Design, Implementation, and Usability of the Electronic Medical Record Search Engine (EMERSE) Tool

Colby N. Reyes, MS[1], Kai Zheng, PhD[1], David A. Hanauer, MD, MS[2]
[1]University of California, Irvine, Irvine, CA, USA; [2]University of Michigan, Ann Arbor, MI, USA

Abstract

Free text forms of clinical documentation stored in electronic health records contain a trove of data for researchers and clinicians alike. However, often these data are challenging to use and not easily accessible. EMERSE, a clinical documentation search and data abstraction tool developed by the University of Michigan, helps users in the task of searching through free text notes in clinical documentation. This study evaluates the usability and user experience of the EMERSE system, and draws inferences for the design of such systems. The study was conducted in 3 phases. In Phase 1, interviews with site administrators investigated factors that facilitate or hinder the implementation and adoption of EMERSE. Phase 2 employed semi-structured interviews to understand the uses, benefits, and limitations of the system from the perspective of experienced users. In Phase 3, system-naive users performed a set of basic workflow tasks, then completed post-activity questions and surveys to evaluate the intuitiveness and usability of the system. Participants rated the system exceptionally high on usability, user interface satisfaction, and perceived usefulness. Feedback also indicated that improvements could be made in visual contrast, affordances, and scope of notes indexed. These results indicate that tools such as EMERSE should be highly intuitive, attractive, and moderately customizable. This paper discusses some aspects of what may contribute to a system having such characteristics.

Introduction

Electronic Health Record (EHR) use has expanded rapidly over the last decade and allowed for the generation and storage of vast amounts of clinical data. However, uses of these data have been limited by the form and nature of this data[1,2]. The natural and expressive nature of free text clinical documentation makes it useful to clinicians in the immediate context of care[1]. Outside of its original context, however, and coupled with the sheer volume of data generated in this format on a daily basis, these data become largely inaccessible—and therefore unusable—to clinicians, healthcare administrators, and researchers[2–5] for secondary use purposes. If properly harnessed, broader use of these clinical data could propel the healthcare system towards a true Learning Health System, a self-learning vehicle that feeds a virtuous cycle of improvement through cohesive, perpetual interplay between research and application[3,6,7]. Various clinical information extraction (CIE) tools have been developed over the years in an attempt to solve this issue and provide solutions that allow users to harness the value of the information contained in clinical documentation. Wang et al[1] present a review of such tools and conclude that the adoption of CIE use is still in its early stages of development, as CIE tools are only used in a small portion of the published papers that use EHR data[1]. Reasons for limited use include: access to EHR data for NLP professionals is limited; many clinical NLP systems rely on rule-based approaches which more easily accommodate the knowledge of clinical domain experts, but are more rigid and require a higher level of clinical literacy; and lack of interoperability that would allow for the training and transfer of CIE tool use across locations and clinical systems. Of note, however, is that EMERSE does not appear in their list of CIE tools, despite its documented use[8] in all of the application usage CIE tool categories they identify[1]. However, this is likely because their review focused on NLP-based tools and applications and excluded articles on "information retrieval"[1]. This exclusion is indicative of the larger trend towards NLP-based tools that may in some respects be more powerful than EMERSE, but generally lack the same measure of usability and user-friendliness[2,9–12], or have not yet been implemented in a tool with the usability and utility that EMERSE has demonstrated[8]. Hill et al reviewed the usability and common use of within-EHR CIE and search tools[2]. While such tools are generally positively perceived by users, they suffer from substantial issues in both usability and utility, due partly to the high cognitive load these tools impose on users[2]. Though the authors do not address it, these issues are likely also due to the complex, confusing, and poorly designed EHRs within which the tools function[13–15].

EMERSE occupies a liminal space between the types of tools reviewed by Wang et al[1] and Hill et al[2]. EMERSE is an information retrieval system for narrative documents stored in EHRs, with features specifically designed to work with medical text and to support studies involving cohorts. Due to its design[4], it is suited to clinical and translational research uses[1] and uses in the context of clinical tasks[2]. EMERSE addresses a number of issues and weaknesses present in NLP-based CIE systems and within-EHR search tools. Over the 15 years since EMERSE was originally introduced and validated[4], a number of studies have evaluated various aspects of the EMERSE system at different points in the ongoing history of its development[3,4,9,16–19]. These studies have demonstrated that EMERSE improves efficiency of chart reviews[19], that collaborative term bundles can decrease cognitive load for clinical users and facilitate continuity and reproducibility of research and other types of work done using EMERSE, and that users consistently find it highly valuable for job efficacy and efficiency[2,3,4,9,16–19]. They also note that EMERSE achieves these outcomes through its focus on a clean and highly intuitive user interface, and through a system design that is flexible and powerful enough to facilitate effective queries for users of varying search engine proficiency and "clinical literacy"[9,17,18]. This study follows upon this ongoing research, seeking to understand factors that facilitate or detract from EMERSE's usability and utility in supporting such works as the EMERSE project is expanding to new sites. The purpose of this study is to solicit feedback about EMERSE's usefulness and ease of use, and to identify implementation and usability issues and potential solutions in order to guide the development and dissemination of this and other similar software.

