Context Sensitive Health Informatics: Human and Sociotechnical Approaches
M.-C. Beuscart-Zéphir et al. (Eds.)

doi:10.3233/978-1-61499-293-6-181

Essential Questions: Accuracy, errors and user perceptions in a drag/drop user-composable Electronic Health Record

Yalini SENATHIRAJAH[a,1], David KAUFMAN[b] and Suzanne BAKKEN[c]
[a] *SUNY Downstate Medical Center, Department of Medical Informatics*
[b] *Arizona State University, Department of Biomedical Informatics*
[c] *Columbia University School of Nursing and Department of Biomedical Informatics*

Abstract. In previous work we have described the creation and user testing of a drag/drop user-composable electronic health record, MedWISE. Any new design poses new potential problems; here we discuss the accuracy, potential for new types of errors, and user reactions to the approach, including their perceptions of ease of use and usefulness. Our results come from a mixed methods laboratory study with 18 clinicians at a large academic medical center. 12 used MedWISE and 6 used the conventional system.
MedWISE users had comparable assessment accuracy to use of the conventional system, low (1/10) risk of diagnosis momentum error, and overall favorable and even enthusiastic user perception. 10/12 said the system eased their cognitive process, Ease of use and usefulness were rated 3.79 and 4.00 on a 5-point Likert scale. Users were unconcerned about the possibility of increased errors due to their trust in colleagues and system similarities to current practice. After first describing the issues, we suggest a method to elucidate risks in innovative and current systems.

Keywords. MedWISE, electronic health record, user-composable, usability, errors, electronic medical record, human-computer interaction, cognitive support

Introduction

We previously described a novel model for creating healthcare information systems, which allows the clinician user to create and share his/her own Electronic Health Records (EHR) system and interfaces, assembling desired elements together by drag/drop, creating patient-specific displays without knowing programming [1]. In other studies, we detail the advantages of this approach for efficiency, cognitive support, time savings, and other aspects of the composable EHR, MedWISE [2]. MedWISE also allows the user to mashup any laboratory results, graphing them on the same axes, create custom lab panels by selecting from the complete list of laboratory tests done at this institution, view available information on timeline visualizations, share created elements with colleagues, set created interfaces as self-updating templates, and other capabilities [3].

[1] Corresponding Author: Yalini SENATHIRAJAH. Email: yas7001@dbmi.columbia.edu

The most commonly expressed caution about giving clinical users the ability to design/compose the EHR themselves is the possibility that omission of important elements, and/or propagation of such interfaces, may lead to medical errors. We believe that such systems are in equipoise with current practice, since aside from attested notes, we know of no institution which monitors and corrects the information viewing of its clinicians (e.g. by sending warnings). In conventional systems the user must get access to multiple pieces of information via multiple screens, while MedWISE allows the user to gather elements on the same screen. Both depend on the user's skill in not omitting items. In MedWISE omissions may be noticeable to others, as a previous users' composition is available (if shared) to colleagues, who may then amend it from the complete set of elements which are always available. By contrast, the browsing history of colleagues is usually not available in conventional systems, though it may be monitored by system administrators, for security purposes. Moreover the ability to juxtapose items can call attention to related items which should be monitored together and serve as a checklist. Error propagation due to shared interfaces with omissions is more serious, since propagation of an omission might have cascading effects if undetected.

Studies using simulated rounds found low (<50%) rates of error detection with a higher rate if subjects were primed by being warned that the case has errors [4]. Patel found experts detect more errors and recover faster [5]. Too-narrow initial differential diagnosis is the prime cause of diagnostic error [6]. The effectiveness of checklists [6] in improving care also points to the importance of omissions.

In order to examine these questions we conducted a small study of clinicians using the novel system to assess real patient cases; one case of which was designed to test the possibility of such errors. 'Diagnosis momentum' [7] is an error type in which a clinician adopts another clinician's line of reasoning and/or diagnosis due to being overly influenced and insufficiently independent, perhaps ignoring subtle clues to a different interpretation. As a small test of this, we presented clinicians with a real-life case (from the textbook 'Learning Clinical Reasoning') [8]. Using a real case ensured error was a realistic possibility.

