Using High-Fidelity Simulation and Eye Tracking to Characterize EHR Workflow Patterns among Hospital Physicians

Julie W. Doberne, BS1, Ze He, MS2, Vishnu Mohan, MD, MBI1, Jeffrey A. Gold, MD1,3, Jenna Marquard, PhD2, Michael F. Chiang, MD1,4
Departments of 1Medical Informatics & Clinical Epidemiology, 3Pulmonology & Critical Care, and 4Ophthalmology, Oregon Health & Science University, Portland, OR; 2University of Massachusetts Amherst, Amherst, MA

Abstract
Modern EHR systems are complex, and end-user behavior and training are highly variable. The need for clinicians to access key clinical data is a critical patient safety issue. This study used a mixed methods approach employing a high-fidelity EHR simulation environment, eye and screen tracking, surveys, and semi-structured interviews to characterize typical EHR usage by hospital physicians (hospitalists) as they encounter a new patient. The main findings were: 1) There were strong similarities across the groups in the information types the physicians looked at most frequently, 2) While there was no overall difference in case duration between the groups, we observed two distinct workflow types between the groups with respect to gathering information in the EHR and creating a note, and 3) A majority of the case time was devoted to note composition in both groups. This has implications for EHR interface design and raises further questions about what individual user workflows exist in the EHR.

Introduction
Electronic health records (EHRs) have become a central component of the modern clinical workflow, serving as a central documentation repository, an ordering mechanism, and a provider communication tool. EHRs have been promoted as a mechanism for improving quality of healthcare delivery, patient safety and provider efficiency. Widespread adoption has been driven in part by substantial governmental incentives through the Centers for Medicare and Medicaid Services (CMS) Meaningful Use program, with 59% of U.S. hospitals and 48% of office-based providers using EHRs as of 2014.

Implementations of EHRs have been shown to dramatically influence clinical workflows. End-user behaviors and training approaches are highly variable. Adaptive end-user behaviors such as excessive use of copy-paste/copy forward, “overdocumentation”, and “upcoding” may compromise health care quality and patient safety; training and standardization may help reduce these practices. Beginning efforts have been made to establish standard practices among EHR end users. However, these efforts have largely focused on documentation and not on information review.

The challenges facing consistency in EHR training and use are diverse. Though back end databases are fairly consistent across instances of an EHR, user interfaces and workflows can be substantially different depending upon the institution and clinical environment in which the EHR is used. Within an institution or practice group, the physician-level characteristics in usage of EHR features and usage intensity have been found to be highly variable and personalized. The strong influence of personal experiences and preferences is thought to partly explain this variance.

Assessment of end-user EHR behaviors has often been conducted via self-reports and surveys, direct observation, and meaningful use measure reporting. Survey and reporting methods can provide a high-level perspective of provider behavior, but do not capture individual workflows. Direct observation may lack the accuracy required to quantitatively evaluate differences between individuals or groups of individuals. Real-time eye tracking technology has been shown to successfully capture user behavior in online website searching, website interface design, visual attention and video games. Within the realm of medicine, it has been employed to study radiologic and electrocardiography interpretation, note reading, and medication administration. The purpose of this study is to address this gap in knowledge by characterizing the workflow patterns of physicians using the EHR. This is done using a mixed methods approach employing a high-fidelity EHR simulation environment equipped with eye and screen tracking, surveys, and semi-structured interviews to characterize the typical EHR usage by a group of hospital physicians (hospitalists) as they encounter a new patient.
Methods

This study was approved by the Institutional Review Board at Oregon Health & Science University (OHSU; Portland, OR). Subjects signed a consent form prior to participation.