Methods

The study consists of three phases. In each phase, participants were either research or healthcare professionals who use EMERSE and/or other research software systems on a daily basis, or oversee such software use and its users. The research protocol for the research activities involving human subjects was reviewed and approved by the Institutional Review Boards of the University of Michigan Medical School (IRBMED).

Participants & Recruitment

Participants consisted of a total of 55 individuals from 7 different sites, including the University of Michigan and 6 large academic medical centers. A total of 10 individuals composed the 6 teams from the external project sites for Phase 1; 20 participants from the University of Michigan were interviewed individually for Phase 2; and 25 participants at 5 external sites were interviewed individually for Phase 3.

Phase 1 participants were site leaders and administrators from UC San Francisco, University of Kentucky, University of Cincinnati, Columbia University, University of North Carolina, and Case Western who were either considering, or in the process of, implementing EMERSE as a tool for their staff. Project teams from each of the six different academic medical centers involved were contacted and asked to participate in a semi-structured interview. Phase 2 of the study identified 20 veteran EMERSE users at the University of Michigan with the highest system usage based on system log data and recruited them to participate in semi-structured interviews. In Phase 3, following initial installation and functional testing of EMERSE, five users each at University of Kentucky, University of Cincinnati, Columbia University, University of North Carolina, and Case Western were identified by the site PIs and recruited to participate in usability studies to explore issues that may arise from the sites' local contexts. These participants were clinician scientists, clinical research coordinators, or healthcare administrators, who frequently perform medical chart review tasks that can be facilitated by the use of the EMERSE system, but who had never used the EMERSE system.

Data Collection

In all three phases, data collection took place in the context of interviews conducted via Zoom video-conferencing, with an additional administration of QUIS and TAM surveys for Phase 3 of the study.

Phase 1: Leadership interviews were conducted with the project team at each participating site. Interviews lasted approximately one hour, and followed a semi-structured protocol containing 12 questions on organizational context, characteristics of anticipated users, progress of implementation, obstacles encountered, and satisfaction with the technical support provided by the University of Michigan team.

Phase 2: Semi-structured interviews were conducted with 20 participants at the University of Michigan who used EMERSE as a routine part of their job. Interviews included questions informed by technology acceptance theories, developed to solicit feedback about the system's usefulness and ease of use, and to identify usability issues. Interviews lasted 30 to 45 minutes. Upon participant consent, interviews were recorded and transcribed to facilitate qualitative analyses. No identifying information was collected during the interviews.

Phase 3: User testing, contextual interviews, and survey administration were conducted with 25 participants from 5 sites. Users were asked to review documentation, then followed a structured test script to perform a set of pre-defined search tasks on a test version of EMERSE containing 10,000 simulated patients and ~500,000 documents consisting of PubMed abstracts and open-access journal case reports. The interviewer collected and catalogued qualitative data in the form of observations[20] as users performed the prescribed tasks. Participants were then asked to report their perceptions and experience with the system in two surveys: (1) the validated and widely used Questionnaire for User

Interface Satisfaction (QUIS)[21], and (2) a validated questionnaire based on the technology acceptance model (TAM) which assesses key determinants of technology acceptance behavior among prospective users[22–24]. Each session, including usability testing and time to respond to the questionnaires, lasted between 30 to 50 minutes. No pieces of identifying information were collected.

Data Analysis

Initial analyses were performed *in situ* as interviews were conducted[21,25], particularly in Phase 3 where anomalies and patterns were identified by observation as participants executed the prescribed tasks of the user evaluation. These observations were further refined and categorized into the event types laid out in Table 1. Outcomes were then funneled into the quantitative analysis further described below. In order to perform qualitative analysis, interviews were transcribed using the Automatic Speech Recognition module in the Google Cloud Platform. Interviews from all three phases were coded using the constant comparative method of qualitative coding. Transcripts from all three phases were analyzed qualitatively following the general procedure posed by Creswell[20] and aligned with the Constant Comparison Method. Following in this process outline, and drawing on tools and techniques from phenomenology (coding significant statements)[20] and grounded theory (open coding and selective coding)[20], each interview transcript with its corresponding field notes was coded by identifying and chronicling significant statements, the interviewer's observation of participant's sentiment toward the software, emerging categories of statements and observations, and the interconnection of codes. Codes were then iteratively refined and re-compared to the data as ideas, patterns, and themes emerged within and across study phases. Qualitative analysis was aided by NLP-based semi-automated qualitative coding techniques used in the analysis of other types of software[25–29] using the python library nltk to aid in the production, examination, refinement, and verification of qualitative codes in the constant comparison method process. Codes were then summarized into larger recurring themes. Quantitative analysis was performed by means of descriptive summary statistics on both the survey data and the catalogued events and outcomes of the steps of the demonstration scenarios.