It is also vital that new approaches support an overall diagnostic accuracy comparable to current systems. To test this, we had a senior clinician review the cases and create a reference standard of case elements (including diagnoses), to which the subjects' assessments were compared. This is a common informatics method [9].

User reactions to the novel approach are important for its viability; since perceptions or experience of non-usefulness or poor ease of use will deter adoption [10]. Therefore we evaluated user experience and perceptions with mixed methods described below. Table 1 lists the research questions and evidence types for the studies.

1. Methods

Twelve clinicians (7 medicine and 2 nephrology residents, 2 attending physicians and a physician assistant (PA)) were recruited via a focus group announcement and email from the hospitalist and nephrology departments of New York Presbyterian hospital (NYP). Users were scheduled for 2-hour sessions and compensated $100. Approval was obtained from the Institutional Review Board of Columbia University. Subjects were given a short survey about clinical experience, demographics, computer proficiency and use of social networking tools (Table 2).

Table 1. Research questions, data sources, criteria for answering the question affirmatively, and examples. T=thinkaloud; D=debriefing; S=statements

Research question (data source)	Evidence criterion or measure
1. Do users given an assembly of case data appraise it with sufficient independence to avoid making 'diagnosis momentum' errors? (T)	User's case 5 summary includes renal decline and recommendation for renal follow-up; request further information
2. Are there significant differences in accuracy (i.e. proportions of diagnosis elements stated by users compared with an expert reference standard) between MedWISE and WebCIS (conventional system)? (T)	Proportion of diagnostic elements stated by users as compared to expert-derived reference standard; T-test used to determine significant differences
3. What are user impressions and perceptions? (With regard to ease of use, usefulness) (T,D)	Thinkaloud statements, response to debriefing questions, Statements re: ease of use and usefulness, adoption, Likert scale in debriefing survey 0.326
4. Which features do users find useful and why? (T,S,D)	Statements re: ease of use/usefulness of specific feature; attempt feature use

Table 2. User demographics and EHR use (N=18)

	Average	Range
Service at NYP	2.5 yr	4mo-6.5yr
Work in field	3.3 yr	4mo-7 yr
WebCIS use	2.4 yr	4mo-6.5 yr
Eclipsys use	1.9 yr	4mo-4.5 yr
Other EHR/CIS use	2.7 yr	0-6.5 yr
Hours/week WebCIS	26.6 hr	8hr-80hr
Hours/week Eclipsys (commercial system at hospital)	25.8 hr	8hr-80hr

One subject rated himself 'expert' in computer knowledge, seven 'above average', and five 'average'. Nine used EHR/CIS from other locations, usually home.

Twelve users were given a short MedWISE training session. They were then given five real patient cases, asked (via oral and printed instructions) to assume that they would be taking over care of the patient, and to use MedWISE to familiarize themselves with the patient's condition and state their assessment, diagnoses, and plan. Data were recorded using Morae video-analytic software which permits detailed screen recording and analysis of the users' speech while carrying out the task, in a 'thinkaloud' protocol common for this type of study [11]. At the end of the five cases, clinicians rated their perceptions of system ease of use and usefulness on a 5-point Likert scale, (where 1=Very hard to use/Not useful and 5=Very easy to use/Very useful), as well as stating their general reactions and comments on specific features.

For the test of diagnosis momentum error the fifth case was prepared from a real-life case in which a patient coming into emergency with very serious emergent hypertension later died from an unnoticed renal failure, although the relevant data were available and clear. The case (with lab panels and notes) was presented to 10 users in printed form (so users could not examine other data) as it would appear in MedWISE.

For the accuracy test six other clinicians (5 medicine residents and one psychiatry resident) used the conventional system in use (WebCIS) to appraise three of the same cases, to serve as a control. Diagnoses and plans of both groups were compared with the reference standard. Differences in the average numbers of elements overlapping the

reference standard were calculated; this information can serve to find an effect size for future hypothesis testing in a larger study.