Development of Simulation Cases

An instance of our EHR environment (EpicCare, Epic Systems Inc., Verona, WI) was created to house simulated patient cases. This simulation environment imports all end-user customizations from the actual EHR environment, so the interface looked exactly as it would for each subject. Simulated cases were based upon real patient cases with common principal diagnoses (i.e., among the top 10 most common ICD-9 diagnoses for adults upon hospital discharge)\(^2\). Two patient cases (Cases A and B) were created and independently reviewed for medical accuracy and clinical realism by domain experts in accordance with previously published recommendations for high-fidelity case creation\(^3\). Both patients had previously established care at our institution, but were now presenting to the emergency department (ED) for evaluation of a new set of symptoms. Each case contained historical data in each of the categories listed in Table 1; the current ED visit contained vitals, intake/output, laboratory data, EKG, chest roentgenogram, and a half-completed ED resident note stating the history of present illness, physical exam findings, and review of systems.

Recruitment and Testing of Subjects

Attending physicians from the OHSU Division of Hospital Medicine comprised the study population. Simulations were conducted on a representative active patient ward to mimic external distractions encountered by physicians as they use an EHR. Subjects were asked to act as the admitting hospitalist, review both patient charts, and create a history and physical (H&P) note complete with assessment and plan for each patient. Simulation time was not limited. Case order was held constant for all subjects throughout the study. After completing the cases, subjects were asked to verbally describe their typical workflow for admitting a patient. Semi-structured interview questions were used to elicit details about when they use the EHR during that process, their principal sources of information, note writing strategies, and the nature of the patient interaction. Lastly, subjects were asked to complete a questionnaire regarding demographic information, EHR experience, and general computer experience.

Eye and screen tracking were conducted using a Tobii X2 60 Eye Tracker (Tobii Systems, Danderyd Municipality, Sweden), a non-invasive tracker mounted below the computer monitor. All testing was conducted using a standardized computer station with consistent and static screen and chair height. Before each simulation, the eye tracker was calibrated to each subject using a 1-minute 9-point calibration algorithm provided by the manufacturer. Upon commencement of the simulation the screen tracking software (Tobii Studio, Tobii Systems, Danderyd Municipality, Sweden) captured screen video, keystrokes, mouse clicks, ocular saccades, and eye fixations. A velocity threshold identification filter was used to identify sets of fixations (gazes), using the standard definition of a fixation as lasting a minimum of 100 ms\(^1\). Each video was coded manually by a member of the research team (JD). Videos were coded by recording the information type upon which the gaze was situated at each second of the case.

Data Analysis

Simulation gaze data were divided into two major categories: informational and navigational. Informational gazes pertained to any kind of clinical data (all entries in Table 1 except 14); navigational gazes were defined as lacking clinical data, and frequently occurred on toolbars, menus, and up/down scrolling arrows. Documentation was considered a subset of the informational gaze category. Comparisons between group means were conducted using two-sided t-tests.
Several metrics were used to evaluate the data. First, we measured the average duration of each participant’s gaze on each information type (Table 1), calculated by their total duration of gazes on that information type divided by the number of gazes on that information type during the case. Next, we calculated the total number of transitions between information types for each patient case, and the total number of informational gazes for each case.

We used first-order Discrete Time Markov Chains (DTMCs) to model transitions between information types. A Markov Chain consists of a series of successive state-to-state transitions (Equation 1, Figure 1), which form a transition matrix. Equation 2 illustrates an example of a transition matrix, which has three states. The row i (i = 1,2,3) shows the transition probability distribution from information-type i to other information-types j. For example, $p_{13}$ is the probability that a subject transitions from information-type 1 to 3. Higher probabilities in the matrices indicate the information-type pairs that are more likely.

Results

Subject characteristics

From a total of 23 eligible subjects, 17 (74%) subjects completed a total of 33 patient cases. Fifty-nine percent of subjects were male; the mean length of time since medical school graduation was 13.3 years. One hundred percent of subjects described themselves as “somewhat” or “very experienced” with computers. The mean length of time using the study EHR (EpicCare) was 6.5 years.

Simulation characteristics

The subjects were divided into two groups based upon how long into the case it took them to begin composing a note. This division was based upon a natural grouping observed in average note start times per subject (Figure 2). Subjects in Group 1 (n=8) began composing a note on average less than 2 minutes into the case; subjects in Group 2 (n=9) began composing a note on average more than 2 minutes into the case.