Table 1. Event codes with corresponding meta-types and descriptions

Event Code	Meta-Type	Description
Error/Failure (F)	*Fail*	Participant "failed" the step (completed it incorrectly, or skipped it)
Help (H)	*Help Request*	Participant asked for help completing the step or helping understand what a step accomplished
Initial Mistake (I)	*Confusion*	Participant initially made a mistake (e.g. wrong click) and then corrected it themselves
Delay (D)	*Confusion*	Participant exhibited notable delay between actions
Searching (S)	*Confusion*	Participant exhibited behaviors (e.g. mouse movements) indicative of searching for interface items
No Issue	*No Issue*	Participant completed step with none of the above events

Results

Qualitative Analysis

Phase 1: Feedback about EMERSE and its implementation was overall very positive. All sites were excited about the prospect of using EMERSE to meet the local needs for retrieving information from free text. All spoke highly of the ease of deploying the software and the responsiveness of the University of Michigan support team. Most improvement suggestions were made by early adopting sites and focused on the inadequacy of software documentation and deployment instructions. These issues have since been addressed, as the sites that started implementation later in the project did not report similar observations. Other improvement suggestions included accommodation of additional database management systems besides Oracle (which is now available), and a few user interface enhancements to facilitate local workflow. Site leads indicated that a key implementation barrier was obtaining organizational approval, which was often the main reason for implementation delays. All sites commented that the process of installing and configuring EMERSE was straightforward. The only technical barrier was to import clinical notes from the electronic health records system and format the notes properly for consumption in EMERSE. Overall, there were no significant concerns regarding system implementation.

Phase 2: Qualitative analysis of Phase 2 interviews surfaced key themes relating to (a) the desire for the system to index more data sources so that EMERSE could be used for more tasks and/or without the need for any other software interactions, and (b) the need for promotion of the system to expand the system's user base.

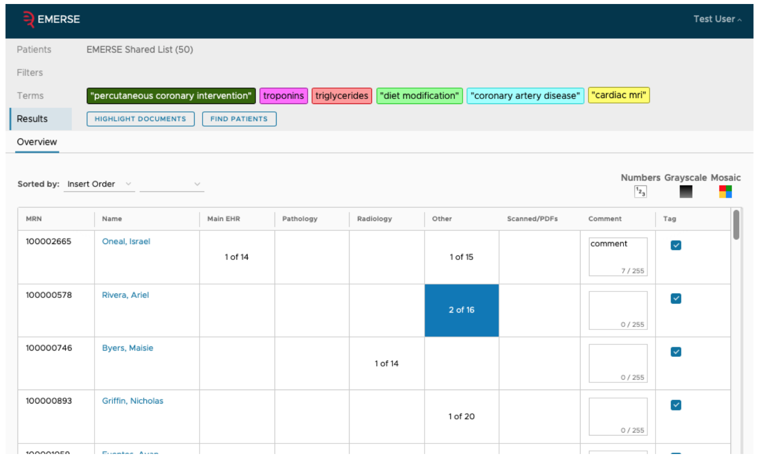

Figure 1. The *Highlight Documents* view in EMERSE

A major theme that emerged from interviews with experienced users was the desire to be able to use EMERSE for even more tasks and on more sources of data. Users indicated that reliance on another tool (e.g. the EHR for data not included in EMERSE, or DataDirect which was the locally used tool for structured data) to get to data that EMERSE does not index slowed their work and hindered usability. Integration, in one form or another, with more sources of data was positioned as a means to make the tool even more powerful and useful to them in their work. The desired forms of integration fell into three general categories: discrete data (labs, vitals, etc), scanned documents, and external site records through health information exchanges (HIEs). A large proportion of users interviewed noted that some form of integration with the EHR's database of discrete variables would make EMERSE even more powerful and increase its utility for them. This database of discrete variables includes lab results, vitals, and other tabular data stored in the EHR that EMERSE currently does not index because it is limited to free-text data. While EMERSE is accessible through the EHR at the University of Michigan by the integration of a button that takes users from the EHR to EMERSE, participants noted that reliance on another tool to get discrete data often slowed their workflow and presented the greatest limitation to the usability and utility of EMERSE in their work. Users also expressed desire for EMERSE to index and present results from more document types, including expansion of access to text attached to or embedded in CT, MRI, EKG, and other reports. However, the lack of some of these data sources is not a limitation of EMERSE itself, but a result of external factors. For example, the University of Michigan has not yet committed resources to run optical character recognition (OCR) on scanned documents, and external HIE documents are not permitted for research use under existing agreements.

Another theme that emerged was the promotion and distribution of the EMERSE system. Users unanimously responded that they would recommend the system to others. However, they repeatedly noted that the distribution of its use and awareness of its existence—even within the University of Michigan where it was developed—was not proportionate to its utility. That is, for a tool they found so useful, they believed that far too few people knew about its existence. Some users even suggested that if large national healthcare or insurance companies knew about EMERSE, they would certainly implement it as their main tool for data abstraction.

Phase 3: Qualitative analysis of observations and participant questions during system testing, transcripts of post-testing interviews, and participant responses in survey comment fields revealed four major themes regarding the participants' perception of and experience with the system. These themes were: (1) the expectation of a brief, but necessary adaptation period to become proficient with use of the system; (2) the utility the participants expected the system would provide in their respective roles; (3) the use and role of colors in the system; and (4) clarity and resetting of patient lists.

