Triangulating Methodologies from Software, Medicine and Human Factors Industries to Measure Usability and Clinical Efficacy of Medication Data Visualization in an Electronic Health Record System

Bora Chang¹, Manoj Kanagaraj¹, Ben Neely², Noa Segall, PhD¹, Erich Huang, MD-PhD¹,²
¹Duke University School of Medicine, ²Duke University Health System Division of Translational Bioinformatics, Durham, NC

Abstract
Within the last decade, use of Electronic Health Record (EHR) systems has become intimately integrated into healthcare practice in the United States. However, large gaps remain in the study of clinical usability and require rigorous and innovative approaches for testing usability principles. In this study, validated tools from the core functions that EHRs serve—software, medicine and human factors—were combined to holistically understand and objectively measure usability of medication data displays. The first phase of this study included 132 medical trainee participants who were randomized to one of two simulated EHR environments with either a medication list or a medication timeline visualization. Within these environments human-computer interaction metrics, clinical reasoning and situation awareness tests, and usability surveys captured their multi-faceted interactions. Results showed no statistically significant differences in the two displays from software and situation awareness perspectives, though there were higher statistically significant usability scores of the medication timeline (intervention) as compared to the medication list (control). This first phase of a novel design in triangulating methodologies revealed several limitations from which future experiments will be adjusted with hopes of yielding further insight and a generalizable testing platform for evolving EHR interfaces.

Background
In 2009, The Health Information Technology for Economic and Clinical Health (HITECH) Act passed as part of the American Recovery and Reinvestment Act to promote the adoption and meaningful use of health information technology. This law spurred the rapid adoption of electronic health record systems (EHRs) across the United States in a short period of time.¹ The National Center for Health Statistics (NCHS) found that the use of EHRs in hospitals, outpatient departments, and office-based practices had major increases in EHR adoption to levels of 80% usage nation-wide by 2013²,³ With considerable incentives for rapid implementation, meeting federal guidelines has been the priority for EHR vendors over provider and patient usability.⁴ With such precipitous EHR adoption come significant changes in the clinical workflow and the nature of medical practice. The increased frequency with which clinicians must interface with EHRs to provide patient care underscores the need for their use to be efficient and effective. According to the Healthcare Information and Management Systems Society (HIMSS) Task Force on Usability, the usability of EHRs has a direct relationship with clinical productivity, error rate, user fatigue and user satisfaction, all of which affect patient safety and health outcomes. To evaluate the usability of a tool requires measuring user efficiency, effectiveness, cognitive load and ease of learning.⁵ The study of usability related to EHRs is a relatively new field. Thus, relevant medical literature is mostly limited to the past decade and a half. Recent systematic reviews [Harrington (2011), Clarke (2013) and Zahabi (2015)] of the literature on EHR usability and safety cover 91 publications between 2000 to 2015.⁶,⁷ Though studies included a range of methods, including heuristic design assessment, questionnaires, semi-structured and in-depth interviews, observations and task and behavior pattern analyses, Harrington (2011) and Zahabi (2015) explicitly describe the “limitations of the existing body of research identifying EMR design problems, including the application of various methods of usability and safety analysis” in which “the majority of studies of EMR usability issues have employed descriptive or qualitative analysis methods.”⁶,⁷ Additionally, they point out that much of the current literature identifies EHR usability issues using comparisons between paper-based systems and EHRs, but few studies include comparisons among multiple EHR designs.⁶,⁷ There is a paucity in research on EHR usability in quantitative methodologies that capture a complete understanding of active user-application interaction and the downstream, clinical impact of EHR designs.

Precedents for quantifiably measuring software efficacy with proven validity can be found within the software industry in the form of controlled experiments. Controlled experiments are designed to establish causal relationships with high probability between certain variables. These experiments have been described in scientific literature since the early 20th century and have been well developed in the statistics community.⁸ However, it was not until the 1990s, with the growth of the Internet that technology giants like Amazon, Facebook, Google, and Yahoo! built a
culture of data-driven user interface design and enhancements. Industry leaders like Google and Microsoft have created a software development infrastructure around controlled experiments, or software A/B testing, at its core and have propelled this method as a standard of development in the industry.