<table>
<thead>
<tr>
<th>Subject Characteristics</th>
<th>Group 1</th>
<th>Group 2</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>n=8</td>
<td>p=0.02*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender, % male</td>
<td>87.5%</td>
<td>33.3%</td>
<td>0.02*</td>
</tr>
<tr>
<td>Self-rated computer experience, level</td>
<td>2.5</td>
<td>2</td>
<td>0.03*</td>
</tr>
<tr>
<td>Years since medical school graduation</td>
<td>15.3</td>
<td>11.5</td>
<td>0.42</td>
</tr>
<tr>
<td>EpicCare experience, years</td>
<td>7.3</td>
<td>5.7</td>
<td>0.27</td>
</tr>
<tr>
<td>Simulation Characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transitions</td>
<td>75.5</td>
<td>57.5</td>
<td>0.04*</td>
</tr>
<tr>
<td>Gazes</td>
<td>81.7</td>
<td>62.3</td>
<td>0.04*</td>
</tr>
<tr>
<td>Documentation, number of gazes</td>
<td>32.9</td>
<td>21.4</td>
<td>&lt;0.01*</td>
</tr>
<tr>
<td>Navigation, number of gazes</td>
<td>17.6</td>
<td>17.8</td>
<td>0.47</td>
</tr>
<tr>
<td>Number of unique information types</td>
<td>12.5</td>
<td>12.4</td>
<td>0.85</td>
</tr>
<tr>
<td>Case length, mm:ss</td>
<td>25:29</td>
<td>24:29</td>
<td>0.72</td>
</tr>
</tbody>
</table>

Table 2. Demographic and simulation characteristics

The proportion of men in Group 1 (87.5%) was significantly greater than the proportion of men in Group 2 (33.3%, p=0.02) (Table 2). Using a Likert-type scale to assess self-rated computer experience (1, less experienced; 2, somewhat experienced; 3, very experienced), Group 1 reported a higher mean experience score compared to Group 2 (2.5 and 2, respectively; p=0.03). Time since medical school graduation and length of EpicCare experience did not differ significantly between the groups.

The average time for Case A was significantly longer than Case B (mean±SD, 28:12±8:05 and 21:56±6:35 respectively, p=0.02). To evaluate a potential learning effect, the note start times were normalized to total case times and compared among the groups from Case A to Case B. There was a slight difference in the decrement in note start time ratio between the groups, with Group 2’s note start ratio dropping more in Case B, but this was not significant.
Number of transitions per second was evaluated as well. Whereas the transitions per second for Group 1 decreased slightly from Case A to Case B, the transitions per second for Group 2 increased slightly, but the difference between the two was not statistically significant (G1 = 0.006, G2 = -0.005; p = 0.14).

Group 1 had significantly more gazes (G1=81.7, G2=62.3; p=0.04) and transitions (G1=75.5, G2=57.5; p=0.04) over the course of each case. There was no difference between the groups in average case time, navigation time, and average number of unique information types accessed within each case.

Semi-structured interviews

Self-described workflows elicited from the semi-structured interviews were consistent with categorization into Group 1 and Group 2. Representative quotations are shown in Table 3.

Clinical content

The information types with the longest total gaze durations are shown in Figure 3. For both Groups 1 and 2, total duration of gazes for documentation (composition of the H&P note) was much higher than all non-documentation information (G1=851 seconds, G2=745 seconds). The non-documentation information types with the greatest average and total duration of gazes (greater 10 seconds) were imaging results, inpatient progress notes, lab values, medications, and ambulatory clinic notes. These five information types were also the most often visited throughout the simulations. Group 1 gazed at laboratory values significantly more often than Group 2 (G1=13.3, G2=8.2, p=0.02). Differences were observed in less frequently visited information types as well. Past medical history (G1=0.75, G2=1.6, p=0.06), problem list (G1=1.4, G2=2.9, p=0.08), family history (G1=0.7, G2=0.2, p=0.04) and other information types (G1=3.88, G2=2.53, p=0.04) all showed slight differences in visitation frequency between groups. These trends in frequency remained consistent when the gaze values were normalized by the total number of visits to all information types.