Theme 1: Participants expected to find the system even easier to use after a brief adaptation period. Many noted that even simply retaining the instructions for the basic scenarios used in this phase for one or two more uses would be sufficient to become completely familiarized with the interface and its functionality. The reasons given that a brief "adaptation period" would likely be necessary to become satisfactorily proficient with the interface were related mainly to navigation and the apparentness of affordances of system elements. For example, the ability to click a cell with a document count indicator in the table as shown in Figure 1 and be taken to an overview of that category of documents was an affordance that was not immediately apparent to a substantial number of users. A similar issue was noted with the *Patients*, *Filters*, *Terms*, and *Results* navigation elements, as seen in the top left of Figure 1. However, participants expressed that, on the whole, the system and interface were very intuitive.

Theme 2: During interviews, participants repeatedly expressed their perception that the EMERSE system would be highly useful to them. They also repeatedly expressed their desire to start using the system in their real work.

Theme 3: Many users addressed the accessibility of the interface, mainly in regards to colors and visual contrast. Positive feedback on the use of colors fell into the categories of interface aesthetics and color-coding. A number of users noted that the color scheme and flat design of the EMERSE interface made the system very visually pleasing and facilitated finding information and understanding its organization. Furthermore, color-coding search terms according to their Boolean logical relation (i.e. "*OR*" Boolean logic—or synonyms— indicated by terms of the same color; "*AND*" indicated by different colors) decreased cognitive load in performing the search task. Such color-coding increased users' understanding of information grouping and improved visual distinction and recognizability when scanning the interface for information and the occurrence of different terms in documents and query results. Critiques of color usage fell into the category of visual contrast, with a subset of these expressing the need for ADA-compliant colors schemes. Participants also expressed that lower visual contrast made it more difficult to understand the function of some interface elements. This contributed especially to the non-obvious affordances of the *Patients*, *Filters*, *Terms*, and *Results* navigation elements (Figure 1). Users also suggested that increasing the visual contrast of the *Find Patients* and *Highlight Documents* buttons when they are available to click would increase the usability of these core features by making their affordances more obvious.

Theme 4: Tasks related to patient lists, while mainly involving navigation elements in the system interface, emerged as a separate theme for two reasons: (1) these are issues that veteran users also noted (and thus did not fit into the category of navigation elements users believed they "just need to get used to"), and (2) understanding of patient lists is central to the EMERSE workflow. While users perceived the importance of patient lists, they raised concerns related to awareness of the current patient list, changing between patient lists, and resetting parameters for a new search across all available patients.

Quantitative Analysis

Results from quantitative analysis indicate that new users of the EMERSE system are able to complete basic but critical workflow tasks in the system with a high success rate (low error rate and low rates of confusion or needing help), are highly satisfied with the interface, and have highly positive perceptions of its expected utility. Analysis of the number of times exhibited behaviors in the *Fail, Help Request,* or *Confusion* category (Table 1) are summarized in Table 2. In the two most "troublesome" steps, 31% and 25% of users, respectively, exhibited a delay and/or searching behavior and only 56% and 69% of users, respectively, completed the step with no issues. Connecting the instructions of this step to the qualitative analysis reveals that these two outcomes were due to difficulty finding the location of the *Highlight Documents* button (Figure 1), confusion over the "clickability" and/or meaning of the document count indicator (Figure 1), or finding the *Saved Terms* tab within the *Terms* view. The step with the highest error rate had only 2 (9%) of the participants with error in completing the step. This step involves the removal of a search term as the first step to start a new query after completing a prior query. In connection with the themes that emerged from qualitative analysis of Phase 3 interview responses, the higher error rate on this step indicates that the affordance of the pencil icon in the corner of each search term is not immediately apparent as the means by which to remove that specific term. In connection with this, specific feedback from users indicated that, because the pencil icon also allows for editing of the term without removal of it, the addition of a button on each term such as an "x" or a "trash can" icon would be useful to separate the action of removing the term from the action of editing it, thus making the option for the user to remove individual terms more obvious. The two steps with the highest help request rate saw 2 (9%) of the participants requesting help on each of these steps. The first of these two steps represents the first instance of the use of the *Highlight Documents* button (see Figure 1), as well as the instruction to click on a specific cell containing a document count indicator (Figure 1). Qualitative analysis also reveals these interface elements were a source of confusion for users due to lack of visual distinction for the *Highlight Documents* button and the functionality of the cells in the *Highlight Documents* table as links not being immediately apparent. The second of these two steps represents the first time users are asked to explicitly change the *Patient List*. The requests for help were due to both an inability to find where options to explicitly change the patient list were located and confusion over why this was a necessary step to proceed with the new query. Of note is that 77% of *all steps* were completed by *all users* with no errors, 53% of *all steps* were completed by *all users* with *no help requests*, and 31% of *all steps* were completed by *all users* with *no issues*. Across all steps of the three scenarios, as summarized in Table 2 the average error rate was 2%, the average help request rate was 4%, and the average rate of completion with no issues was 88%.