Based on the need for a robust quantifiable system of measurement coupled with clinical objectives, we sought to develop a framework of integrating methodologies of software A/B testing with simulated clinical decision measurements. Integrated into an infrastructure of a controlled experiment, we have built simulated electronic health record environments with two different designs that present medication history, alongside a clinical questionnaire module that prompts users to interact with the information presented in the simulated EHRs. This questionnaire seeks to measure degrees of clinical understanding and reasoning based on the way information is presented. We have utilized human factors and usability principles as defined in heuristic evaluation methodology to address several layers of reasoning. We hypothesized that interactive software designs developed with a focus on user-experience will aid in efficient visualization of clinical information, affect understanding of patient history, improve clinical management, and ultimately provide patients with optimized medical care. Furthermore, this novel triangulated approach to comprehensively understand and measure user interactions with health informatics technologies provides a rigorous and quantifiable framework that can be agnostically applied to a variety of clinical decision support tools that may be developed in the future.

Methods

Study Design

This study was a randomized, controlled trial aimed to quantifiably measure the impact of a medication timeline visualization on usability and clinical reasoning. Simulated electronic medical record environments, built as web applications, hosted the control and intervention variables (Figure 1 & 2). Users were randomized to one of the two simulated EHR environments, provided a clinical scenario, then asked to interact with the EHR environment by answering a uniform set of clinical questions based on the medication information in the simulated environment. This clinical questionnaire and simulated EHR environment is referred to as the experimental “module” in this paper. The user interactions with the medication application were tracked by a web analytics platform called Optimizely, which provides robust A/B, multivariate, multi-page testing capabilities to improve user engagement. The objective metrics that Optimizely tracked in this study were: (1) the number of unique, engaged users and (2) the number of clicks inside the simulated EHR environment. After the module was completed, users were asked to complete a multi-part survey, which reflected user opinion of the usability of the medication applications presented in the simulated EHR environments.

Methods

Study Design

This study was a randomized, controlled trial aimed to quantifiably measure the impact of a medication timeline visualization on usability and clinical reasoning. Simulated electronic medical record environments, built as web applications, hosted the control and intervention variables (Figure 1 & 2). Users were randomized to one of the two simulated EHR environments, provided a clinical scenario, then asked to interact with the EHR environment by answering a uniform set of clinical questions based on the medication information in the simulated environment. This clinical questionnaire and simulated EHR environment is referred to as the experimental “module” in this paper. The user interactions with the medication application were tracked by a web analytics platform called Optimizely, which provides robust A/B, multivariate, multi-page testing capabilities to improve user engagement. The objective metrics that Optimizely tracked in this study were: (1) the number of unique, engaged users and (2) the number of clicks inside the simulated EHR environment.

After the module was completed, users were asked to complete a multi-part survey, which reflected user opinion of the usability of the medication applications presented in the simulated EHR environments.

Figure 1. Simulated EHR with Control Medication List

![Simulated EHR with Control Medication List](image-url)
Figure 2. Simulated EHR with Intervention Medication Timeline

Setting and Eligibility
We obtained approval for this study through the Duke University Health System Institutional Review Board. We were granted a request to waive consent from users since there was no collection of sensitive personal health information from study subjects. Participants included medical trainees who had completed the Duke Health System’s EHR training module that is mandatory before engaging in direct patient care. In the first phase of this experiment these medical trainees were in the following programs: Duke Physician Assistant (PA) Program and Duke University School of Medicine. The training levels of students included second year PA students and medical school students in their second year of training and above. The second phase of the study will be completed with GME Duke residents in the future. This paper only discusses the results from the first phase of this experiment.

Control and Intervention
The simulated EHR environment that acted as the control in this experiment was modeled after the medication list that currently exists in the major EHRs in healthcare. Currently, much of the medication information for patients exists as a list of medication names, start and end dates, instructions for taking the medications and the providers’ names that prescribed the medications. Our control web page looks similar to this model, portraying a list of a patient’s medications and their corresponding details (Figure 1). The intervention was a simulated EHR environment with a visualization of the same medication information in an interactive, chronological timeline – one that the user can zoom in and out to see various timeframes, with a representative “time bar” at the bottom of the screen to give users easy control to define the time period of interest (Figure 2). This was modeled similarly to the Belden, Plaisant, et al (2015) “Inspired EHR”, medication list. In addition to visualizing this information with a historical and chronological lens, the intervention included filters that simplified the information. The intervention medication application included a filter for conditions that allowed users to filter the list by the condition for which a medication was prescribed. Another sorting feature was filtering by drug type. Lastly, an important visual component was the color of the timeline bars which indicated two concepts: (1) the gradation of color communicated a relative change in dose of medication, with darker shades signaling increases in dose and lighter shades signaling decreases in dose, (2) the purple and orange colors denoted pharmacy “fill status” of the medication, indicating that the medication was purchased at the pharmacy by the patient; purple indicated a filled prescription and orange indicated a non-filled prescription. For both simulated EHR settings, users could obtain detailed information by double-clicking on the medication row in the control or the medication timeline bar in the intervention. This function would open a pop-up window with more detailed information about the prescription such as the name of the provider, dosing details, directions for medication intake, number of refills and pharmacy fill status (Figure 3).
Software A/B Testing and Objective Metrics

