Journal of the American Medical Informatics Association, 24(e1), 2017, e35–e39
doi: 10.1093/jamia/ocw092
Advance Access Publication Date: 3 July 2016
Research and Applications

Research and Applications

A framework for evaluating electronic health record vendor user-centered design and usability testing processes

Raj M Ratwani,[1,2] A Zachary Hettinger,[1,2] Allison Kosydar,[1] Rollin J Fairbanks,[1,2] and Michael L Hodgkins[3]

[1]National Center for Human Factors in Healthcare, MedStar Health, Washington, DC, [2]Department of Emergency Medicine, Georgetown University School of Medicine, Washington, DC, and [3]American Medical Association, Chicago, Illinois

Corresponding Author: Raj Ratwani, PhD, Scientific Director, National Center for Human Factors in Healthcare, MedStar Health, 3007 Tilden St, Suite 7M, Washington, DC 20009, USA. E-mail: raj.ratwani@medicalhfe.org; Tel: 202-244-9815

Received 30 November 2015; Revised 17 April 2016; Accepted 9 May 2016

ABSTRACT

Objective: Currently, there are few resources for electronic health record (EHR) purchasers and end users to understand the usability processes employed by EHR vendors during product design and development. We developed a framework, based on human factors literature and industry standards, to systematically evaluate the user-centered design processes and usability testing methods used by EHR vendors.

Materials and Methods: We reviewed current usability certification requirements and the human factors literature to develop a 15-point framework for evaluating EHR products. The framework is based on 3 dimensions: user-centered design process, summative testing methodology, and summative testing results. Two vendor usability reports were retrieved from the Office of the National Coordinator's Certified Health IT Product List and were evaluated using the framework.

Results: One vendor scored low on the framework (5 pts) while the other vendor scored high on the framework (15 pts). The 2 scored vendor reports demonstrate the framework's ability to discriminate between the variabilities in vendor processes and to determine which vendors are meeting best practices.

Discussion: The framework provides a method to more easily comprehend EHR vendors' usability processes and serves to highlight where EHR vendors may be falling short in terms of best practices. The framework provides a greater level of transparency for both purchasers and end users of EHRs.

Conclusion: The framework highlights the need for clearer certification requirements and suggests that the authorized certification bodies that examine vendor usability reports may need to be provided with clearer guidance.

Key words: electronic health records, usability, human factors, health information technology

INTRODUCTION

Background and Significance. With clinicians' widespread use of electronic health records (EHRs), dissatisfaction with EHRs has become more apparent and the usability of this technology has been identified as a major challenge.[1–3] Many EHRs are not prospectively designed, developed, and implemented to support the cognitive needs of clinicians, resulting in stress, frustration, reduced efficiency, and patient safety hazards.[4,5] In an effort to promote EHR usability, the Office of the National Coordinator for Health Information Technology (ONC) established *safety-enhanced design* (SED) certification requirements that require EHR vendors to attest to a user-centered design (UCD) process and conduct a formal (summative)

usability test on 8 EHR capabilities in order to certify their products.[6] EHR vendors' SED reports describing their UCD process, testing methodology, and testing results are required to be made public on the ONC's Certified Health IT Product List (CHPL) for all vendors that meet the certification requirements.

However, there is tremendous variability in the UCD processes and testing methodologies used by vendors. In a previous study, we examined the usability processes of 11 EHR vendors and found that some vendors do not have any true UCD process in place, while others have sophisticated processes with dedicated usability staff.[7] Analysis of the SED reports from vendors has shown that many vendors do not attest to a UCD process and are not meeting industry-accepted usability testing standards, yet these products are still certified has having met the requirements.[8]

Currently, there is the perception of a lack of transparency around the usability processes employed by vendors and the usability of EHR products. One purpose of the ONC making the SED reports publicly available is to allow purchasers and users of EHRs to understand the usability processes that vendors employ. A rigorous UCD process and usability testing methodology are important in developing usable technology.[9,10] Unfortunately, these reports are complex and difficult to comprehend, with some reports spanning over 100 pages, and with inconsistent presentation of information and a lack of context as to what information is important and how the vendor's process compares to evidence-based standards. For people who are not well trained in the science of usability, the reports do not provide information in an easily digestible format that can inform purchasing decisions, and for those who are versed in usability science, interpreting the reports is time-consuming. This lack of transparency contributes to purchasers selecting EHR products that have suboptimal usability and also hinders competition in the marketplace.

