Design Principles

Design Changes to Prototype	Corresponding Interface Design Principle(s) ²³
(1) Prefacing CRs with a 'P' or 'N' for primary care provider or nurse	Consistency and Standards, Error Prevention
(2) Making CR dialog boxes accessible directly from the cover sheet	Flexibility and Efficiency of Use
(3) CR dialog box formats were standardized	Consistency and Standards, Simplicity and Aesthetic Integrity
(4) An electronic visit checklist to summarize tasks	Memory, Visibility of System Status

CR = Clinical Reminder.

Table 2 ■ Clinical Reminders for Patient Scenarios

Scenarios	Clinical Reminders
1A, 1B*	Pain Screening
2A, 2B	Colorectal Cancer Screen, Hypertension Screen, Influenza Vaccine, Pneumococcal
3A, 3B	Advance Directives Screen, Colorectal Cancer Screen, Depression Screen, Hypertension BP >140/90, Hypertension Screen/BP Check, Influenza Vaccine, Pain Screening, Pneumococcal, Tobacco Use Screen
4A, 4B	Alcohol Use Screen, IHD Aspirin Therapy Use, Hypertension Screen/BP Check, Nutrition/Obesity Screen, PTSD Screen
5A, 5B	Hypertension Screen/BP Check, Pneumococcal, Tobacco Screen
6A, 6B	Diabetic Eye Exam, Diabetic Foot Exam, Exercise Screen, Healthwise Book, Hypertension BP>140/90, Hypertension Screen/BP Check, Influenza Vaccine, Nutrition/Obesity Screen, Pain Screen, Tetanus Diphtheria

*CRs prefaced with 'N' or 'P' for nurse or provider, respectively, in all design B scenarios (1B-6B).

because they did not understand why the CR applied to that patient. Displaying the variables which triggered the CR was expected to reduce frustration and workload.

The final design change, an electronic checklist that summarized tasks to complete during a patient encounter, was designed primarily for primary care providers rather than intake nurses. However, since the checklist would be viewable by nursing personnel as the interface is not tailored to roles, we were interested to see if the presence of the checklist would cause any unintended effects, positive or negative, for the study participants.

Procedure

Exploration Session

Each participant was introduced to designs A and B, in a counter-balanced fashion (i.e., participant 1 used design A first, participant 2 used design B first, etc.). Each participant was given brief written instructions to satisfy a Pain Screening CR for each design, with relevant patient information necessary to satisfy the CR. Time to satisfy a single Pain Screening CR without prior training was recorded automatically through the time-stamped digital video recording, with a maximum time limit of five minutes per design. We used this time measurement as an a priori measure of learnability for first-time users. Participants were presented with both designs so that they had equal exposure to each, but only the data from the first design presented, design A or B, was used in the statistical analysis to eliminate the carryover learning effect. After completing the CR task for both design A and B and before the experimental session, participants received five minutes of training on completing CRs using each design, with the same counterbalanced design order. We provided training to participants to give them the similar training related to the CR system that new VA employees receive prior to using the CPRS.

Experimental Session

Each participant was given instructions to complete five patient scenarios for each of the designs A and B. The fictitious patients included both males and females, ranging in age from 33 to 70 years old, each with multiple health complaints. Differences between the paired patient scenarios across A and B were "surface-level" only (e.g., name, social security number, chief complaint) to reduce variability with user performance not related to the design of the CRs. With the assistance of a practicing physician and co-author (MF), and well as an experienced RN, the patient scenarios were developed that covered a range of disease areas, gender, age, medications, and other clinical factors that would substan-

tiate the inclusion of the various CRs listed in Table 2. During the development of each scenario, the physician and nurse reviewed the scenario separately and suggested improvements to maximize realism in an iterative fashion until the current scenarios used for the simulation were established.

Each scenario simulated an actual patient encounter of a nursing intake (check-in) appointment, with one of the authors (JW) playing the role of every patient.