Prior to the experiment we determined several “Overall Evaluation Criteria” (OEC) as referred to in the software industry, or metrics of defined “success” that we wished to track in the study. We captured these objective metrics by A/B testing software, Optimizely, while users interacted with the module. The metrics we selected were: (1) the total number of engaged, unique users, defined by the Optimizely platform as users who clicked on an element of the website and tracked by a cookie unique to their devices, and (2) the total number of clicks within the medication information application.

Clinical Questionnaire

The clinical questionnaire was developed to elucidate the clinical reasoning and decision-making process in the simulated EHR environment. When developing the questions and answers, we followed the principles of situational awareness assessment. Endsley, et al (2003) define three levels of situation awareness: (1) perception of the elements in the environment, (2) comprehension of the current situation, and (3) projection of future status (Table 1). The first tier, perception, entails perceiving status, attributes, and dynamics of relevant elements in the environment. In our simulation, users perceived various data elements such as medication names, start and end dates for prescriptions, colors of fill status, etc. The second tier, comprehension, involves understanding what the data and cues perceived mean in relation to relevant goals and objectives. In our simulation, users were asked to synthesize disjointed data elements to answer questions related to ultimate treatment goals. Lastly, the projection tier demands the prediction of what those data elements from the first two layers will do in the short-term future. A person can only achieve this third level of situation awareness by having a solid understanding of the situation and the dynamics of the system within which she or he is working. In our simulation, users were asked to project what treatment option would be best given their understanding of the patient’s medication history. These three layers of situation awareness integrate perception of knowledge and the ability to prioritize and synthesize information. These are critical for efficient decision-making and effective action.

The clinical scenario was based on a fictitious patient whose situation mirrors a typical American patient in the fourth quartile of life with several chronic conditions that require numerous medications for treatment. Answers to clinical questions were based on standard of care guidelines for the pertinent clinical management of diabetes mellitus type 2, hypertension and unipolar depression. The questionnaire was an embedded Qualtrics survey, a survey software platform, within the simulated EHR web environment. Qualtrics captured all questionnaire data in addition to the total time spent on the survey. The scoring of the clinical questionnaire are as follows: a total score (a measure of correctly answered questions), a tier 1 score (subset of the overall score including only perception questions), a tier 2 score (subset of the overall score including comprehension questions), a tier 3 score (subset of the overall score including projection questions), and a total time-to-complete questionnaire metric.
Clinical Vignette

Meet Mr. Smith. He is a 62 year old gentleman living in Durham, NC with his family. He has been seeing Dr. Bradley as his primary care provider for over 20 years. After a long, prolific career, Dr. Bradley, who is one of your partners in the practice, is retiring. You are now adopting Mr. Smith as a new patient.

His past medical history includes:

- Obesity
- Hypertension
- Type 2 Diabetes
- Gout
- Depression
- Community-Acquired Pneumonia

He has come in today for three reasons. He would like to:

1) Discuss his most recent HbA1c of 8.4% despite taking his diabetes medications
2) Discuss his last elevated blood pressure readings despite changes in his medications
3) Discuss his depressed mood and decreased interest in activities recently

Your objectives:

1) Concerning his diabetes, you can trust that he has been on a diabetic diet with regular exercise. You would like to assess his past and current regimen of medications for his Type 2 Diabetes and see if an additional agent is necessary to prescribe today.
2) For his hypertension, despite taking his new hypertensive medication diligently for the last year, his hypertension has not changed. Trends of blood pressure readings in the last year and a half have shown >145/95 (JNC 8 recommends treatment for >140/90). He has been trialed on several different types of medications. You would like to assess what additional changes in his regimen may be required today.
3) For his depression, after assessing any changes that may have caused an increase in depressive symptoms, you would like to assess his current depression medications and understand his treatment course up until now.