While independent usability evaluations of fully developed EHR products would be ideal for market transparency and information on product fit, this type of evaluation is a significant challenge due to a lack of access to EHR systems to conduct end-product usability assessments. The publicly available SED reports, while limited in scope, do present an opportunity to gain insight into vendor usability processes if the information in these reports can be distilled into scientifically supported criteria.

Objective. Our goal was to develop a standardized, evidence-based framework for evaluating the UCD process, summative testing methodologies, and summative testing results of EHR vendors based on their required submissions for ONC certification. By examining the details that EHR vendors are required to report and comparing this information to industry standards and best practices that that have been validated in the human factors and usability literature, the framework will provide an appropriate context in which to determine the rigor of EHR vendors' UCD and testing processes relative to the certification requirements in a standardized format.

The framework, based on certification requirements and supporting human factors data found in the psychology and industrial systems engineering literature, is described in the Methods section, along with the specific scoring process. In the Results section, we describe how the framework is applied to 2 EHR SED reports that were retrieved from the ONC CHPL site.

MATERIALS AND METHODS

We first reviewed the ONC 2014 certification requirements to determine the specific UCD process, summative testing methodology, and summative testing results information that EHR vendors are required to report as part of SED. We then examined the human factors and usability literature to determine the standards for each of the required data elements to be reported by vendors. Based on this information, we formulated a framework for evaluating the vendor UCD and testing process.

Table 1. List of usability relevant certification requirements and identified dimensions

Certification Requirements	Mapped Usability Process Dimension
State UCD process	UCD Process
Report total number of participants in summative test	Summative Testing Methodology
Provide a description of participants including experience and demographic characteristics	
Describe the user tasks that were tested	
List specific metrics to test effectiveness, efficiency, and satisfaction	
Report results of the test, including major test findings	Summative Testing Results
Specify areas for improvement	

Safety-enhanced design certification requirements. The ONC 2014 certification requirements stipulate that vendors must attest to a UCD process by providing a statement of the process, and must conduct a summative usability test on 8 EHR capabilities.[6] The method and results of the summative usability test must be reported using the National Institute for Standards and Technology Customized Common Industry Format Template for Electronic Health Record Usability Testing (NISTIR 7742).[11] Usability-relevant aspects that vendors are required to report are listed in the left column of Table 1.

The information that vendors are required to report was segmented into 3 dimensions (right column of table 1): UCD process, summative testing methodology, and summative testing results. These 3 dimensions were formulated based on the information that the ONC requires EHR vendors to report. We examined the literature on UCD and summative testing to determine the existing guidance and validated standards for each component of each dimension.

Usability process standards and scoring criteria. Based on a review of the usability and human factors literature, we identified the appropriate usability process standards for each component (Table 2). We then created scoring criteria for each usability process dimension and subcomponent. Five points were assigned to each dimension, for a total possible score of 15 points for each vendor report. Within each dimension certain factors were weighted depending on the importance of the factor (rationale for the weighting is described below).

For the UCD process dimension, the ONC certification requirement is that vendors provide a statement of their process. If vendors provided this statement, 5 points were assigned for this dimension; if not, no points were assigned.

For the summative testing methodology dimension, there were 4 subcomponents, and points were assigned as follows:

Total number of participants: Usability researchers have found that using 15 participants in summative usability testing leads to the discovery of 90% of usability design challenges.[12] Given the

Table 2. Summary of usability recommendations and scoring criteria

Dimension	Subcomponents	Recommendation	Scoring Criteria
UCD process	Not applicable	A user-centered design process puts the needs of the user at the forefront of design and development, resulting in a product that is more likely to meet users' needs.[9,10]	5 pts for statement of UCD process 0 otherwise
Summative Testing Methodology	Number of participants	Fifteen participants will reveal >90% of the problems when conducting summative testing.[12]	1 point for 15+ participants 0.5 point for 10–14 participants 0 points for <10 participants
	Clinical background	Participants should represent the end-user demographic of the product.[10,14,17]	2 points if all clinical 0.5 point for at least 1 clinical 0 points for no clinical
	Use case rigor	The use cases should be as representative as possible of the use cases in the live environment. Use cases should allow evaluation of clinical and usability aspects and include challenging scenario elements.[10,13]	1 point for detailed use cases 0.5 point for vague use cases 0 points for no use case description
	Appropriate measures	Usability measure of effectiveness, efficiency, and satisfaction should be used in summative testing.[15] • Effectiveness: percent of correct actions[11,15] • Efficiency: time on task[14] • Satisfaction: SUS or SUMI[14]	1 point if all measures accurately captured 0 points if any measure is not accurately captured
Summative Testing Results	Percent effectiveness	Success rate for first-time users during summative testing should be 80–95%.[16]	3 points if effectiveness is ≥80% 0 points if <80%
	Areas for improvement identified	Detailed areas for improvement should be provided to drive the next iteration of design and development.[10,15]	2 points if substantive description 0.5 point if minimally addressed 0 points if no information

safety-critical functions that are being tested as part of the ONC's requirements, it is important to have a sample size that will achieve the 90% threshold. A maximum of 1 point was assigned if 15 or more participants were used, on average, across all the capabilities tested by the EHR vendor. Ten participants captured 80% of design challenges, thus half a point was assigned if 10–14 participants were used. If fewer than 10 participants were used or no information was reported, then zero points were assigned.