Patient scenario presentation order was randomized for each participant, and paired across designs A and B to reduce extraneous variability (i.e., the same randomized patient scenario order was used for both designs). Design presentation order A and B were counterbalanced across participants as in the exploration session and in the same order of design (i.e., the first participant used design A first in the exploration session and the experimental design; the second participant used design B first in both sessions). Participants completed all five patient scenarios for one design (A or B) before completing the five scenarios for the other design. The computerized version of the NASA Task Load Index (TLX)^{24, 25} was administered to the participants after each patient scenario, a total of ten times per participant. The usability questionnaire was administered after the participant had finished all five patient scenarios for both designs. After the experimental conditions, we conducted a semi-structured debrief interview to gather additional feedback on the redesigned interface.

Data Collection / Dependent Measures

We measured time to complete a CR in the exploration and experimental sessions for both designs through the time-stamped digital video recordings. Completion time in the exploration session was a surrogate measure of "learnability" without prior training. Time measurement in the experimental session corresponded to "efficiency". We adapted a commonly used usability questionnaire^{26, 27} (see Table 3) that uses a 7-point Likert-type scale to measure user satisfaction during the experimental session.

Workload measures were provided by the computerized version of the NASA-TLX during the experimental session as well. The NASA TLX is a rating scale commonly used in human factors research to assess perceived workload across several dimensions (Figure 2). This scale has been validated as being sensitive to detect changes in perceived workload across varying levels of task difficulty²⁴ and has been used in several domains to assess operator workload in complex environments, including aviation,²⁸ nuclear power plants,²⁹

Table 3 ■ Usability Survey*

Questions and Groupings
Group 1
1. Overall, I am satisfied with how easy it is to use this system.
Group 2
15. Overall, I am satisfied with this system.
Group 3
2. It was simple to use this system.
3. I can effectively complete my work using this system.
4. I am able to efficiently complete my work using this system.
5. It was easy to learn to use this system.
6. Whenever I make a mistake using the system, I recover easily and quickly.
7. It is easy to find the information I needed.
8. The information in the system is easy to understand.
9. The information in the system is effective in helping complete the tasks & scenarios.
10. The organization of the information on the systems screens is clear.
11. The display layouts simplify tasks.
12. The sequence of displays is confusing.
13. Human memory limitations are overwhelmed.
14. I like using the interface of this system

Questions 11, 12 from Shneiderman questionnaire⁴⁴; all others from Lewis questionnaire.^{26,27} See reference sources for instrument psychometrics.

*Adapted from Common Usability Questionnaires^{26,27,44}

ground transportation,³⁰ and healthcare,³¹ as well as tasks involving long-durance vigilance,³² for example. We used unweighted TLX scores as the TLX dimensional weighting procedure has been found to be of limited benefit.^{33, 34} Finally, experimenter notes were recorded during each patient scenario detailing incidents such as inaccurate user assumptions and incorrect CR completion.

Statistical Methods

The experimental design was a within-subject A vs. B comparison (the single factor was Design Type). To test hypothesis (1), that design B would be more “learnable” than A, we used a t-test to compare the time to complete a CR with A and B from the exploration session. We used only the data from the first design each participant was presented to control for the carryover learning effect. For hypothesis (2), that design B would be more efficient to use than A, we used paired t-tests to compare the data for corresponding scenarios across A and B in the experimental session. To test hypothesis (3), that participants would have greater user satisfaction with design B than with A, we grouped similar usability questions together and used the non-parametric Wilcoxon Signed Ranks Test to compare the 7-point Likert-type scale responses across A and B. Responses to a single Likert-type item are normally treated as ordinal data in which case a non-parametric test is appropriate. However, summed responses to several Likert items may be treated as interval data measuring a latent variable. Thus in the case of grouped usability questions, we used a paired t-test. For hypothesis (4), that participants would have a lower perceived workload using design B than with A, we conducted analysis of variance (ANOVA) for Design Type on the TLX ratings and conducted paired t-tests for each TLX subscale (mental demand, physical demand, temporal demand, performance, effort, and frustration level) across design A and B. We used a 0.05 level of significance for all statistical tests. One participant’s TLX data was determined to be an outlier as it trended considerably in the opposite direction of the same data from the other participants. Further, the partici-

part’s TLX data trended in the opposite direction from her TLX data recorded in the exploration session; the TLX was administered to the participants in the exploration session as practice. These two contradictions suggest that the data was mislabeled when it was archived post-test. All data from this participant, including the performance data and usability survey responses, were excluded in the analyses and an additional nurse was recruited in order to retain 16 study participants. However, we also provide the results with this outlier included after presenting the results without the outlier.