** Of note: For the purposes of the experiment, FILL STATUS will imply compliance to medications.**

### Table 1. Clinical Questionnaire Based on Situation Awareness

| Clinical Vignette | Tier 1 - Perception Example | Tier 2 - Comprehension Example | Tier 3 - Projection Example |
|-------------------|-----------------------------|-------------------------------|-----------------------------|
| Meet Mr. Smith. He is a 62 year old gentleman living in Durham, NC with his family. He has been seeing Dr. Bradley as his primary care provider for over 20 years. After a long, prolific career, Dr. Bradley, who is one of your partners in the practice, is retiring. You are now adopting Mr. Smith as a new patient. | Which medication(s) is the patient currently taking for diabetes? | Over time, Mr. Smith’s blood pressure still continued to increase despite his compliance with his hydrochlorothiazide (HCTZ). Another medication was added to his hypertension medication regimen. Which new class of hypertensive medication was added? | Why do you think the provider chose bupropion to prescribe in 2006? |
| His past medical history includes: | a. metformin | a. Calcium channel blockers | a. The patient may have been experiencing side effects of Sertraline, therefore switched to Bupropion as monotherapy instead. |
| % Obesity % Hypertension % Type 2 Diabetes % Gout % Depression % Community-Acquired Pneumonia | b. insulin | b. ACE inhibitors | b. Bupropion is in the SSRI drug class. The provider may have desired to try another SSRI agent. |
| He has come in today for three reasons. He would like to: | c. lisinopril | c. Alpha agonist | c. Sertraline was used as a monotherapy and was escalated in dose. It may not have been effective at the high dose, and therefore another agent, Bupropion was added. |
| 1) Discuss his most recent HbA1c of 8.4% despite taking his diabetes medications | d. glipizide | d. Thiazides | d. Bupropion was effective for this patient when it was used 10 years ago. |
| 2) Discuss his last elevated blood pressure readings despite changes in his medications | e. amlodipine | | |
| 3) Discuss his depressed mood and decreased interest in activities recently | | |

** Usability Questionnaire

The usability survey was based on several validated tools in the human factors field. The survey addressed three major components: (1) ease of use (2) cognitive workload (3) user satisfaction and perception of impact. For measuring ease of use, we tailored the System Usability Survey (SUS), which has been used to evaluate a wide variety of products and services, including hardware, software, mobile devices, websites and applications. The SUS has become an industry standard, with references in over 1300 articles and publications. The benefits of using SUS are its ease of administration to participants (short length), its ability to deliver reliable results even with small sample sizes, and its proven validity21,22 (Figure 5). One minor change in the survey was the substitution of ‘system’ with ‘application’ for the sake of language consistency used throughout the study. To assess cognitive workload, we used the NASA Task Load Index (NASA-TLX). It includes six subscales that represent somewhat independent clusters of variables: Mental, Physical, and Temporal Demands, Frustration, Effort, and Performance. Since its initial development over 20 years ago, the NASA-TLX has been cited by over 5000 articles and has been used in numerous industries23,24 (Figure 6). To assess users’ satisfaction and perception of impact of the application, we asked three questions on their perception of the application’s effect on their cognition, clinical management and projected patient outcomes from those decisions and actions (Figure 7). Finally, there was a free text space provided for any comments, suggestions, or questions. All answer choices were on a Likert scale of one to five ranging from strongly disagree to strongly agree for ease of use (SUS) and user satisfaction, and ranging from very low to very high for cognitive workload (NASA-TLX). The published scoring rubrics for SUS and NASA-TLX were used to score these data21-24 A modification of the NASA-TLX involved weighting each question equally, and scoring answers from one to five based on whether the question was “positive” or “negative.”23,25
### Figure 5. System Usability Scale (SUS)

| Statement                                                                 | Strongly Disagree | Somewhat Disagree | Neither | Somewhat Agree | Strongly Agree |
|---------------------------------------------------------------------------|-------------------|-------------------|---------|----------------|----------------|
| I think that I would like to use this application frequently.             | ○                 | ○                 | ○       | ○              | ○              |
| I found the application unnecessarily complex.                            | ○                 | ○                 | ○       | ○              | ○              |
| I thought the application was easy to use.                                | ○                 | ○                 | ○       | ○              | ○              |
| I think that I would need the support of a technical person to be able to use this application. | ○                 | ○                 | ○       | ○              | ○              |
| I found the various functions in the application were well integrated.    | ○                 | ○                 | ○       | ○              | ○              |
| I thought there was too much inconsistency in this application.           | ○                 | ○                 | ○       | ○              | ○              |
| I would imagine that most people would learn to use this application very quickly. | ○                 | ○                 | ○       | ○              | ○              |
| I found the application very cumbersome to use.                          | ○                 | ○                 | ○       | ○              | ○              |
| I felt very confident using the application.                             | ○                 | ○                 | ○       | ○              | ○              |
| I needed to learn a lot of things before I could get going with this application. | ○                 | ○                 | ○       | ○              | ○              |

