Challenges and Opportunities to Improve the Clinician Experience Reviewing Electronic Progress Notes

Gretchen M. Hultman1  Jenna L. Marquard2  Elizabeth Lindemann3  Elliot Arsoniadis1,3  Serguei Pakhomov1,4  Genevieve B. Melton1,3

1Institute for Health Informatics, University of Minnesota, Minneapolis, Minnesota, United States
2College of Engineering, University of Massachusetts, Amherst, Massachusetts, United States
3Department of Surgery, University of Minnesota, Minneapolis, Minnesota, United States
4College of Pharmacy, University of Minnesota, Minneapolis, Minnesota, United States

Address for correspondence  Genevieve B. Melton, MD, PhD, Department of Surgery, Institute for Health Informatics, 420 Delaware Street SE, Mayo Mail Code 450, Minneapolis, MN 55455, United States (e-mail: gmelton@umn.edu).

Abstract

Background  High-quality clinical notes are essential to effective clinical communication. However, electronic clinical notes are often long, difficult to review, and contain information that is potentially extraneous or out of date. Additionally, many clinicians write electronic clinical notes using customized templates, resulting in notes with significant variability in structure. There is a need to understand better how clinicians review electronic notes and how note structure variability may impact clinicians’ note-reviewing experiences.

Objective  This article aims to understand how physicians review electronic clinical notes and what impact section order has on note-reviewing patterns.

Materials and Methods  We conducted an experiment utilizing an electronic health record (EHR) system prototype containing four anonymized patient cases, each composed of nine progress notes that were presented with note sections organized in different orders to different subjects (i.e., Subjective, Objective, Assessment, and Plan, Assessment, Plan, Subjective, and Objective, Subjective, Assessment, Objective, and Plan, and Mixed). Participants, who were mid-level residents and fellows, reviewed the cases and provided a brief summary after reviewing each case. Time-related data were collected and analyzed using descriptive statistics. Surveys were administered and interviews regarding experiences reviewing notes were collected and analyzed qualitatively.

Results  Qualitatively, participants reported challenges related to reviewing electronic clinical notes. Experimentally, time spent reviewing notes varied based on the note section organization. Consistency in note section organization improved performance (e.g., less scrolling and searching) compared with Mixed section organization when reviewing progress notes.

Keywords  ► electronic health records and systems  ► clinical documentation and communication  ► qualitative methods  ► workflow  ► sociotechnical aspects of information technology

Discussion  Clinicians face significant challenges reviewing electronic clinical notes. Our findings support minimizing extraneous information in notes, removing information that can be found in other parts of the EHR, and standardizing the display and order of note sections to improve clinicians’ note review experience.

Conclusion  Our findings support the need to improve EHR note design and presentation to support optimal note review patterns for clinicians.

Background and Significance

Clinical notes within electronic health record (EHR) systems remain a key element of care documentation and communication. However, clinicians using EHRs consistently report poor system usability, time-consuming data entry, and degradation of clinical documentation as key challenges associated with EHRs. In part, this is because EHRs enable the inclusion of large amounts of structured information into electronic notes (“notes”) via auto-filling or “carry-forward” capabilities, including “copy and paste” functionality. While these features can promote thoroughness and can be efficient to use, these data can hinder note readability and can result in inaccurate or irrelevant information in notes (e.g., copying and pasting text that says “yesterday” without updating the reference each day). Lengthy notes can also overload users cognitively, and impede them from deciphering and retrieving key information. This problem is not isolated to providers only and is an important challenge for a range of clinicians. For example, one study showed that pharmacists spend significant time searching for and reading information in notes, highlighting the challenges with extracting information from notes. As such, notes are now considered by some to be data-rich and information-poor.

Challenges related to use of clinical notes and EHR systems are not limited to the United States. Kaipio et al documented usability challenges with EHR systems in Finland. In a Dutch study, authors implemented a clinical notes application. While they had some success, the study demonstrated that there was room for improvement with respect to usability of such systems.

Despite this, clinicians value the narrative expressivity of unstructured data in notes because notes can provide insight into clinical decision-making and point out salient patient case data. However, unstructured data can be difficult to search and read. Because of these concerns, there is a need to make the most clinically relevant data within notes easy to find and read.

Progress notes, the most typical type of clinical note used to document care management for established patients, most often follow the Subjective, Objective, Assessment, and Plan (SOAP) note format that was established by Dr. Lawrence Weed in the 1960s as part of the Problem-Oriented Medical Record framework. Anecdotally, clinicians report that the Assessment and Plan sections of the note are most important when reviewing a progress note and that it is common for providers to read these sections first, regardless of where they appear in the note. Some have suggested that note reading could be improved by rearranging the sections of the notes so that the Assessment and Plan sections are at the top of a note.