Results

We used one-tailed statistical tests to test our hypotheses since they are directional. However, we also report the two-tailed test results to give the reader a more complete view of the findings.

Learnability for First-Time Users

We used time to complete the first CR as a measure of “learnability” of the interface for first-time users. For the eight individuals who used design A first, only one participant was able to complete the CR task within the five minute limit. Conversely, for those eight individuals who used design B first, only one participant did not complete the CR task within the five minute limit. A one-tailed t-test shows that time in seconds to satisfy a CR with design B ($M = 141$, $SD = 84.6$) was significantly less than time with design A ($M = 286$, $SD = 40.3$), $t(14) = 4.37$, $p < 0.001$. The two-tailed result also gives $p < 0.001$.

Efficiency

In the experimental session, time to complete the patient scenarios with design A and design B was not statistically significant for three of the five patient scenarios, likely due to learning effects from the exploration session. However, participants did take significantly less time to complete two patient scenarios for design B (4B and 5B) than the corresponding patient scenarios for design A (4A and 5A; Table

Figure 2. NASA Task Load Index (TLX) for Perceived Workload Assessment.^{24, 25}

2). A paired one-tailed t-test showed that time in seconds to complete scenario 4B ($M = 264$, $SD = 114.6$) was significantly less than with scenario 4A ($M = 340$, $SD = 137.0$), $t(15) = 2.23$, $p = 0.02$. The two-tailed result is also significant at $p = 0.04$. The same test also showed time in seconds to complete scenario 5B ($M = 112$, $SD = 43.0$) was significantly less than with scenario 5A ($M = 160$, $SD = 61.2$), $t(15) = 2.81$, $p = 0.007$ (two-tailed result significant at $p = 0.01$). It is unclear why these scenarios were different than the other three. Note that the number of CRs was not the largest (see Table 2) and that patient scenario order was randomized, so explanations related to these factors are unlikely.

Usability

We grouped the questions from the 7-point Likert-type usability survey (1 = strongly disagree, 4 = undecided, 7 = strongly agree) for analysis by overall ease of use (question 1), overall satisfaction (question 15), and by grouping the responses for the more detailed usability constructs (questions 2–14). We established these groupings to avoid excessive paired comparisons, thus reducing the risk of commit-

ting a Type I error. Since questions 1 and 15 relate to ‘overall’ assessment, we decided to group these questions separately. Table 3 lists the specific questions. Design B scored significantly better than A on the single Likert-type items for overall ease of use (design A: $M = 4.1$, $SD = 1.6$; design B: $M = 5.5$, $SD = 0.8$) and overall satisfaction (design A: $M = 3.9$, $SD = 1.7$; design B: $M = 5.6$, $SD = 1.2$) by the Wilcoxon Signed Ranks Test, $p < 0.05$ (two-tailed). Design B scored significantly better than A for the average responses across the 13 detailed usability constructs (design A: $M = 4.3$, $SD = 0.3$; design B: $M = 5.4$, $SD = 0.3$), $t(15) = -5.18$, $p < 0.001$ (one-tailed). The two-tailed result also gives $p < 0.001$.

Workload

An analysis of variance (ANOVA) was performed on the ratings of the NASA TLX workload dimensions. ANOVA did reveal a significant main effect of the TLX subscales on workload, $F(5, 75) = 6.76$, $p < 0.001$, but no significant main effect of overall Design Type on workload, $F(1, 15) = 2.53$, $p = 0.133$. An individual comparison of each subscale across A and B using paired t-tests did reveal significantly lower ratings for mental demand, $t(15) = 1.84$, $p = 0.04$ and frustration, $t(15) = 2.04$, $p = 0.03$ (one-tailed), but not for the other subscales. The two-tailed results for mental demand and frustration are $p = 0.09$ and $p = 0.06$, respectively.