### Figure 6. NASA - Task Load Index (NASA-TLX)

| Question                                                                 | Very Low | Somewhat Low | Neutral | Somewhat High | Very High |
|--------------------------------------------------------------------------|----------|--------------|---------|---------------|-----------|
| How mentally demanding was the task?                                     | ○        | ○            | ○       | ○             | ○         |
| How much time pressure did you feel?                                     | ○        | ○            | ○       | ○             | ○         |
| 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?    | ○        | ○            | ○       | ○             | ○         |

### Figure 7. User Satisfaction Questions

| Question                                                                 | Strongly Disagree | Somewhat Disagree | Neither | Somewhat Agree | Strongly Agree |
|--------------------------------------------------------------------------|-------------------|-------------------|---------|----------------|----------------|
| I believe this medication display would help improve my ability to manage patients’ medications. | ○                 | ○                 | ○       | ○              | ○              |
| I believe this medication display would aid me in providing a better treatment regimen for my patients. | ○                 | ○                 | ○       | ○              | ○              |
| I believe this medication display would improve patient outcomes.        | ○                 | ○                 | ○       | ○              | ○              |

### Analysis

In this randomized experiment, 132 participants who completed the module were randomized to a standard and an innovative EHR environment. Testing the superiority of the two arms was accomplished with a two-sided t-test for the average of two independent samples. The significance level (alpha) was set at 0.05. Each data collection instrument captured different, yet important data, from which various metrics were derived. All of these metrics were considered equal and no primary analysis was defined *a priori*.

### Results

**Demographics**

Of the 132 participants in Phase 1 who completed the entirety of the module, 72% (95 participants) were second to fifth year medical students and 28% (37 participants) were second-year physician assistant students. There are fluctuations in the total number of participants in the various portions of the study module. This limitation is discussed in the following section. All participants received EHR training through modules set by the Duke University Health System as part of patient safety certification before engaging in clinical duties. All participants had at least 10 months of EHR experience at the time this study was conducted.

The results of the control and intervention arms from all three areas of methodology (software, clinical and situation awareness, and human factors) are included in Table 2. There is a difference in total number of recorded engaged users as measured by Optimizely and the total number of recorded participants for the clinical questionnaire and usability surveys. This difference is attributed to the attrition of users who engaged with the module but did not complete the entirety of the module. These results show several noteworthy items: 1) The intervention showed more than double the click rate per user than the control; 2) there was an insignificant difference between both arms in clinical questionnaire scores or timing, however, there was 3) a significant difference showing better usability for the intervention than the control as measured by SUS and user satisfaction surveys (p-value 0.02 and 0.00014 respectively).
Table 2. Control and Intervention Data from Triangulation of Methodologies

|                                      | Control                  | Intervention              | P-value (<0.05) |
|--------------------------------------|--------------------------|---------------------------|----------------|
| (1) Software Analytics – Optimizely   |                          |                           |                |
| Engaged users                        | 64                       | 78                        |                |
| Average Clicks Per User              | 32                       | 69                        |                |
| (2) Clinical Reasoning & Situational Awareness - Clinical Questionnaire |                          |                           |                |
| Participants                         | 65                       | 68                        |                |
| Total Average Score                  | 20.4 out of 24 (SD 4.6)   | 19.0 out of 24 (SD 4.6)   | 0.09           |
| Tier 1 Score (Perception)            | 5.6 out of 6 (SD 1.18)    | 5.37 out of 6 (SD 1.19)   | 0.27           |
| Tier 2 Score (Comprehension)         | 10.95 out of 11 (SD 2.67) | 10.06 out of 11 (SD 2.71) | 0.06           |
| Tier 3 Score (Projection)            | 3.8 out of 5 (SD 1.17)    | 3.38 out of 5 (SD 1.28)   | 0.31           |
| Average Total Duration               | 12.77 min (SD 5.04)       | 13.59 min (SD 5.68)       | 0.39           |
| (3) Human Factors - System Usability Scale (SUS), NASA - Task Load Index (TLX), User Satisfaction |                          |                           |                |
| SUS                                  | 43.9 out of 50 (SD 12.7)  | 49.1 out of 50 (SD 13.0)  | 0.02           |
| NASA – TLX                           | 72.4 out of 100 (SD 14.4) | 76.2 out of 100 (SD 14.9) | 0.14           |
| User Satisfaction                    | 10.73 out of 15 (SD 3.13) | 12.28 out of 15 (SD 2.25) | 0.0014         |

User Comments and Suggestions

Medication List (Control)

"This application is significantly improved over the current medication display in our EHR. One issue I ran into is that a lot of the questions were about drugs prescribed for particular conditions, and there was no clear way to identify an indication within this display. I would also like it if the current drugs were more clearly identified, as by a different color, or something like that."