Table 2. Summary of event occurrences in user testing scenarios.

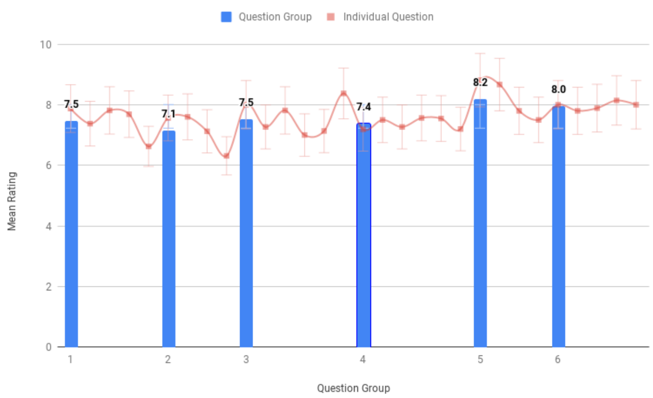

Figure 2. Quis mean response score by question group and individual question

Scenario	Errors (%)	Help Requests (%)	Search or Delay (%)	Corrected Mistakes (%)	*No Errors (%)*	*No Help or Errors (%)*	*No Issues (%)*
1	2	3	12	3	*98*	*95*	*90*
2	2	3	11	3	*98*	*96*	*90*
3	2	3	19	5	*97*	*95*	*85*
Overall	**2**	**3**	**14**	**4**	***98***	***95***	***88***

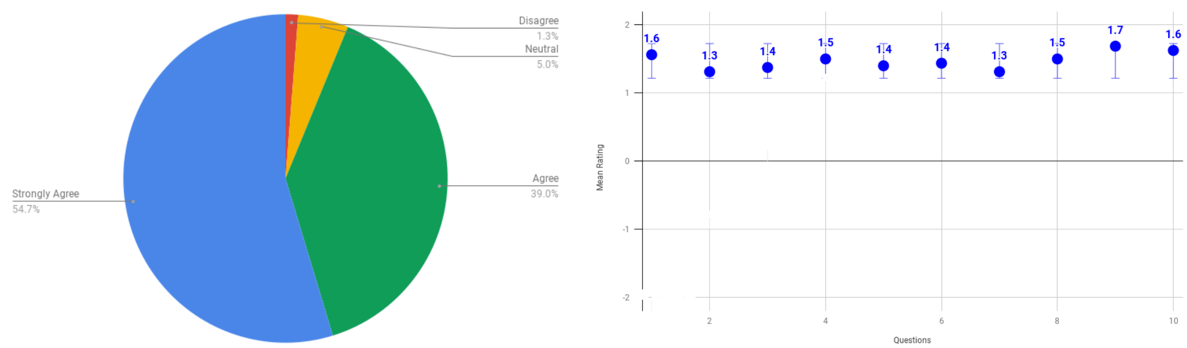

Figure 3. Summary Measures of Likert Scale responses for TAM Survey. Total response percentage of each answer category across all questions (left) and mean rating for each question (right).

QUIS survey results indicate very high ratings of usability and satisfaction after only one session of use. On the scale of 0 through 9 on the QUIS survey, users on average rated their overall impression of the software at 7.5 ± 1.3. The lowest-performing question was the ranking of the sequence of screens on the scale from "confusing" (0) to "very clear" (9), with an average rating of 6.3 ± 2.0. The highest-performing questions were both related to system capabilities. System speed was rated at 8.8 ± 0.4, and quality of search results was rated at 8.7 ± 0.7. Notably, these are core features of the EMERSE system's design. Also of note is that the highest-performing groups of results on the QUIS survey were the "Usability & UI" and "System Capabilities", rated at 7.9 ± 1.2 and 8.2 ± 1.0, respectively. Average scores of each question and each question group can be seen in Figure 2. TAM survey results indicate high acceptance of the EMERSE system. Likert Scale ratings were assigned the numerical values from -2 for *Strongly Disagree* to 2 for *Strongly Agree*. The average score across all 10 questions was 1.5 ± 0.1. The highest performing questions were the final two, asking whether users would find it easy to become skilled at using the system and whether they would find the system easy to use. These questions were rated at 1.7 ± 0.5 and 1.6 ± 0.5, respectively. The lowest rated question was in regards to whether users believed the system would improve their job performance, but was still rated at 1.3 ± 0.8. A summary of the ratings can be seen below (Figure 3).

Discussion

Themes and outcomes from the three phases of user evaluation will be discussed and evaluated in light of Norman's idea of affordances[21,30–32] and Nielsen's design heuristics[33–36].