Results with Outlier Included

The results of the analysis with the outlier included do not change the significant results found for learnability, efficiency, or usability. However, the results for the NASA TLX subscales, ‘mental workload’ and ‘frustration’, would not be statistically significant with this outlier included: mental demand, $t(15) = 1.36$, $p = 0.10$ and frustration, $t(15) = 1.23$, $p = 0.12$ (one-tailed). The two-tailed results for mental demand and frustration with the outlier included are $p = 0.19$ and $p = 0.24$, respectively.

Unanticipated Effects

The audio and video recordings revealed two unanticipated effects of the CR design modifications in design B, one positive and one negative. First, a design modification allowed participants to satisfy the CRs directly from the initial cover sheet on the CPRS. An unexpected benefit from this change was that users were able to view the summarized patient information such as the patient’s problem and medication lists during CR tasks. Participants reported this information as helpful in assessing CR relevance to the patient and entering required information to remove a CR. In the current design, this data is only visible by switching tabs when completing CRs. Second, a design modification streamlined the CR dialog box formatting with a What-When-Who format to standardize the presentation information across CRs. Although most participants responded positively to this change in general, two participants misinterpreted the meaning of “When” to represent the last time the current patient received the intervention instead of the frequency the intervention is due for all patients.

Discussion

We explored how learnable, efficient, and usable a redesigned interface was for ambulatory clinic nurses, as well as the impact of the redesign on their workload. We redesigned the CPRS and CR interface with four design modifications to specifically target barriers of nurse-physician coordination

and poor usability found in previous research.¹²⁻¹⁴ A redesigned interface for CRs with these four modest changes was found to significantly increase learnability for first-time users as measured by time to complete the first CR. The redesigned interface was also found to significantly increase efficiency as measured by task completion time for two of five patient scenarios, usability as determined by all three groupings of questions taken from a commonly used survey instrument, and two of six workload subscales of the NASA TLX survey: mental workload and frustration.

As with most user-interface testing, there were individual differences observed in the user data that are useful in interpreting the results. One strategy to accommodate user differences is to redesign the interface with the goal of changing design features responsible for large performance differences among users.³⁵ This strategy of designing a more robust interface was the approach we chose by implementing the four design modifications to the redesigned prototype, especially the modification enabling users to access and satisfy CRs directly from the coversheet. This modification was meant to reduce the performance gap between first-time and more experienced users, as we learned from previous research that many novice users did not know how to access the CRs.¹² The redesigned prototype likely helped reduce individual performance differences as evidenced by the learnability results from the exploration session. The largest performance differences in the experimental session were on the order of 3:1 for time to complete all five scenarios for a given design. This difference is quite modest, as performance differences for computer tasks on the order of 20:1 are not uncommon.³⁵

One positive and one negative unintended effect of the changes in design B were identified. The positive effect complements the efficiency findings since being able to view the summarized patient information from the cover sheet while completing the CRs likely reduced time to complete the CR dialog box fields and thus the patient scenarios. No changes are recommended in response to the positive effect. The negative effect of confusion about the label "When" could likely be addressed by relabeling the text as "Recommended frequency." With the exception of this change, the first three design modifications are recommended for immediate implementation in the national VHA CPRS software. The fourth design change, the electronic checklist that summarized tasks to complete during a patient encounter, was designed primarily for primary care providers and needs to be tested with physician participants prior to implementation. There were no unintended effects, negative or positive, identified for the electronic checklist on the nursing participants. These findings could also be relevant for other software packages, particularly ones where adoption is dependent upon learnability during first-time use, as well as efficiency and usability for experienced users.

From a methodological perspective, this paper demonstrates human factors methodologies such as workload assessment using the NASA TLX, user satisfaction measurement, and time-accuracy measurement through a meaningful, immediately applicable example. Simulation studies such as this one, as well as other human factors and sociotechnical analyses, can provide a well-suited approach for investigating human and organizational influences on the use of a

clinical information system.³⁶ These analyses should be conducted on clinical information systems prior to deployment to adequately support clinicians³⁷ and to increase their acceptance of the system.³⁸