"It would be better if medications were categorized according to which chronic condition they were prescribed for. You could click on the chronic condition, and then see the medications that had been prescribed."

"A few more columns with additional information available would be nice to avoid having to click on small rows to get a specific medication. Also, it might be nice to have prescriptions ranging from most recent to oldest. This would alleviate having to repeatedly scroll up and down a long list and look closely at the dates to see which one you want to click on. Finally, some additional drug information would be nice - maybe a free text box for why it was prescribed, dose changed, etc."

Medication Timeline (Intervention)

"Really fantastically designed and easy to use MAR. However, to be truly useful, it would require patients to have been part of a single EMR (or a linked EMR) for the duration of their treatments, which often is not the case. I believe that a situation similar to the case presented here, in which a patient was followed for 20-years, would be extremely rare."

"I liked how the medication information was split up by class and condition. This made it easier to track medication changes. While I liked being able to see all the medication across their lifespan of being prescribed, I didn't like how I had to click on the bar to get more information about the prescription. I also didn't like how I had to zoom in/out in order to see the information I wanted."

"[The intervention] was extremely intuitive and easy to use. I watched the video tutorial, but think that I could've easily figured it out if I was asked to use it without prior instruction. I want this program in my life now!!!"
Discussion

Software Analytics and Usability

The fields of software analytics and human-computer interaction generally correlate a high number of clicks with increased user workload. In this experiment, Optimizely data demonstrated more than double the number of clicks in absolute total and per user number of clicks in the intervention compared to the control. From the conventional perspective, this could be interpreted as the intervention visualization being more “burdensome” on users. However, the features of the medication timeline, which are interactive by nature, include an array of additional functionality, which require users to interact with the information (zoom, drag and drop, filter, set scope of timeframe). This is in contrast with the medication list, which has a single function of scrolling vertically up and down to see a full list of medications. Other than the granular window feature (Figure 3), the two arms of the study had drastically different presentations of medication data. In combining this insight with the usability survey results, which indicate a higher statistically significant score for the medication timeline in its ease of use and user satisfaction, it is clear that despite the larger number of clicks for the intervention, its reported usability was greater. The NASA-TLX scores show a higher average score for the intervention as well though this is not statistically significant. In sampling user comments and suggestions, there was a greater number of comments overall for the control medication list (30 comments) as compared to those received for the intervention medication timeline (17 comments). Though limited by subjective assessment, several of the comments for the medication list involved suggestions that were being executed in the medication timeline such as filters for conditions, current medication list, color-coded fill and unfilled prescription data. Contrarily, many comments in the intervention were positive reactions to the new interface with adjustments for optimized ergonomic design that would be helpful in the future.

Clinical Reasoning and Total Duration

The primary hypothesis was that the medication timeline would be more effective in aiding users to answer the questions in all three tiers of situation awareness with higher accuracy. Interestingly, there is no statistically significant difference in the clinical questionnaire scores between control and intervention. This holds true for the total average score as well as the scores by situation awareness tier. Users generally scored well on the clinical questionnaire with average scores of 80-85%. This indicated that on all levels, despite the type of program used to obtain medication data, medical trainees are able to find the information, interpret it and project clinical reasoning based on the information to a high degree of accuracy. We wanted to explore whether the medication list or timeline formats varied with respect to total time spent on the clinical module. We hypothesized that participants would require more time to complete the clinical module using the control interface than when using the intervention. Again, time to complete the module was similar across interfaces. These findings indicate that, when given a generous time limit (30 minutes), all questions can be answered accurately in both the intervention and control formats. There may be a “ceiling” effect where users, when given ample time, are able to overcome the difficulty of interacting with the control interface and correctly respond to clinical questions.

Limitations

Completion Rates - While the web application distribution of the clinical study allowed us to capture an overall completion rate of approximately 132 participants as defined by those who completed the last step of the study, the major limitation of such a distribution was the inconsistency in completion rates of the entire module, where there was a greater attrition rate as the module progressed.

Optimizely Set-up - When comparing the number of clicks between the control and intervention arms, due to the inherent difference in how the information is presented, there is a stark contrast in the way users interact with the medication list versus the medication timeline. Because of this difference, the objective metric of total number of clicks was not a fair metric of comparison between the two arms of the study.