Expertise of participants: The participants used in summative testing should represent the end users of the system being tested.[10,13,14] Our focus was EHR capabilities that are used by clinicians. Consequently, participants in vendor summative tests should have a clinical background, which we operationally defined as any role that involves contact with patients, such as physician, nurse, technician, or medical assistant. Using participants with the appropriate background is a critical aspect of usability testing, and we reflected this in the scoring process by weighting this factor. If all the participants in the clinical capabilities that were tested had a clinical background, 2 points were assigned. If at least 1 participant had a clinical background, half a point was assigned, and no clinical participants or failure to report resulted in zero points.

Use cases: Summative usability testing should include tasks and use cases that are representative of real-world uses of the system being tested and that test relevant aspects of the system.[10,13] If vendors used detailed use cases that were clinically appropriate, 1 point was assigned. If vague use cases were provided, half a point was assigned, and if no use case information was provided, zero points were assigned.

Appropriate metrics: Effectiveness, efficiency, and satisfaction during summative usability testing are important outcomes variables that are critical to understanding the usability of a product.[15] Effectiveness is traditionally measured by examining task success as the number of correctly completed tasks divided by the number of attempted tasks.[11,15] Efficiency metrics examine the time to complete a task.[14] Satisfaction metrics include either the system usability scale or the Software Usability Measurement Inventory.[14] The appropriate use of these metrics is explicitly stated in the NISTIR 7742 template that vendors are required to use to report their results.[11] If vendors appropriately measured effectiveness, efficiency, and satisfaction, 1 point was assigned, and if any measures were not appropriately used, zero points were assigned.

For the summative results dimension, there were 2 subcomponents and points were assigned as follows:

Measure of effectiveness: Following a UCD process and using first-time users in a summative usability test, effectiveness ratings are expected to be ≥80%.[16] If vendors reported an average effectiveness value across all tested use cases for the capabilities tested of ≥80%, 3 points were assigned, and if the effectiveness rating was <80%, zero points were assigned. This subcomponent was weighted because it reflects the application of a UCD process throughout design and development.

Identified areas for improvement: Recognizing system deficiencies and identifying areas for improvement in the software

undergoing testing is a major outcome of the testing process and is important for advancing usability.[10,15] If there was a substantive discussion of areas for improvement in the vendor report, 2 points were assigned. If areas for improvement were minimally addressed, half a point was assigned, and no discussion of areas for improvement resulted in zero points.

Measures of efficiency and satisfaction were not included in the scoring process because there are no clear guidelines or standards on expected efficiency or satisfaction outcomes following summative testing that could be applied to the EHR vendor reports.

RESULTS

We demonstrate use of the scoring process for 2 EHR SED vendor reports that were retrieved from the ONC's CHPL website, which is the official list of EHR products for meaningful use. Each report was first examined by a human factors scientist (PhD), a clinician (MD) trained in informatics, and a research assistant (BS). The human factors scientist extracted information from the reports and assigned points based on the criteria described in the methods section for all dimensions and subcomponents, excluding the description of the use cases. The research assistant also reviewed each report to confirm the scoring that was assigned by the human factors scientist. The use cases were extracted from the reports by the research assistant and provided to the clinician, who was blinded to the vendor names. The clinician was given the scoring criteria and asked to evaluate the use cases and assign scores based on the rigor of the use case.

To examine the consistency of applying the framework, 2 additional committees, each composed of a clinician (MD) and a human factors expert (MS), were provided with a description of the framework and applied the framework to the 2 vendor reports. The clinician scored the use cases and the human factors expert coded the other components. The committees scored the reports independently and were blinded to the initial scoring. The extracted information and scores from the committees were compared to the original scores from the first review to assess the consistency of application. Application of the framework across the 3 committees was completely identical in terms of information extracted and points assigned.

UCD process dimension. Vendor 1 reported following the NISTIR 7741 UCD process, while vendor 2 did not state a UCD process in its report. Vendor 1 was assigned 5 points and vendor 2 was assigned zero points.