User feedback on color, visual contrast, and visual distinction is largely aligned with the increasing awareness and concern for accessible and inclusive design[37–41]. As other websites and companies have sought to comply with ADA guidelines for internet accessibility, awareness of the benefits of accessible and inclusive design has trickled down to users who may not necessarily *"require"* accessible design. This is consistent with a key idea of accessible design: designing for inclusion at baseline will benefit *all* users[37–41]. A feature in EMERSE that hearkens to this principle is the color-coding of terms. This feature has shown highly positive reception among users not only because it improves visual contrast and clarity of information presented, but it also serves to clarify the logic of Boolean search terms that users less familiar with advanced search engine functionality[9] may not intuitively understand. Furthermore, this feature indicates the potential of color-coding and other similar uses of system colors in clinical information software to comply with Nielsen's 6th heuristic, which stresses the value of recognition over recall. In addition to the benefits of color-coding, positive feedback on interface colors and flat, minimalist design demonstrate the validity and importance of Neilsen's 8th heuristic in clinical software tools—an area where the field is currently sorely lacking.

Confusion surrounding patient lists is best understood and evaluated in light of Nielsen's 1st, 3rd, and 4th heuristics[36]. It is most useful here to begin with Nielsen's fourth heuristic: *Consistency and Standards*. Users are highly influenced by the norms of designs and functionalities seen in the numerous other systems they interact with on a daily basis in their personal and professional lives and "should not have to wonder whether different ... actions mean the same thing"[36]. While "EHR search is substantially different from Web search"[18], the persistence of patient lists across the changing of search terms may not be intuitive to users because of expectations set by other common search engines. That is, traditional search engines typically default to search across all available sources unless otherwise specified, whereas once a patient list is added or selected in EMERSE, it persists until it is explicitly removed. This brings us to the issues related to Nielsen's first heuristic: *Visibility of System Status*. Although in regards to patient lists EMERSE technically does meet the criteria of this heuristic by keeping a display of the name and patient count of the current patient list (Figure 1), it may be that this display does not *functionally* fulfill the criteria. That is, in order to supersede user assumptions, this display may need to be larger and "louder". This brings us finally to the third heuristic: *User control and Freedom*. In this heuristic, Nielsen highlights the importance of *Undo*, *Redo*, and *"Emergency Exit"* capabilities of the system. While EMERSE does this quite well in some other respects, participants often requested additional, more accessible, or more apparent "reset" buttons, either to completely restart their search, or to reset patient lists, filters, and/or search terms. While EMERSE provides options to restart the search or reset patient lists or terms, the options were considered either too hidden (as discussed below) or required more steps than desired.

Hindrances to immediate intuitiveness of the EMERSE system were related mainly to navigation and the apparentness of system element affordances. A commonality across themes was the non-obvious affordances of interface elements. The hidden nature of some shortcuts and customizations align with Nielsen's 7th and 8th heuristics of "Flexibility and efficiency of use" and "Minimalist design", but often these affordances were too well hidden (and hidden in plain sight) from users who desired to find them. In addition to "reset" buttons or options not being immediately apparent to users, participants in both Phases 2 and 3 often "requested" features that already existed, such as control over color-coding of terms. In these instances, "Minimalist Design" may have overshadowed the apparentness of system features. Furthermore, feedback related to the visual contrast of the *Find Patients* and *Highlight Documents* buttons highlights the particular importance of Nielsen's 1st heuristic in clinical software tools.

Desire for expanded document access reflects the current state of interoperability in U.S. healthcare, as information blocking practices and the segmentation of EHR systems and clinical databases is highly prohibitive to the free but secure flow of medical record information between locations. This is further complicated by pre-existing regulations and agreements which limit the degree to which data can be used even if interoperability mechanisms are in place, and that patients themselves may not actually desire some forms of interoperability. To use an applicable example, the University of Michigan has data sharing mechanisms in place with external locations that Phase 2 users desired to see data from, but current regulations and agreements prohibit their use in ways that would allow the data to be indexed by EMERSE. Nonetheless, tools like EMERSE with exceptional usability and user-friendliness that promote their utility for users, will become all the more important and necessary as interoperability legislation expands the scope of data upon which clinical databases (and consequently search engines and other CIE tools) can draw, since the increased capacity for data does not inherently bring with it the increased availability of useful information from that data[3–5,9].

Finally, while some assign the task of overcoming current limitations of current EHR-based search engines to NLP and AI techniques[2], EMERSE is able to overcome a number of these issues related to search engine result quality through its query methods[9] and effective, highly intuitive design without the potential drawback of NLP approaches. One issue related to search engine results[2], which was also noted by participants, has not yet been solved through the design of EMERSE and may require an NLP-based solution. This is the issue of negations[2], in which a search engine like EMERSE with highly literal results may return a "positive" hit for a term when the context would negate that hit as a true positive. For example, searching for patients whose charts contain the term *"Coronary Artery Disease"* may return as "false positives" patients whose charts include the literal string match, but actually read as *"no family history of Coronary Artery Disease"*. Consequently, a solution to this issue may still require the use or integration of an NLP-based functionality. The EMERSE team is working on including negation into the system and is anticipated to be completed within a year. EMERSE also supports a way to "exclude terms" which allows for certain exclusions that are not necessarily negation but could still result in false positives. This often occurs within the context of "boilerplate" language such as those from instructions or warnings (e.g., search for nausea but ignore the term in the context of "the patient was warned that nausea is a common side effect of this medication").