Clinical Reasoning and Time Pressure - There is a limit to the difficulty of a question that can be asked in the clinical questionnaire before it verges on relying on medical expertise. The purpose of the clinical questionnaire and three tiers of reasoning was to ask clinical questions typical of those a primary care physician asks when attempting to understand a patient’s history via the EHR. However, as demonstrated in this experiment, the ample time given to users allowed them to accurately answer many of these questions in both arms of the study and did not create the common “time pressure” experienced by physicians in a real clinical setting.

Adjustment to New Models of Visualization - There is yet a question of how much exposure to or practice with a new application is needed before a user is well adjusted. As with any new skill, with learning a new application,
there is a learning curve: users first recognize the features of the application, learn to effectively use them, and then practice to use them efficiently. As one user in the medication timeline arm of the study noted: “I got more used to it while using it just in the 20 or so minutes.” This is a significant variable that distinguishes the intervention from the control. The control medication list is familiar to users who have been trained in a "typical" EHR with a similar medication data presentation. There is not an interactive visualization in a typical EHR system that includes the types of functionalities (zoom, filtering, setting timeline parameters) presented in our intervention. Hence, the learning curve of first exposure may have been a key limitation in accurately measuring user interaction with the intervention compared to the control.

Future directions
In order to address the limitations outlined in the previous section, we will adjust the design of the second phase of this experiment. The second cohort will include medical trainees in the Duke Graduate Medical Education (GME) Internal Medicine and Psychiatry Residencies.

Unique User Tracking and Research Subject Anonymity - The inconsistent completion rates were one limitation of the web distribution of the study module. There was no mechanism to track unique user data from one portion of the study (ie. the clinical questionnaire) to the data of the usability. Hence, data analysis in Phase 1 was limited to the aggregate averages with varying total number of participants per section. In Phase 2 of this study, in order to link and track unique users who complete the entire module from beginning to end, while still maintaining anonymity by keeping the contact information collected for award purposes, we will use the methodology of “piping text” to carry over unique response IDs through the survey collection system (Qualtrics). This will allow us to tie clinical reasoning data to usability data in a 1:1 ratio for rigorous integrity in the data analysis of these surveys. In addition, the feature to “prevent ballot box stuffing,” or to allow one survey attempt per user will be enabled via Qualtrics.

Optimizely Set-up for Fair Comparison - Because of the explicit differences in the interaction required for the medication list versus the timeline, in order to have a fair comparison of the control and intervention for the metric of the number of clicks, we devised a “goal” measurement on a feature that is consistent between the two arms. The granular window that pops-up when users double-click on the medication list item or bar contains identical information. The act of “mining” for more granular data is parallel in the control and intervention. Hence, we will add a click tracker to the “close” button on the window of both arms via Optimizely. This will be a proxy to how many times users were clicking to access more detailed data in the control versus the intervention.

Clinical Reasoning and Simulated Clinical Time Pressure - The results of the clinical questionnaire showed a high percentage of accuracy on the clinical questionnaire in both arms with no significant difference between their average scores. The average amount of time spent on the module was approximately 13 minutes for both arms, with a range of 1 minute to 30 minutes. A majority of the users spent between 10-15 minutes on the clinical module portion, which demonstrates the approximate amount of time necessary to thoroughly go through the clinical vignette, interact with the medication information and answer the clinical questions. Based on this time average from phase 1 of the study, in the proceeding phase we will create a time pressure that gives users a cut-off time limit of 8 minutes which will simulate the typical time pressure felt by clinicians in situations of patient care. We believe this type of pressure will allow us to spread the user accuracy curve and hypothesize that we will see a significant difference in accuracy between the control and intervention arms when under such pressure to quickly analyze and mine information.

Learning Curve of a New Visualization - In order to address the learning curve of a first time exposure, we are setting up a “playground” in which users will be able view their respective, randomized video tutorial of how to navigate the simulated EHR and medication information, then link to their respective practice simulated environment to allow users to gain familiarity with the type of visualization and its features before beginning the clinical module. This playground will be limited to 2 minutes on each arm of the study.