Think-aloud protocols from the users were transcribed and coded according to two coding schemata, for diagnostic reasoning [4] and human-computer interaction and user comments, derived from the literature on the intelligent use of space in workplaces [5]. Here we discuss only the codes pertaining to user approval/disapproval and usefulness of specific features. We summarized survey data using descriptive statistics.

2. Results

2.1. Diagnosis Momentum Test

In our single test case only one of ten subjects failed to detect the renal problem, unlike the real-life situation (from which the test case was derived) in which the patient died due to clinicians' failure to follow up independently.

Surprisingly, comments indicated subjects themselves were unconcerned about the possibility of increased errors and/or error propagation. They believe that clinicians are careful enough, that the current systems similarly depend on propagation of notes among colleagues, which presents the same issues, and that decreased copying, note splitting, and other features might decrease errors. They also stated that sharing might result in greater completeness and error catching due to the 'many eyes' involved, and pointed out the advantages of local group configuration and control.

2.2. Accuracy

Those using MedWISE reported slightly higher numbers of major diagnostic elements vis a vis the reference standard, compared to WebCIS users (averages 6.833 v. 6, 5.5 v. 5.333 and 4.4 v. 4.333 for cases 2, 3 and 4 respectively), but this difference was not statistically significant ($p>0.2$ in all cases). Numbers are too small to be conclusive but the trend is that accuracy was not affected.

2.3. User Comments

User comments were mostly positive and even enthusiastic, as they felt overall that the system would save time and frustration, and match their thinking.

On a 5-point Likert scale for ease of use average user rating was 3.79. On a 5-point scale for usefulness, average user rating was 4.0. Some stated that they would give a higher score if more information (e.g., vital signs, lab summaries) were available, or if minor bugs were fixed. Subjects expected to become more proficient with practice, saying there is a slight learning curve (a criticism in terms of web 2.0 'walk up and use' approaches but better than reactions to many conventional EHRs). 10/12 said it made their mental process easier. Some expressed interest in using it for different use cases e.g. clinic duty, or transplant.

Users had definite individual opinions about the usefulness of various features; the most liked and cited one being the ability to gather and view information together on a single screen. One emergent theme was enthusiasm, (expressed in comments such as

"oh, awesome", "save this tab and share it, that would be like, I mean that'd save 10 min". Another theme was engagement, expressed in behaviors such as spending much time exploring, stating opinions, experimenting, or speculating on the usefulness or potential tool creation for different contexts or problems they experience at work.

The ease of use appeared to allow progressive learning and gradual adoption. A timid user could first use others' shared tabs, then incorporate additional elements with a click, then create his/her own interfaces, then make custom panels, and so on. It was an explicit part of our model design that users can participate at the level with which they are comfortable, from none (using the interface like the legacy system, which our first subject did), to extensive creation, customization, and social networking.

2.4. Usefulness of specific features

Table 3 shows user approval of specific features.

Table 3. Counts of comments coded 'approval'; re: useful specific features

View items on same screen 19	Patient-specific displays 6	Multiple sources 4
Note writing 9	Custom lab summaries 4	Display as checklist 4
		Templates 5
Time savings 6	Quickly summarize 4	Sharing with colleagues 4

3. Discussion

The low proportion of users in which diagnosis momentum seemed to occur is certainly not conclusive; whether user-configured shared interfaces result in greater likelihood of error than current systems is a major concern regarding the whole philosophy of the project. Larger, precise laboratory studies and tests of propagation of potential errors (by having some users devise interfaces and others use them in cases), are needed to fully resolve this question. Likewise for concerns about consistency: if informal work practices are externalized in the EHR, greater standardization may result. Also, common domain knowledge may result in consistency of created items.

However, some issues about innovation and ethics are illustrated by a common reaction to this new approach. We believe that MedWISE is in equipoise with current practice, in that all information is available at any time, users decide what to view in both systems, and currently institutions do not monitor user viewing except for security purposes and do not warn clinicians about inadequate information viewing. Our small accuracy study also supports the null hypothesis that there is no significant difference in users' ability to reach valid conclusions using MedWISE compared to conventional systems. Why then, the insistence by some that potential errors may be a fatal flaw?