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>“I often start my note right away as I go about my chart review.”</td>
<td>“I review all the current data, labs, and imaging. Then look at last</td>
</tr>
<tr>
<td>“I tend to be linear… I jump around.”</td>
<td>clinic note, meds, and clinical history. I start writing a note after</td>
</tr>
<tr>
<td>“I usually start with a note because it autopopulates with the</td>
<td>that.”</td>
</tr>
<tr>
<td>information I need.”</td>
<td>“I look at the meds, prior notes and imaging, then start putting a</td>
</tr>
<tr>
<td></td>
<td>note skeleton together.”</td>
</tr>
<tr>
<td></td>
<td>“I do a quick review of the chart before I go see the patient. Then I</td>
</tr>
<tr>
<td></td>
<td>build my problem list and when I create the note, it auto-imports the</td>
</tr>
<tr>
<td></td>
<td>information I’ve collected.”</td>
</tr>
</tbody>
</table>
Transition visualizations

Figures 4a and 4b show circle visualizations of normalized Markov Chain frequencies of information types for both groups. Nodes situated around the rim of the circle represent the various information types, ordered by size moving counter-clockwise. The sizes of the nodes are proportional to their gaze number distribution frequencies, and the thickness of the lines connecting nodes indicate the normalized frequencies of transitions (transition probability) between the two information types. For clarity, only the top 80% of total transitions are depicted here.

The general patterns are the same between the two groups: most of the transitions are centered around documentation, which is also the most frequently visited information type. We can identify the top 5 most frequently visited information types, which are consistent between the two groups: documentation, lab values, inpatient progress note, vital signs and imaging results. In addition, because documentation played a larger role in group 1 than it did in group 2, there were fewer high frequency transitions compared with group 2.

One important observation from the visualization is which information types were closely related. For example, the social history, past medical history, family history, and past surgical history nodes are co-located because they were similar in visitation frequencies. In Group 1, there are more bold lines from documentation to other information types, capturing the notion that Group 1 transitioned more frequently between documentation and other information types.

Discussion

This study assessed characteristics of physicians’ information search patterns in an EHR as they created a note for a new patient. There were two discrete types of
users based upon information review and documentation tasks. The key findings from our analyses are: 1) There were strong similarities across the groups in the information types the physicians looked at most frequently; 2) While there was no overall difference in case duration between the groups, we observed two distinct workflow types between the groups with respect to gathering information in the EHR and creating a note; and 3) A majority of the case time was devoted to note composition in both groups.

Both groups showed the same preferences for a small subset of information types, in terms of how long they looked at each information type (Figure 3). Imaging results, progress notes, laboratory values, medications, and prior clinic notes were looked at longest. The total number of unique information types also did not differ between groups. This suggests some uniformity in clinical reasoning that may be explained by their mutual medical specialty, common clinical environment, and/or similarities in medical training. Differences between groups were only found in lower-frequency information types; members of Group 1 spent less time reviewing past medical history and problem lists, and more time reviewing the family history.

Groups 1 and 2 exhibited significantly different workflow types, despite no overall difference in case completion time. Group 1, characterized by early note creation, transitioned frequently between information types in the EHR after starting the note (Figure 4). Group 2 physicians, characterized by later note creation, tended to dwell on information longer before starting to compose the note. Group 1 showed a markedly higher number of transitions and gazes compared to Group 2, confirming a higher rate of switching from one information type to another. We found significant differences between the groups in how the simulation time was used. Overall, Group 1 spent substantially longer time in the documentation phase of the simulation. Subjects’ self-described workflows (Table 3) supported a dichotomy between early and note creation. Subjects in Group 1 mentioned starting a note as one of their first activities prior to information gathering (and “jumping around” in the EHR when gathering information) whereas Group 2 members described reviewing a variety of information prior to note creation.