This work was guided by a cognitive systems engineering framework.³⁹⁻⁴¹ Cognitive engineering is broader in scope than traditional usability engineering, such that the unit of analysis includes human and technological agents, as well as their interactions in the context of the larger sociotechnical system. However, this study is primarily driven by an atheoretical approach, as much of the HCI work is based at this time. Indeed, Landauer contends that for HCI, task-performance and simulation methods coupled with formative design evaluation are sufficient and strong guides; theories for HCI are imprecise, cover only limited aspects of behavior, and applicable to only some parts of the system studied,⁴² as clinical information systems are a conglomeration of many elements. On the other hand, there is a place for theory-based methods for this type of work as they can improve certain aspects of computer and informatics design. For example, cognitive psychology theories such as have had clear impact on some interface design.⁴³ For this study, the redesign modifications tested were informed by previous field observation¹² to help us first understand CR functioning in the greater sociotechnical system, or healthcare system, in which they are used. This type of qualitative field observation followed by scenario-driven, comparative usability testing of experimental prototypes in a simulated setting is complementary. Clinical software development may benefit from this paradigm if more widely followed.

Limitations

Participants were first-time users with no prior VHA experience and limited to nursing personnel. One participant's data was classified as an outlier and therefore not included in the statistical analysis. One of the authors (JW) played the role of the patient in each scenario and, by knowing the study design, could have inadvertently influenced the results. Certain participant characteristics such as computer experience could have potentially impacted the learnability results. A simulation study does not capture the full complexity of a sociotechnical system as it exists in real life. Differences observed in efficiency, user satisfaction, and workload measures between the current design and redesigned prototype reflect the redesigned prototype with the four designed changes as a whole; i.e., the four design changes to the redesigned prototype were not individually distinguishable in a statistical analysis. While the design change of prefacing the CRs with role assignments to reduce confusion was a meant to address the barrier found in previous research of coordination between nurses and providers, this particular study only provides evidence of how nurses, not providers, performed with the new and old design. Further study is needed to assess the impact of design on actual coordination between nurses and providers. Also, this study is limited by a relatively small number of participants; it is possible a larger N may have detected additional statistical differences. The baseline design A is the current VHA CPRS, not the improved "reengineered" (JAVA) version currently being developed by the VHA, which will likely incorporate the recommendations stemming from this study. However, after these recommenda-

tions are incorporated, further evaluation will be warranted at that point to ensure successful implementation.

We also expect findings from this study to have relevance and provide some guidance to other healthcare systems that are considering implementing electronic medical records with automated decision support functionality, as recommended by the US Congress. However, generalizability of our results to non-VHA hospitals may be limited due to unique factors of the VHA's CPRS and CR system and extensive informatics infrastructure. For example, the reminders in the electronic medical record at Brigham and Women's Hospital in Boston do not have the documentation aspect observed with the VA CR system.⁹ The VHA's CPRS is a mature system, and the CRs were added as a "layer" to this mature system, rather than designed during the inception of CPRS. However, as the VHA is currently transforming the underlying structure from the older Delphi programming language to a new JAVA-based platform, this will afford a unique opportunity to consider design changes that were previously constrained by the old architecture of CPRS.

Conclusions

We demonstrated that modest modifications to the VHA's current CR system significantly increased learnability, efficiency, and usability in a robust simulation, while decreasing mental workload and user frustration, thereby reducing barriers to adoption of CRs and effective CR use. Addressing these computer interface issues during the ongoing redesign of CPRS will likely require only modest additional resources and may reduce user frustration and increase effective CR use, thereby increasing adoption and reducing workarounds. The results reported in this paper related to nurses' use of CRs and cannot necessarily be extrapolated to physicians' use of CRs. Finally, this study may provide useful design information for healthcare organizations during the transition from paper-based to electronic medical records, and for those organizations that have already implemented such systems and are considering adding automated decision support functionality such as CRs.