Limitations

One limitation of this study is the number of participants for Phase 3. However, these participants were drawn from multiple institutions, which helps improve the generalizability of our findings. Another limitation is that recruitment of participants was not completely random. Future studies could perform more random recruitment of participants for the types of testing done in Phases 2 and 3 of this study to decrease potential "positive user" bias that this study may have been subject to. However, this was a tradeoff made intentionally in this study: while high-frequency users (phase 2) and semi-volunteers (phase 3) may have a more positive bias or inclination toward the software, these pools were used for recruitment for phases 2 and 3 to increase participation and quality of feedback.

Conclusion

Overall, the EMERSE system performs very highly, was rated very positively by new and experienced users alike, and was met with great enthusiasm for its adoption in new sites. In preparing for implementation of the system at new sites, key concerns lie in the realms of obtaining organizational approval and the ability to effectively export free-text note data from the site's EHR so that they could be imported into EMERSE. While experienced users would unanimously recommend EMERSE to others and desire to see the continued expansion of the sources of documents indexed, they note that its use and utility is not as widespread and well-publicized as it could be. Participants currently using the system expressed the belief that expansion of the document sources indexed would further increase the utility of the system in their work, and that better promotion and advertisement of the EMERSE system would facilitate the adoption of EMERSE by more users. Novice users noted in testing the system that navigation and affordances of the interface were not always completely intuitive, but believed they could quickly gain proficiency and expressed enthusiasm for the adoption of the tool in their work. Finally, experienced and novice users of EMERSE alike noted issues of clarity regarding which set of patients is currently selected and being searched through. Potential reasons for and solutions to this issue have been discussed. In summary, lessons learned from this study indicate the importance of clarity of navigation and apparentness of affordances in the usability of clinical documentation search tools. While small improvements can still be made in these areas, the EMERSE system and the results of this study present a template for using system colors and a highly intuitive user interface to create high-utility CIE tools, which are desperately needed in a field dominated by low-usability EHRs.

References

1. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: A literature review. J Biomed Inform [Internet]. 2018 Jan 1 [cited 2020 Nov 9];77:34–49. Available from: http://www.sciencedirect.com/science/article/pii/S1532046417302563
2. Hill JR, Visweswaran S, Ning X, Schleyer TK. Use, Impact, Weaknesses, and Advanced Features of Search Functions for Clinical Use in Electronic Health Records: A Scoping Review. Appl Clin Inform [Internet]. 2021 May [cited 2021 Jul 28];12(03):417–28. Available from: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0041-1730033
3. Hanauer DA, Mei Q, Law J, Khanna R, Zheng K. Supporting Information Retrieval from Electronic Health Records: A Report of University of Michigan's Nine-Year Experience in Developing and Using the Electronic Medical Record Search Engine (EMERSE). J Biomed Inform [Internet]. 2015 Jun [cited 2021 Mar 9];55:290–300. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4527540/
4. Hanauer DA. EMERSE: The Electronic Medical Record Search Engine. AMIA Annu Symp Proc AMIA Symp. 2006;941.
5. Jha AK. The promise of electronic records: around the corner or down the road? JAMA. 2011 Aug 24;306(8):880–1.
6. Friedman C, Rigby M. Conceptualising and creating a global learning health system. Int J Med Inf. 2013 Apr;82(4):e63-71.
7. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. Sci Transl Med. 2010 Nov 10;2(57):57cm29.
8. University of Michigan PE. Publications Using EMERSE [Internet]. EMERSE. [cited 2021 Nov 10]. Available from: https://project-emerse.org/publications.html#more
9. Hanauer DA, Wu DTY, Yang L, Mei Q, Murkowski-Steffy KB, Vydiswaran Vgv, et al. Development and Empirical User-Centered Evaluation of Semantically-based Query Recommendation for an Electronic Health Record Search Engine. J Biomed Inform. 2017 Mar;67:1–10.
10. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. J Biomed Inform. 2004 Feb;37(1):56–76.
11. Yen P-Y, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc. 2012 May 1;19(3):413–22.
12. Zheng K, Vydiswaran VGV, Liu Y, Wang Y, Stubbs A, Uzuner Ö, et al. Ease of adoption of clinical natural language processing software: An evaluation of five systems. J Biomed Inform. 2015;
13. Friedberg MW, Chen PG, Van Busum KR, Aunon F, Pham C, Caloyeras J, et al. Factors Affecting Physician Professional Satisfaction and Their Implications for Patient Care, Health Systems, and Health Policy. Rand Health Q. 2014 Dec 1;3(4):1.
14. Babbott S, Manwell LB, Brown R, Montague E, Williams E, Schwartz M, et al. Electronic medical records and physician stress in primary care: results from the MEMO Study. J Am Med Inform Assoc. 2014 Feb 1;21(e1):e100–6.
15. Sinsky CA, Beasley JW, Simmons GE, Baron RJ. Electronic Health Records: Design, Implementation, and Policy for Higher-Value Primary Care. Ann Intern Med. 2014;
16. Hanauer DA, Barnholtz-Sloan JS, Beno MF, Del Fiol G, Durbin EB, Gologorskaya O, et al. Electronic Medical Record Search Engine (EMERSE): An Information Retrieval Tool for Supporting Cancer Research. JCO Clin Cancer Inform. 2020 Nov 1;(4):454–63.
17. Zheng K, Mei Q, Hanauer DA. Collaborative search in electronic health records. J Am Med Inform Assoc. 2011 May;18(3):282–91.
18. Yang L, Mei Q, Zheng K, Hanauer DA. Query log analysis of an electronic health record search engine. AMIA Annu Symp Proc AMIA Symp. 2011;2011:915–24.
19. Seyfried L, Hanauer DA, Nease D, Albeiruti R, Kavanagh J, Kales HC. Enhanced identification of eligibility for depression research using an electronic medical record search engine. Int J Med Inf. 2009 Dec;78(12):e13–8.
20. Creswell JW. Research design: qualitative, quantitative, and mixed methods approaches. 4th ed. Thousand Oaks: SAGE Publications; 2014. 273 p.
21. Chin JP, Diehl VA, Norman KL. Development of an instrument measuring user satisfaction of the human-computer interface. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems [Internet]. New York, NY, USA: Association for Computing Machinery; 1988 [cited 2021 Nov 12]. p. 213–8. (CHI '88). Available from: https://doi.org/10.1145/57167.57203