As we move to the second phase of this study, with the improvements listed above, we note that this is the first report of a platform for objectively measuring EHR user interfaces and experience using a triangulated approach for measuring objective and subjective data. Our pilot experiment suggests that, with further refinements, this is a viable and scalable tool with which to move EHRs and their associated workflows to more user-friendly and efficient experiences.
References
1. Mandl, K. D., Khorasani, R., & Kohane, I. S. (2012). Meaningful use of electronic health records. Health Affairs, 31(6), 1365.
2. Jamoom E, Hing E. (2015). Progress with electronic health record adoption among emergency and outpatient departments: United States, 2006–2011. NCHS data brief, no 187. Hyattsville, MD: National Center for Health Statistics. 2015.
3. Hsiao C-J, Hing E. (2014). Use and characteristics of electronic health record systems among office-based physician practices: United States, 2001–2013. NCHS data brief, no 143. Hyattsville, MD: National Center for Health Statistics. 2014.
4. Wachter R. (2015). The Digital Doctor: Hope, Hype, and Harm at the Dawn of Medicine's Computer Age. McGraw Hill Professional; 2015.
5. Belden J, Grayson R, Barnes, J. (2009). Defining and Testing EMR Usability: Principles and Proposed Methods of EMR Usability Evaluation and Rating. https://www.himss.org/sites/himssorg/files/HIMSSOrg/Content/files/HIMSS_DefiningandTestingEMRUsability.pdf. June 2009.
6. Harrington L, Kennerly D, Johnson C. (2009). Safety issues related to the electronic medical record (EMR): synthesis of the literature from the last decade, 2000-2009. J Healthc Manag. 2011;56(1):31-43.
7. Clarke, M, Steege, L, Moore, J, Belden, J, Koopman, R, Kim, M. Addressing Human Computer Interaction Issues of Electronic Health Record in Clinical Encounters. Design, User Experience, and Usability: User Experience Design for Everyday Life Applications and Services, Third International Conference, DUXU 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings. Springer; 2014.
8. Zahabi, M, Kaber, D, Swangnetr, M. (2015). Usability and Safety in Electronic Medical Records Interface Design: A Review of Recent Literatures and Guideline. Human Factors: The Journal of the Human Factors and Ergonomics Society, 57(5), 805-834.
9. Box, G, Hunter, J, William G. (2005). Statistics for Experimenters: Design, Innovation, and Discovery. John Wiley & Sons, Inc.
10. Kohavi, R. (2012). Online controlled experiments. Proceedings of the Sixth ACM Conference on Recommender Systems - RecSys '12.
11. Tang, D, Agarwal, A, O'Brien, D, Meyer, M. (2010). Overlapping experiment infrastructure. Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '10.
12. Kohavi, R, Longbotham, R, Sommerfield, D, Henne, R. (2008). Controlled experiments on the web: Survey and practical guide. Data Mining and Knowledge Discovery Data Min Knowl Disc, 18(1), 140-181.
13. Dumas, J, Salzman, M. (2006). Usability Assessment Methods. Reviews of Human Factors and Ergonomics, 2(1), 109-140.
14. Optimizely: Optimizely: Let’s optimize digital experiences for your customers. (2016). Retrieved May 16, 2016, from https://www.optimizely.com/.
15. Belden, J, Plaisant, C, Johnson, T. (2014). Inspired EHRs: Designing for Clinicians. (2014). Retrieved May 16, 2016, from http://inspiredehrs.org/.
16. Roy, R. K. (2001). Design of experiments using the Taguchi approach: 16 steps to product and process improvement. New York: Wiley.
17. Endsley MR, Bolte B, Jones DG. (2003). Designing for situation awareness: An approach to user-centered design. Boca Raton, FL: CRC Press; 2003.
18. McCulloch, D, Nathan, D, Mulder, J. (2015). Initial management of blood glucose in adults with type 2 diabetes mellitus. Retrieved May 16, 2016, from http://www.uptodate.com/contents/initial-management-of-blood-glucose-in-adults-with-type-2-diabetes-mellitus.
19. Mann, J. (2015). Choice of drug therapy in primary (essential) hypertension. Retrieved May 16, 2016, from http://www.uptodate.com/contents/choice-of-drug-therapy-in-primary-essential-hypertension.
20. Simon, G, Ciechanowski, P. (2015). Unipolar major depression in adults: Choosing initial treatment. Retrieved May 16, 2016, from http://www.uptodate.com/contents/unipolar-major-depression-in-adults-choosing-initial-treatment.
21. Brooke, J. (1986). SUS - A quick and dirty usability scale. Retrieved May 16, 2016, from http://www.usabilitynet.org/trump/documents/Suschapt.doc.
22. System Usability Scale (SUS). (2016). Retrieved May 16, 2016, from http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html.
23. Hart, S, Staveland, L. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) Human Mental Workload. Amsterdam: North Holland Press.
24. Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. PsyCETRA Dataset.
25. Geddie JC, Boer LC, Edwards RJ, et al. NATO Guidelines on Human Engineering Testing and Evaluation. Neuilly-sur-Seine Cedex: NATO; 2001.