References ■

- Asch SM, Fremont AM, Turner BJ, et al. Symptom-based framework for assessing quality of HIV care. *Int J Qual Health Care* 2004;16(1):41-50.
- McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *N Engl J Med* 2003;348(26):2635-45.
- Wennberg JE. Practice variations and health care reform: connecting the dots. *Health Aff (Millwood)* 2004;[Suppl Web Exclusives]:VAR140-VAR144.
- McDonald CJ. Protocol-based computer reminders, the quality of care and the non-perfectability of man. *N Engl J Med* 1976;295(24):1351-5.
- Hunt DL, Haynes RB, Hanna SE, Smith K. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *JAMA* 1998; 280(15):1339-46.
- Agrawal A, Mayo-Smith MF. Adherence to computerized clinical reminders in a large healthcare delivery network. *Medinfo* 2004;11(Pt 1):111-4.
- Demakis JG, Beauchamp C, Cull WL, et al. Improving residents' compliance with standards of ambulatory care: results from the VA Cooperative Study on Computerized Reminders. *JAMA* 2000;284(11):1411-6.
- Garg AX, Adhikari NK, McDonald H, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA* 2005; 293(10):1223-8.
- Sequist TD, Gandhi TK, Karson AS, et al. A randomized trial of electronic clinical reminders to improve quality of care for diabetes and coronary artery disease. *J Am Med Inform Assoc* 2005;12(4):431-7.
- Shea S, DuMouchel W, Bahamonde L. A meta-analysis of 16 randomized controlled trials to evaluate computer-based clinical reminder systems for preventive care in the ambulatory setting. *J Am Med Inform Assoc* 1996;3(6):399-409.
- Patterson ES, Nguyen AD, Halloran JP, Asch SM. Human factors barriers to the effective use of ten HIV clinical reminders. *J Am Med Inform Assoc* 2004;11(1):50-9.
- Saleem JJ, Patterson ES, Militello L, Render ML, Orshansky G, Asch SM. Exploring barriers and facilitators to the use of computerized clinical reminders. *J Am Med Inform Assoc* 2005;12(4):438-47.
- Fung CH, Woods JN, Asch SM, Glassman P, Doebbeling BN. Variation in implementation and use of computerized clinical reminders in an integrated healthcare system. *Am J Manag Care* 2004;10(11 Pt 2):878-85.
- Militello L, Patterson ES, Tripp-Reimer T, et al. Clinical Reminders: Why Don't They Use Them? *Proc Human Factors and Ergonomics Society's 48th Annual Meeting* 2004;1651-5.
- Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. *J Biomed Inform* 2004;37(1):56-76.
- Horsky J, Kaufman DR, Oppenheim MI, Patel VL. A framework for analyzing the cognitive complexity of computer-assisted clinical ordering. *J Biomed Inform* 2003;36(1-2):4-22.
- Zhang J, Johnson TR, Patel VL, Paige DL, Kubose T. Using usability heuristics to evaluate patient safety of medical devices. *J Biomed Inform* 2003;36(1-2):23-30.
- Beuscart-Zephir MC, Watbled L, Carpentier AM, Degroisse M, Alao O. A rapid usability assessment methodology to support the choice of clinical information systems: a case study. *Proc AMIA Symp* 2002;46-50.
- Coble JM, Karat J, Orland MJ, Kahn MG. Iterative usability testing: ensuring a usable clinical workstation. *Proc AMIA Symp* 1997;744-8.
- Rodriguez NJ, Murillo V, Borges JA, Ortiz J, Sands DZ. A usability study of physicians interaction with a paper-based patient record system and a graphical-based electronic patient record system. *Proc AMIA Symp* 2002;667-71.
- Wachter SB, Agutter J, Syroid N, Drews F, Weinger MB, Westenskow D. The employment of an iterative design process to develop a pulmonary graphical display. *J Am Med Inform Assoc* 2003;10(4):363-72.
- Patterson ES, Doebbeling BN, Fung CH, Militello L, Anders S, Asch SM. Identifying barriers to the effective use of clinical reminders: bootstrapping multiple methods. *J Biomed Inform* 2005;38(3):189-99.
- Nielson J. Enhancing the explanatory power of visibility heuristics. *CHI '94 Proceedings* 1994;152-8.
- Hart S, Staveland L. Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock PA, Meshkati N, eds. *Human Mental Workload*. North-Holland: Elsevier Science Publishers, 1988, pp. 139-183.
- NASA Task Load Index (TLX) for Windows. Available at: <http://www.nrl.navy.mil/aic/ide/NASATLX.php>. Accessed on: December 15, 2004.
- Lewis JR. IBM Computer Usability Satisfaction Questionnaires: Psychometric evaluation and instructions for use [Report]. 54.786. 1993. Boca Raton, FL, IBM Corporation, Human Factors Group.