A recent study examined clinician satisfaction after adoption of Assessment, Plan, Subjective, and Objective (APSO) notes with 13 outpatient clinicians. The study found that clinicians largely favored the APSO notes despite the inconvenience of adopting the new method for creating progress notes. Another study used eye-tracking data to determine what information clinicians focused on while reviewing electronic progress notes. The study found that clinicians spent most of their time reading the Assessment and Plan section. In most institutions, however, clinicians continue to write progress notes with customized section orders (and thus unpredictable section ordering for readers). This unpredictability may challenge note readers as they search for relevant information. It is well-known in the user interface design community that user preference and task performance are often not aligned. For this reason, electronic clinical note formatting should not be changed purely based on user preference.

Objectives

This study sought to gain insight into how clinicians read and use electronic progress notes. Our objectives were to understand how section order impacts note-reading patterns and to gain qualitative insights into physician note-reading behaviors and preferences. We conducted an experiment where clinicians read electronic progress notes randomized in different orders and provided case summaries for a realistic set of patients to better understand how the ordering of information in progress notes impacted how clinicians read and synthesize these notes. We also analyzed closed-form and open-ended questionnaire responses, and semi-structured interview responses.

Materials and Methods

Study Design

This institutional review board-approved study was conducted at a large Midwestern academic health center. A previously described EHR system prototype designed to look like the VistA Computerized Patient Record System (CPRS) was populated with four deidentified patient cases, each with nine progress notes. The patient cases were
designed to be of similar complexity, and to realistically represent patients being managed for chronic medical conditions frequently encountered in adult primary care clinics. These cases were used in previous experiments, which assessed their complexity and reading time, how time constraints changed patient case synthesis, and the effect of highlighting new information in the notes.18–20

Each of the nine progress notes within the cases were arranged and presented in four different section orders: SOAP, APSO, SAPO, and Mixed section format. In the Mixed section format, the section orders were inconsistent across the nine notes, with three notes presented in each of the above orders. In addition to the classic SOAP sections, notes contained other sections that could be classified as subjective, objective, assessment, or plan. Examples of these other elements included History of Present Illness (classified as Subjective), and Visit Diagnosis (classified as Assessment).

The four cases were presented in the same order, but each case had a different note order, such that each participant saw all four note orderings but with randomized note-ordering assignments for each case. For example, participant 1 viewed case 1 in the Mixed order, followed by SAPO and APSO, while participant 2 viewed case 1 in the APSO order, followed by SOAP, Mixed, and SAPO. Note orderings were randomized using a Latin squares design. The note orderings within the Mixed order were also randomized using a Latin squares design. Note orderings were randomized to account for the possible effect of participant behavior changing as the experiment progressed. For example, participants may have systematically read notes in the first patient case for longer than later patient cases. If the same note ordering always occurred first, the results may be confounded.

Participants (n = 23) were mid-level residents at a large Midwestern training program (internal medicine [n = 15], surgery [n = 8]). This participant population was chosen so that we could recruit a larger sample of participants and have participants with similar experience levels. Participants were recruited via emails and flyers placed in workrooms. This choice was made to control for differences in experience level that might affect how users read and interpret clinical notes. Sessions were facilitated by two researchers (G.H. and O.I.). The study was pilot tested with two participants. Sessions lasted between 60 and 90 minutes.

Study Procedure
Participants were seated at a desktop computer with the EHR interface opened to the note list for the first patient case. Participants were asked to review the notes for the patient as they normally would and then provide a verbal summary of the case. Upon completing the verbal summary, participants rated their perceived workload for the case using the validated National Aeronautics and Space Administration Task Load Index (NASA-TLX) workload instrument.21 Participants repeated this process for the three other cases. We used screen capture software to capture participants’ navigation patterns, a typical approach in EHR usability studies.22 At the end of the session, participants completed a questionnaire with demographic questions, questions regarding their experience with different EHR systems, when and how long it takes them to review a set of electronic notes for a single patient, the importance they place on different information types within electronic notes, how well current electronic notes support the retrieval of different information types, and general barriers to accessing the information they need in electronic notes. The questionnaire was developed based on a previous experimental study where hospitalists reviewed information in the EHR prior to admitting a patient.23 The questionnaire was pilot tested to ensure the questions were understandable. All sessions were audio recorded.