We believe it has to do with the nature of innovation. Healthcare is necessarily concerned with avoiding harm. Any innovation must demonstrate that in this respect it is comparable to current practice (i.e. equipoise). However new innovations may be more closely scrutinized, particularly for ways in which they do not compare to current practice. Missing an accustomed feature is taken more seriously than the possibility that current practice has liabilities compared to the innovation. These liabilities may be difficult to imagine, but as soon as the innovation is proven possible, the ethical question becomes which is preferable, based on the whole constellation of advantages and liabilities. A new capability may render current practice unethical if continued [12].

One method to ascertain the liabilities of current practice may be to imagine via a thought experiment that the innovation is the norm, and imagine changing to the current practice. This may be a useful exercise in teasing out the various opportunities and liabilities to help identify areas of risk in the current system, facilitating research design to test whether these risks in fact occur. Thus for the MedWISE case, one would compare the risks of having to retain information in working memory between screens, and search repeatedly, with the risks of item omission, and the errors possible with each set of features.

Our contention is that only empirical research can determine whether, why and how errors occur in novel systems, but that a rush to declare a novel system error-prone and current systems not, may sometimes be more a result of lack of examination/imagination than the result of careful examination of actual behavior.

4. Limitations

Limitations include the small numbers of cases and clinicians, the study at one medical center (but with data from two) and the fact that the controls for the accuracy study were conducted two years after the intervention study, but with exactly the same case information. The use of real patient cases with typical clinicians are strengths.

5. Conclusions

User composable approaches to the electronic health record present many potential benefits [2]. Our small studies show comparable assessment accuracy to use of the conventional system, low risk of diagnosis momentum error, and overall favorable and even enthusiastic user perceptions of individual features and the overall approach. Further work with larger numbers of clinicians and cases is needed to test the possibility of errors. We presented a possible method of identifying risks in current systems, which could be used to facilitate experiment design.

References

[1] Y. Senathirajah, S. Bakken, Important ingredients for health adaptive information systems, *Studies in health technology and informatics* **169** (2011), 280-284.
[2] Y. Senathirajah, D. Kaufman, S. Bakken, Cognitive Analysis of a Highly Configurable Web 2.0 EHR Interface, *AMIA Annual Symposium proceedings* (2010), 732-6.
[3] Y. Senathirajah, S. Bakken, Architectural and usability considerations in the development of a Web 2.0-based EHR, *Studies in health technology and informatics* **143** (2009), 315-21.
[4] E. Razzouk, T. Cohen, K. Almoosa, V. Patel, Approaching the limits of knowledge: the influence of priming on error detection in simulated clinical rounds, *AMIA Annual Symposium proceedings*, 2011, 1155-1164.
[5] V.L. Patel, T. Cohen, T. Murarka, J. Olsen, S. Kagita, S. Myneni, et al. Recovery at the edge of error: debunking the myth of the infallible expert, *Journal of Biomedical Informatics* **44** (2011), 413-424.
[6] J.W. Ely, M.L. Graber, P. Croskerry, Checklists to reduce diagnostic errors, *Academic medicine: Journal of the Association of American Medical Colleges* **86** (2011), 307-313.
[7] J. Groopman, *How Doctors Think*, Houghton Mifflin Harcourt, Bellmawr NJ, 2007.
[8] J. Kassirer, R.I. Kopelman, *Learning Clinical Reasoning,* Williams and Wilkins, Baltimore MD, 1991.
[9] G. Hripcsak, A. Wilcox, Reference standards, judges, and comparison subjects: roles for experts in evaluating system performance, *JAMIA* **9** (2002), 1-15.

[10] L. Grabenbauer, A. Skinner, J. Windle, EHR adoption: Maybe It's not about the money – Physician Super-users, Electronic Health Records and Patient Care, *Applied Clinical Informatics* **2** (2011), 460-471.
[11] A.W. Kushniruk, V.L. Patel, Cognitive and usability engineering methods for the evaluation of clinical information systems, *Journal of Biomedical Informatics* **37** (2004), 56-76.
[12] P.A. Ubel, R. Silbergleit, Behavioral equipoise: a way to resolve ethical stalemates in clinical research, *The American Journal of Bioethics* **11** (2011), 1-8.