22. Magal S, Mirchandani D. Validation of the Technology Acceptance Model for Internet Tools. :7.
23. Lee Y, Kozar K, Larsen K. The Technology Acceptance Model: Past, Present, and Future. Technology. 2003 Jan 1;12.
24. Venkatesh V, Davis FD. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Manag Sci. 2000 Feb 1;46(2):186–204.
25. Maalej W, Kurtanović Z, Nabil H, Stanik C. On the automatic classification of app reviews. Requir Eng. 2016 Sep 1;21(3):311–31.
26. Johann T, Stanik C, B AMA, Maalej W. SAFE: A Simple Approach for Feature Extraction from App Descriptions and App Reviews. In: 2017 IEEE 25th International Requirements Engineering Conference (RE). 2017. p. 21–30.
27. Maalej W, Nabil H. Bug report, feature request, or simply praise? On automatically classifying app reviews. In: 2015 IEEE 23rd International Requirements Engineering Conference (RE). 2015. p. 116–25.
28. Marathe M, Toyama K. Semi-Automated Coding for Qualitative Research: A User-Centered Inquiry and Initial Prototypes. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems [Internet]. Montreal QC Canada: ACM; 2018 [cited 2021 Mar 12]. p. 1–12. Available from: https://dl.acm.org/doi/10.1145/3173574.3173922
29. Dąbrowski J, Letier E, Perini A, Susi A. Finding and Analyzing App Reviews Related to Specific Features: A Research Preview. In: Knauss E, Goedicke M, editors. Requirements Engineering: Foundation for Software Quality. Cham: Springer International Publishing; 2019. p. 183–9. (Lecture Notes in Computer Science).
30. Norman D. Emotion & Design: Attractive Things Work Better. Interact Mag. 2002 Jul 1;9:36–42.
31. Massaro DW. Review of The Psychology of Everyday Things. Am J Psychol. 1990;103(1):141–3.
32. Norman D. The psychology of everyday things. 1990;
33. Nielsen J, Molich R. Heuristic evaluation of user interfaces. In: Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI '90 [Internet]. Seattle, Washington, United States: ACM Press; 1990 [cited 2020 Nov 10]. p. 249–56. Available from: http://portal.acm.org/citation.cfm?doid=97243.97281
34. Nielsen J. Heuristic evaluation. In: Usability inspection methods. USA: John Wiley & Sons, Inc.; 1994. p. 25–62.
35. Nielsen J. Usability Engineering. Morgan Kaufmann; 1994. 382 p.
36. Experience WL in R-BU. 10 Usability Heuristics for User Interface Design [Internet]. Nielsen Norman Group. [cited 2021 Nov 13]. Available from: https://www.nngroup.com/articles/ten-usability-heuristics/
37. Zhu H, Gruber T, Dong H. Value and Values in Inclusive Design. In: HCI. 2020.
38. Microsoft Design. Microsoft Inclusive Design Toolkit [Internet]. Microsoft Design; [cited 2021 Nov 12]. Available from: https://www.microsoft.com/design/inclusive/
39. Microsoft Design [Internet]. [cited 2021 Nov 12]. Available from: https://www.microsoft.com/design/inclusive/
40. Persson H, Åhman H, Yngling A, Gulliksen J. Universal design, inclusive design, accessible design, design for all: different concepts—one goal? On the concept of accessibility—historical, methodological and philosophical aspects. Univers Access Inf Soc. 2014;
41. Burnett M. Doing Inclusive Design: From GenderMag in the Trenches to Inclusive Mag in the Research Lab. AVI. 2020;