Analysis of Experimental Data
The analyses of the experimental data included participants’ time to review notes, time spent verbally summarizing cases, and perceived workload. Each measure was calculated independently for the SOAP, APSO, SAPO, and Mixed note orders. Participant videos were reviewed and coded for scrolling time (screen moving), still time (screen not moving), and navigating time (using a selection box to move between patient notes). Times were tabulated and averages calculated. Two coders reviewed videos for two participants to determine a standard coding process.

Analysis of Qualitative Data
After completing the four cases, we conducted a semistructured debriefing interview with each participant. Interviews were conducted by G.H., a researcher with experience in qualitative analysis, and O.I., a researcher with a background in medicine. Interviews were transcribed, and coded for emerging themes by two coders with experience in qualitative analysis (G.H. and E.L.). One participant was excluded because the interview did not record properly. Interview answers were coded into categories and reviewed in collaboration with a physician expert. A subset of transcripts was coded by both coders to assess the internal reliability of coding (6, 26% of all transcripts). There was overall high agreement between coders (kappa = 0.81, percentage agreement = 98%).

Below, we present our analysis of participants’ perceived note reading patterns based on the interview and questionnaire data. We then present our analysis of participants’ actual note reading patterns based on the experimental data.

Results
Participant Characteristics
Fourteen participants were male and 9 were female. Participants were on average 29.9 years old (standard deviation [SD] = 2.48) and graduated from medical school an average of 2.82 years prior (SD = 1.47). Participants reported their levels of experience with eight different EHRs. Participants reported the highest self-rated experience using Epic; all participants considered themselves average or expert Epic users. VistA CPRS, on which the prototype used in this experiment was based, was used by 20 of the 22 participants. Most (n = 17, 74%) said they considered themselves
average VistA CPRS users; two considered themselves expert users.

Timing of and Time Spent Reading
Participants were asked in a free response question when they typically read patient notes. Responses fell into one or more of five broad categories. Participants reported that they review notes prior to a patient encounter ($n = 22, 96\%$), during an encounter ($n = 1, 4\%$), after an encounter ($n = 1, 4\%$), when writing a note ($n = 2, 9\%$), or when trying to answer specific questions ($n = 2, 9\%$).

Participants were asked to report three values related to how long they typically spend reviewing notes for a single patient: (1) minimum time, (2) average time, and (3) maximum time. Fig. 1 shows these responses with the box located at their average time and the whiskers noting their minimum and maximum values. The horizontal line shows the overall average across participants, though variability within and across participants was substantial.

Reading Order
When asked in the semistructured interview how they typically read through a single patient note, 22 participants responded, and when asked how they typically read through a set of notes for one patient, 21 participants responded. When asked how they read a single note, 14 mentioned starting with Subjective, and 6 mentioned starting with the Assessment and Plan. When asked how they read a set of notes, participants discussed different strategies such as looking for notes from specific specialties or looking for the most recent notes. Responses are summarized in Tables 1 and 2.

Value of Information Types in Notes
Participants were asked to list the five information types they considered the most valuable to them within notes, and ranked the information-types from 1 to 5, with 1 being the highest priority and 5 being the lowest (of the top 5). Shown in Fig. 2, the most recent Assessment and Plan, past medical history, and chief complaint were the three types of information most often ranked as a top 5 priority. Chief complaint and most recent Assessment and Plan were the two categories most frequently listed as priority of one. Labs, Imaging, and Medications were frequently listed in the top five, but as lower priorities. Past surgical history was a priority for surgical but not medicine residents.

Participants were asked how well each section type provides the clinical information they need. Almost all (22/23, 96\%) participants said that the Assessment section met their information needs very well, somewhat well, or neither well or poorly. All participants (23/23, 100\%) said the Plan section meets their information needs very well, somewhat well, or neither well or poorly. The Objective section was rated as meeting information needs somewhat poorly more than any other section. Other sections participants provided included: Medications (very poorly), Imaging (very well), and “blown in data” (auto-populated data) (somewhat poorly).

Table 1 How participants typically read through a single patient note

<table>
<thead>
<tr>
<th>Start with subjective</th>
<th>14</th>
<th>64</th>
<th>“Typically, when assessing a patient note for any given specialty, I’ll look at their HPI or initial subjective assessment, then go and jump to the assessment and plan”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start with A/P</td>
<td>6</td>
<td>27</td>
<td>“I read the assessment and plan and then scroll back up depending on what I need”</td>
</tr>
<tr>
<td>Start with something else</td>
<td>2</td>
<td>9%</td>
<td>“I try to hone into something that person wrote, like, the actual things that are not auto-populated”</td>
</tr>
<tr>
<td>Depends on context</td>
<td>1</td>
<td>5</td>
<td>“It does depend a little bit on the setting that I’m reading it in”</td>
</tr>
<tr>
<td>Mentioned information skipped</td>
<td>4</td>
<td>18</td>
<td>“I usually skim through the lab results, imaging studies, and especially the physical exam I go through quickly”</td>
</tr>
</tbody>
</table>