Summative usability testing methodology dimension. The testing methodology in each report was examined to extract the subcomponents of this dimension and assign the scores.

Number of participants. Across the capabilities tested and the specific tasks, vendor 1 averaged 18.2 participants and was assigned 1 point, while vendor 2 averaged 4 participants and was assigned zero points.

Clinical background of participants. All the participants used by vendor 1 had a clinical background, and the vendor was assigned 2 points. Fifty percent of vendor 2's participants had a clinical background, and the vendor was assigned half a point.

Use case rigor. The use cases used by vendor 1 were rated as rigorous, and the vendor was assigned 1 point, while vendor 2 did not describe use cases, resulting in no points assigned.

Appropriate metrics. Both vendors measured efficiency, effectiveness, and satisfaction appropriately, resulting in 1 point being assigned.

Summative usability testing results dimension. The summative testing results were examined for the average effectiveness across all tasks tested and for discussion of areas of improvement.

Effectiveness. Vendor 1's average effectiveness was 84.8 and vendor 2's effectiveness was 82.1, both above the 80% threshold, resulting in the assignment of 3 points.

Areas for improvement. Vendor 1 provided a detailed list of areas of improvement that were discovered through the usability testing, for an assignment of 2 points. Vendor 2 provided little detail, resulting in scant identification of areas of improvement and half a point being assigned.

Out of 15 possible points, vendor 1 was assigned 15 points and vendor 2 was assigned 5 points. Table 3 provides a summary of the points that were assigned to each vendor and the data from the framework for each dimension and subcomponent. The table also shows the information extracted from each of the 3 committees to demonstrate the consistency of applying the framework.

DISCUSSION

The framework we have developed utilizes existing vendor SED reports, as required for certification by the ONC, to systematically examine vendor UCD and summative testing processes. By identifying the SED certification requirements and aligning them with standards that are recognized in the human factors literature, the framework provides a method to quickly understand and compare vendor usability processes based on publicly available CHPL reports.

The EHR UCD framework provides a method for purchasers and end users to better understand the usability processes of EHR vendors and can serve as a method to highlight where EHR vendors are falling short. The consistency in application by the 3 committees demonstrates clarity of the framework and ease of application. Looking at the 2 vendors that were coded and described in the results section, it is clear that there are still vendors that are not utilizing a rigorous UCD process when compared to industry best practices. We have applied this framework more broadly to 20 EHR vendor products.

One of the limitations of the framework is that it is based on the reported UCD process, summative testing methodology, and summative testing results as provided in the SED certification reports that are self-reported by each vendor. The scores reflect the UCD and testing processes based on these reports and do not reflect the usability of the actual vendor EHR product. In addition, the information we have about the UCD process, testing methodology, and testing results is limited by what is required by the certification requirements and available in the SED report from the CHPL website provided by the ONC. While we had 3 committees apply the framework to 2 vendor reports and this resulted in completely identical information and point assignments, this was a limited application; for statistical validity, a broader application will be required.

Conclusion. There are several implications that follow from the framework and resulting EHR vendor scores. The demonstrated variability in scores suggests that the certification requirements from the ONC may need to explicitly state particular requirements, such as the number of participants, the demographics of participants, and how to measure effectiveness. The ONC has proposed expansion of

Table 3. Vendor information and point assignment across the 3 dimensions

		UCD Process	Summative Testing Methodology				Summative Testing Results		Total Score
		Reported process	No. of participants	Clinical background	Use case rigor	Appropriate metrics	Effectiveness	Areas for Improvement	
Vendor 1 Report	Committee 1	*5 pts*	*1 pt*	*2 pts*	*1 pt*	*1 pt*	*3 pts*	*2 pts*	*15 pts*
		NISTIR 7741	18.2	100%	Detailed	Yes	84.8	Detailed	
	Committee 2	*5 pts*	*1 pt*	*2 pts*	*1 pt*	*1 pt*	*3 pts*	*2 pts*	*15 pts*
		NISTIR 7741	18.2	100%	Detailed	Yes	84.8	Detailed	
	Committee 3	*5 pts*	*1 pt*	*2 pts*	*1 pt*	*1 pt*	*3 pts*	*2 pts*	*15 pts*
		NISTIR 7741	18.2	100%	Detailed	Yes	84.8	Detailed	
Vendor 2 Report	Committee 1	*0 pts*	*0 pts*	*0.5 pts*	*0 pts*	*1 pt*	3 pts	*0.5 pts*	*5 pts*
		No stated process	4	50%	No details	Yes	82.1	Scant	
	Committee 2	*0 pts*	*0 pts*	*0.5 pts*	*0 pts*	*1 pt*	3 pts	*0.5 pts*	*5 pts*
		No stated process	4	50%	No details	Yes	82.1	Scant	
	Committee 3	*0 pts*	*0 pts*	*0.5 pts*	*0 pts*	*1 pt*	3 pts	*0.5 pts*	*5 pts*
		No stated process	4	50%	No details	Yes	82.1	Scant	