27. Lewis JR. IBM Computer Usability Satisfaction Questionnaires: Psychometric evaluation and instructions for use. *Int J Human-Comp Inter* 1995;7(1):57–78.
28. Averty P, Collet C, Dittmar A, Athenes S, Vernet-Maury E. Mental workload in air traffic control: an index constructed from field tests. *Aviat Space Environ Med* 2004;75(4):333–41.
29. Park J, Jung W. A study on the validity of task complexity measure of emergency operating procedures of nuclear power plants—Comparing with a subjective workload. *IEEE Trans Nucl Sci* 2006;53(5):2962–70.
30. Hakan A, Nilsson L. The effects of a mobile telephone task on driver behaviour in a car following situation. *Acc Anal Prev* 1995;27(5):707–15.
31. Wachter SB, Johnson K, Albert R, Syroid N, Drews F, Westenskow D. The evaluation of a pulmonary display to detect adverse respiratory events using high resolution human simulator. *J Am Med Inform Assoc* 2006;13(6):635–42.
32. Temple JG, Warm JS, Dember WN, Jones KS, LaGrange CM, Matthews G. The effects of signal salience and caffeine on performance, workload, and stress in an abbreviated vigilance task. *Hum Factors* 2000;42(2):183–94.
33. Hendy KC, Hamilton KM, Landry LN. Measuring subjective workload: When is one scale better than many? *Hum Factors* 1993;35(4):579–601.
34. Nygren TE. Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Hum Factors* 1991; 33(1):17–33.
35. Egan DE. Individual differences in human-computer interaction. In: Helander M, ed. *Handbook of Human-Computer Interaction*. North-Holland: Elsevier Science Publishers; 1988, pp. 543–68.
36. Kaplan B. Evaluating informatics applications—clinical decision support systems literature review. *Int J Med Inform* 2001;64(1):15–37.
37. Doebbeling BN, Chou AF, Tierney WM. Priorities and strategies for the implementation of integrated informatics and communications technology to improve evidence-based practice. *J Gen Intern Med* 2006; 21[Suppl]2:S50–7.
38. Sittig DF, Krall MA, Dykstra RH, Russell A, Chin HL. A survey of factors affecting clinician acceptance of clinical decision support. *BMC Med Inform Decis Mak* 2006;6:6.
39. Hollnagel E, Woods DD. Cognitive systems engineering: New wine in new bottles. *Int J Human-Comp Stud* 1999;51:339–56.
40. Patterson ES, Bozzette SA, Nguyen AD, Gomes JO, Asch SM. Comparing findings from cognitive engineering evaluations. *Proc Hum Factors Ergonom Soc 47th Annu Meeting* 2003;483–487.
41. Roth EM, Patterson ES, Mumaw RJ. Cognitive engineering: Issues in user-centered system design. In: Marciniak JJ, ed. *Encyclopedia of Software Engineering*, 2nd ed. New York: Wiley-Interscience, John Wiley & Sons; 2002, pp. 163–79.
42. Landauer TK. Let's get real: A positional paper on the role of cognitive psychology in the design of humanly useful and usable systems. In: Baecker RM, Grudin J, Buxton WAS, Greenberg S, eds. *Readings in Human-Computer Interaction: Toward the Year 2000*. San Francisco: Morgan Kaufmann Publishers, Inc.; 1995, pp. 659–65.
43. Barnard P. The contributions of applied cognitive psychology to the study of human-computer interaction. In: Baecker RM, Grudin J, Buxton WAS, Greenberg S, eds. *Readings in Human-Computer Interaction: Toward the Year 2000*. San Francisco: Morgan Kaufmann Publishers, Inc.; 1995, pp. 640–58.
44. Shneiderman B, Plaisant C. *Designing the user interface: Strategies for effective human-computer interaction*. 4th ed. Boston, MA: Pearson/Addison-Wesley; 2004.