Fig. 1 Self-reported note reading time. Participants reported their average, maximum, and minimum note reading times. The average time for each plotted and the whiskers represent the maximum and minimum reading time. The line represents the average of participants’ average reading time.
Barriers to Reading Notes
Participants were asked about six types of information barriers to accessing information they needed from notes. Many of the barriers selected were previously identified in other research, including necessary information not being in notes, there being too much information in notes, and information not being accurate. Participants rated each barrier on a scale of 1 to 5, with 1 being not a barrier and 5 being a severe barrier. As summarized in Fig. 3, most participants (n = 21) perceived that having information in notes poorly displayed or difficult to interpret was a moderate, large, or severe barrier to accessing patient clinical information while not having key needed information in notes was also perceived as a moderate, large, or severe barrier for most participants (n = 16, 73%). Additionally, all participants (n = 22) reported that too much information in notes presented a barrier to note reading.

Experiment Results
As summarized in Table 3, differences in reading time were observed across the four presentation orders, with participants taking the shortest amount of time to read APSO ordered notes and the longest amount of time to read Mixed ordered notes. On average, participants took 1.9 minutes longer to read the Mixed ordered notes versus the APSO ordered notes. On average, participants also took 1.0 minutes less to read APSO ordered notes compared with traditional SOAP notes. The observed times spent reading notes for the cases were in the range of participants’ self-reported time spent reading patient notes. Our analysis revealed that participants tended to spend a significant amount of time

Table 2 How participants typically read through a set of notes for single patient note

<table>
<thead>
<tr>
<th>Number</th>
<th>%</th>
<th>Representative quote</th>
</tr>
</thead>
<tbody>
<tr>
<td>Look for certain specialties</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>Look for notes relevant to Chief Complaint</td>
<td>8</td>
<td>36</td>
</tr>
<tr>
<td>Look for certain note types</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>Start with most recent notes first</td>
<td>6</td>
<td>27</td>
</tr>
<tr>
<td>Start with older notes first</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>Mentioned notes skipped</td>
<td>4</td>
<td>18</td>
</tr>
</tbody>
</table>

Fig. 2 Value of information-types in notes. Participants were asked to list the five types of information they found most valuable and rank them 1–5. This bar chart shows the types of information mentioned and the color denotes its priority.

Fig. 2  Value of information-types in notes. Participants were asked to list the five types of information they found most valuable and rank them 1–5. This bar chart shows the types of information mentioned and the color denotes its priority.
scrolling (average $\approx 4.23$ minutes per case, or $\approx 39\%$ of reading time for each case), regardless of how note sections were ordered. Differences in verbal summary times are also presented in - Table 3. NASA-TLX perceived workload scores (0–50 scale) across the case orders were similar. The NASA-TLX score was lowest for SOAP notes, even though reading time was lowest for APSO notes. Scores are summarized in - Table 3. Reading times, verbal summary times, and TLX scores were similar for medicine and surgery residents within each section order.

Participants were asked if the experimental patient cases were realistic. All participants (23/23, 100\%) said the cases were realistic, stating that the cases represented common medical conditions, and that the flow of care from one provider to another seemed realistic.

Discussion

This study provides important insights about how clinicians read and review progress notes and the challenges they face. Our study also provides insight into the impact that section order has on note reading. Many have documented current concerns with electronic documentation.24,25 These concerns have been documented outside the United States where EHRs are used.10,11 One study examined two potential avenues for improving documentation: having residents attend a lecture or having them attend a lecture coupled with individual

Table 3 Key times and NASA-TLX scores

<table>
<thead>
<tr>
<th></th>
<th>Reading time (min)</th>
<th>Proportion with screen still</th>
<th>Proportion scrolling</th>
<th>Verbal summary time (min)</th>
<th>Average workload scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>SOAP</td>
<td>11.6</td>
<td>61%</td>
<td>38%</td>
<td>2.1</td>
<td>30.6 (10.57)</td>
</tr>
<tr>
<td>APSO</td>
<td>10.6</td>
<td>60%</td>
<td>39%</td>
<td>1.9</td>
<td>31.3 (8.75)</td>
</tr>
<tr>
<td>SAPO</td>
<td>11.3</td>
<td>57%</td>
<td>40%</td>
<td>2.3</td>
<td>31.9 (7.04)</td>
</tr>
<tr>
<td>Mix</td>
<td>12.5</td>
<td>59%</td>
<td>39%</td>
<td>2.1</td>
<td>31.7 (7.78)</td>
</tr>
<tr>
<td>Average</td>
<td>11.5</td>
<td>59%</td>
<td>39%</td>
<td>2.1</td>
<td>31.4 (8.52)</td>
</tr>
</tbody>
</table>