the SED requirements to include a minimum of 10 participants; however, the other criteria are not explicitly stated. In addition, the ONC's authorized certification bodies may need to have explicit guidelines on what constitutes a sufficient UCD process and rigorous summative usability testing. This framework can be used by the ONC to further improve the certification program and serve as a method to track improvements in EHR vendor products.

FUNDING

This work was funded by a contract between the American Medical Association and Raj Ratwani, PhD.

COMPETING INTERESTS

The authors have no competing interests to declare.

CONTRIBUTOR

R.R. led the study design and execution, and all authors supported this effort. A.Z.H and A.K. reviewed literature to support the framework. All authors contributed to writing the manuscript and approved the final version.

ACKNOWLEDGMENTS

We would like to thank Kate Kellogg, MD, Jessica Howe, MS, and Natalie Benda, MS, for assistance in scoring the vendor reports.

REFERENCES

1. Likourezos A, Chalfin DB, Murphy DG, Sommer B, Darcy K, Davidson SJ. Physician and nurse satisfaction with an Electronic Medical Record system. *J Emerg Med*. 2004;27(4):419–424.
2. McGinn CA, Grenier S, Duplantie J, *et al*. Comparison of user groups' perspectives of barriers and facilitators to implementing electronic health records: a systematic review. *BMC Med*. 2011;9(1):46.
3. Friedberg MW, Chen PG, Van Busum KR, *et al*. *Factors Affecting Physician Professional Satisfaction and Their Implications for Patient Care, Health Systems, and Health Policy* RAND Corporation: Santa Monica, CA; 2013.
4. Meeks DW, Smith MW, Taylor L, Sittig DF, Scott JM, Singh H. An analysis of electronic health record-related patient safety concerns. *J Am Med Inform Assoc*. 2014; 21:1–7.
5. Benda NC, Meadors ML, Hettinger AZ, Ratwani RM. Emergency physician task switching increases with the introduction of a commercial electronic health record. *Ann Emerg Med*. 2015. http://www.sciencedirect.com/science/article/pii/S019606441501104X. Accessed October 13, 2015.
6. Department of Health and Human Services. *Health Information Technology: Standards, Implementation Specifications, and Certification Criteria for Electronic Health Record Technology, 2014 Edition* Vol. 77. Washington, DC: Federal Register; 2012.
7. Ratwani RM, Fairbanks RJ, Hettinger AZ, Benda N. Electronic health record usability: analysis of the user centered design processes of eleven electronic health record vendors. *J Am Med Inform Assoc*. 2015; 22:1179–1182.
8. Ratwani RM, Benda N, Hettinger AZ, Fairbanks RJ. Electronic health record vendor adherence to usability certification requirements and testing standards. *JAMA*. 2015;314(10):1070–1071.
9. Norman D, Draper S. *User Centered Design* L. Erlbaum Associates, Hillsdale; 1986.
10. Nielsen J. *Usability Engineering* San Diego, CA: Academic Press; 1993.
11. Schumacher RM, Lowry SZ. *Customized Common Industry Format Template for Electronic Health Record Usability Testing Testing (NIST 7742)* the National Institute for Standards and Technology: Gaithersburg, MD; 2010.
12. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. *Behav Res Methods, Instruments, Comput*. 2003;35(3):379–383.
13. Kushniruk AW, Patel VL. Cognitive and usability engineering methods for the evaluation of clinical information systems. *J Biomed Inform*. 2004;37(1):56–76.
14. Sauro J, Lewis JR. *Qunaitifying the User Experience: Pracitcal Statistics for User Research*. Morgan Kaufman, Elsevier; 2004.
15. Tullis T, Albert B. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Morgan Kaufman, Elsevier; 2013.
16. Weinger MB, Wiklund ME, Gardner-Bonneau DJ, eds. *Handbook of Human Factors in Medical Device Design*. Taylor and Francis Group, CRC Press; 2010.
17. Kushniruk A, Patel VL, Cimino JJ. Usability testing in medical informatics: cognitive approaches to evaluation of information systems and user interfaces. In: *Proceedings of the 1997 AMIA Annual Fall Symposium*. Nashville, TN; 1997:218–222.