Both groups spent a much higher amount of time on note composition than any other task, including reviewing clinical information (Figure 3). The time and burden associated with documentation is noted in the literature. What this study highlights is the finding that documentation time may overshadow all other tasks, including time to read or review the clinical data. This raises the question of whether the time and burden of documentation relates to the untoward effects of EHR implementation on patient care that have been observed. Further research is needed to elucidate the interplay between EHR documentation burden, clinical reasoning, and patient care.

This research raises several questions about the nature of EHR information seeking and patterns in end-user behavior. Several models have been proposed to describe this process. Traditionally information seeking is viewed as a sense-making process in which the user formulates a personal perspective through finding meaning. A common thread shared by most proposed models of information seeking is that it is a dynamic process, where the users move non-linearly through levels of certainty based upon information encountered and judgments of relevancy and specificity. The resultant perspective or decision is not necessarily the same among individuals and is dependent upon the effectiveness of the user’s information retrieval. This becomes problematic in the realm of clinical medicine, where standard of care dictates that there be some baseline level of uniformity in clinical reasoning to ensure patient safety. Ensuring some baseline level of competency in information retrieval becomes crucial when considering the complexity of modern EHR systems, where it has already been demonstrated that end-user behavior is highly variable. This research suggests that users take different pathways to arrive at a common endpoint. The information used and the time to complete the task may not differ, but the order in which the end product (clinical note) is created may differ depending on the user. One of Nielsen’s five criteria for usability is affordance. The different workflows described in this study support the need for interfaces to afford for 1) fluctuation between varying levels of certainty in the meaning-finding process and 2) a variety of approaches in clinical documentation.

Of note, significant differences were observed in the gender composition and self-rated computer experience level between Group 1 and Group 2. With respect to the gender difference, though the sample size of the current study is too small to draw firm conclusions from this observation, our results raise the question of whether different genders may approach the tasks of EHR information gathering and documentation differently. Gender differences in clinical reasoning and information processing have been explored previously. Myers-Levy’s theory of selectivity and information-processing research in other disciplines suggest that women are more likely to employ elaborative information processing strategies regardless of the task complexity, whereas men are more likely to utilize heuristic processing strategies, only switching to elaborative strategies on more complex tasks. Conversely, in the medical literature no significant gender-related difference has been found in diagnostic reasoning. With the insertion of the EHR into clinical workflows we must consider the electronic interface as an additional layer of complexity. Research
conducted on website audiences has shown that differences in perception and satisfaction can vary greatly among gender groups\(^5\). Females have demonstrated greater proficiency in computer display navigation and optical cue responsiveness on the screen.\(^2\) Taken together, prior research suggests that there may be several factors that contribute to gender differences in EHR usage; further research is needed. With respect to the difference in self-rated computer experience, it is unclear whether this outcome is independent of the strong gender differences between the groups. Conflicting research exists on gender differences in perceptions of self-efficacy and attitudes toward computers.\(^4\)\(^5\). If it is an independent outcome, it is not clear that the higher level of computer experience in Group 1 translated into better performance in the simulation task.

There are several important limitations to this study. First, though simulations were conducted in a high-fidelity environment, subjects were aware that the cases were fictional patients. There were no actual patients to interact with and subjects relied upon the content of the history of present illness, review of systems, and physical exam as it was documented in the EHR. Though this may have diminished the realism of the cases, it is expected that it would exert a uniform effect on the subjects. Cases were made to be slightly less complicated than the “average” patient seen at our institution, a tertiary care facility, for the purposes of simulation time. Second, interpretation of the eye tracking gaze data is still in its infancy in clinical informatics research, and it is unclear how well gazes and transitions represent clinical reasoning processes. Third, the notes created during the simulations were not evaluated for accuracy and completeness, thus we cannot comment on whether differences in search patterns affect clinical reasoning and medical decision making. Lastly, the study was conducted among one specialty of physicians at one institution.

Conclusion

This study demonstrates the presence of two information-gathering and documentation workflows among hospitalists using the EHR to admit a new patient. This has important implications for EHR interface design, specifically with respect to affordances for multiple information-gathering pathways. Future studies must continue to examine the workflow differences among individuals, specifically pertaining to note quality, clinical accuracy, and efficiency.

References