Abbreviations: APSO, Assessment, Plan, Subjective, and Objective; NASA-TLX, National Aeronautics and Space Administration Task Load Index; SAPO, Subjective, Assessment, Objective, and Plan; SOAP, Subjective, Objective, Assessment, and Plan.
Currently, the use of custom note templates means that the display order for notes tends to vary across clinicians. We observed that notes with mixed section order negatively impacted participant performance. While there are some advantages to flexible documentation, participants also reported frustrations with notes appearing in differing formats. This indicates that clinicians and organizations would benefit from having a standard order for sections in which all notes appear. The display of note sections could be uncoupled from the order of sections in how notes are written (i.e., custom note entry; standard note presentation). Overall, this could allow clinicians to create notes using their preferred style while displaying notes for optimal reading and use. This approach could address both user preferences and the need for standardization of note display.

This study also provides some confirmatory evidence to support concerns related to the negative impacts of auto-populated data and copy and pasted data in notes. Not only did participants report frustrations and barriers due to auto-populated data and extraneous information, but moving this information to the end of the note decreased time spent reviewing note sets. This supports findings from other studies that indicate clinicians are generally more satisfied with notes in APSO order instead of SOAP. Many participants suggested that auto-populated data can be found more easily and in more readable form elsewhere within EHRs. Further work should examine the impact of removing auto-populated data from notes, especially information that is not relevant to the chief complaint or current active problems, perhaps creating links within the note to other areas of the EHR. Regulations that require certain information be in notes and the impact of changing them could be examined as part of this work.

Interestingly, we did not observe large differences in perceived cognitive load when notes were organized differently, and participants reported the lowest cognitive load score for SOAP notes. Similar to a study of another type of EHR functionality (i.e., ambulatory navigators), we observed that perceived cognitive load was unrelatable to participant performance.

This study has several limitations that reduce its generalizability including its single organization, and relatively small sample size. Our convenience sample was limited to mid-level resident physicians which may affect generalizability. Participant experience with note reading can be affected by previous experience, preferences, and training. Additional confounders such as specialty may be present.

Conclusion

Overall, participants reported that information in notes remains poorly displayed and difficult to interpret. Future work should identify in more detail the specific aspects of the note designs that led participants to consider the notes poorly designed or difficult to interpret. Future work should also validate and expand upon these findings with other user groups and EHR systems, and test alternative note designs.

Clinical Relevance Statement

This work represents a step toward understanding how clinicians review electronic clinical notes and the barriers that they face. This work also represents a step toward understanding how the ordering of information in notes impacts clinician experience reviewing notes.

Multiple Choice Questions

1. Which note order was the most challenging for participants to read?
   a. SOAP.
   b. APSO.
   c. SAPO.
   d. Mix.

   Correct Answer: The correct answer is option d. In this study, the mixed order was the most challenging for participants to read. This order required the most reading time for participants. Participants also commented on frustrations with inconsistently formatted notes in the qualitative portions of this study.

2. Based on the finding of this study, what steps could improve the experience with reading notes?
   a. Allow physicians to write notes in any order according to reading style.
   b. Add additional options for auto-populating data in notes.
   c. Have a standard note writing template.
   d. Add additional data to notes.

   Correct Answer: The correct answer is option c. In this study, participants were negatively affected by notes that were highly variable. Participants were also negatively affected by notes that were in the mixed order. They were also reported frustrations with long notes and notes that contained a lot of auto-populated data.

Protection of Human and Animal Subjects

This study was performed in compliance with World Medical Association Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects. This work was reviewed by the University of Minnesota Institutional Review Board.

Funding

This work was supported by the Agency for Healthcare Research and Quality Award R01HS022085 (to G.M.) and National Science Foundation Award CMMI-1150057 (to J.M.).

Conflict of Interest

None declared.

Acknowledgments

We would like to acknowledge Osadebamwen Ighile, MBBS, MS, and Oladimeji Farri, MBBS, PhD, for their contributions to this study.
References


22 Turf Usability Software [Computer Program]. Version 4.0. Huston, Texas. University of Texas Health Science Center at Huston; 2015


