

NISTIR 7804

Technical Evaluation, Testing, and Validation of the
Usability of Electronic Health Records

Svetlana Z. Lowry
Matthew T. Quinn
Mala Ramaiah
Robert M. Schumacher
Emily S. Patterson
Robert North
Jiajie Zhang
Michael C. Gibbons
Patricia Abbott

NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

**Technical Evaluation, Testing, and Validation of the
Usability of Electronic Health Records**

Svetlana Z. Lowry
Matthew T. Quinn
Mala Ramaiah
National Institute of Standards and Technology

Robert M. Schumacher
User Centric, Inc.

Emily S. Patterson
Ohio State University

Robert North
Human Centered Strategies, LLC

Jiajie Zhang
University of Texas – Houston

Michael C. Gibbons
Patricia Abbott
Johns Hopkins University

February 2012

U.S. Department of Commerce
John E. Bryson, Secretary

National Institute of Standards and Technology
Patrick D. Gallagher, Under Secretary for Standards and Technology and Director



Acknowledgments

The authors gratefully acknowledge the contributions of Ronald Kaye (Food and Drug Administration), Molly F. Story (Food and Drug Administration), Quynh Nhu Nguyen (Food and Drug Administration), Ben-Tzion Karsh (University of Wisconsin-Madison), Muhammad Walji (University of Texas-Houston), Debora Simmons (University of Texas-Houston), Ben Shneiderman (University of Maryland), David H. Brick (Medical School New York University), Kirsten Jerch (User Centric), Janey Barnes (User View), Amy Franklin (University of Texas-Houston), Krisanne Graves (University of Texas-Houston), and Emile Morse (National Institute of Standards and Technology) to the development and expert review of this report. The authors appreciate the many informative comments from a broad spectrum of stakeholders on the Draft Electronic Health Record (EHR) Usability Protocol (EUP) during the public comment period, as well as numerous reviewers for their constructive feedback on prior drafts of the document.

Disclaimer: Certain commercial entities, equipment, or materials may be identified in this report in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

Table of Contents

Table of Contents	4
1 Executive Summary	7
2 Concept for an EHR Usability Protocol (EUP)	10
2.1 General Approach	12
2.2 EUP Foundation: Research Findings Defining EHR Usability Issues and Their Impact on Medical Error (EHR Patient Safety Model)	13
2.4 What the EHR Usability Protocol Is Not	29
3 Step I: EHR Application Analysis	30
3.1 Describe Application Scenarios of Use, User Tasks, and Application Interface Interaction	30
3.2 Define and Describe Application User Groups	30
3.3 Describe Application Environment of Use	30
3.4 Identify Critical Safety-Related Tasks	30
3.5 Formative and/or Preliminary Usability Assessment Overview	31
4 Step II: EHR Interface Expert Review	32
4.1 Overview of Interface Expert Review Process	32
4.2 Overview of Design Principles Used in Reviews	32
4.3 Selection and Composition of the Review Team	33
4.4 Review Process Overview	34
5 Step III: EHR User Interface Validation Test	36
5.1 Purpose of Summative Validation Testing and General Approach.....	36
5.2 Overview of Validation Testing Steps	40
6 Conclusion	57
Appendix A: Government Best Practices of Human-System Integration	58
Appendix B: Form for Expert Review	64
Expert Review of EHR	65
The system should protect the user and patient from potential use errors. Items 1A through 1H are principles of good design that help identify areas that might engender use error in EHRs. Items 2-13 are general principles of good user interface design.....	66
1A. Patient identification error	66
Actions are performed for one patient or documented in one patient’s record that were intended for another patient.....	66
1B. Mode error	68
Actions are performed in one mode that were intended for another mode.....	68
1C. Data accuracy error	69
Displayed data are not accurate.....	69
1D. Data availability error	71
Decisions are based on incomplete information because related information requires additional navigation, access to another provider’s note, taking actions to update the status, or is not updated within a reasonable time. 71	
1E. Interpretation error	72
Differences in measurement systems, conventions and terms contribute to erroneous assumptions about the meaning of information.....	72

1F. Recall error	73
Decisions are based on incorrect assumptions because appropriate actions require users to remember information rather than recognize it.	73
1G. Feedback error	74
Decisions are based on insufficient information because lack of system feedback about automated actions makes it difficult to identify when the actions are not appropriate for the context.	74
1H. Data integrity error	75
Decisions are based on stored data that are corrupted or deleted.	75
2. Visibility of System Status	76
The system should always keep the user informed about what is going on, through appropriate feedback within reasonable time.	76
3. Match between System and the Real World	77
The system should follow the user’s language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.	77
4. User Control and Freedom	78
Users should be free to select and sequence tasks (when appropriate), rather than having the system do this for them. Users often choose system functions by mistake and will need a clearly marked “emergency exit” to leave the unwanted state without having to go through an extended dialogue. Users should make their own decisions (with clear information) regarding the costs of exiting current work. The system should support undo and redo.	78
5. Consistency and Standards	79
Users should not have to wonder whether different words, situations or actions mean the same thing. Follow platform conventions.	79
6. Help Users Recognize, Diagnose and Recover From Errors	80
Error messages should be expressed in plain language (NO CODES).	80
7. Error Prevention	81
Even better than good error messages is a careful design that prevents a problem from occurring in the first place. 81	
8. Recognition Rather Than Recall	82
Make objects, actions and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.	82
Accelerators-unseen by the novice user-may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions. Provide alternative means of access and operation for users who differ from the “average” user (e.g., physical or cognitive ability, culture, language, etc.)	83
10. Aesthetic and Minimalist Design	84
Dialogues should not contain information that is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.	84
11. Help and Documentation	85
Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user’s task, list concrete steps to be carried out, and not be too large.	85
12. Pleasurable and Respectful Interaction with the User	86
The user’s interactions with the system should enhance the quality of her or his work-life. The user should be treated with respect. The design should be aesthetically pleasing- with artistic as well as functional value.	86

13. Privacy	87
The system should help the user to protect personal or private information belonging to the user or his/her patients.	87
Appendix C: Scenario 1: Ambulatory Care – Chronic Complex Patient; Mid-Level Provider	88
Appendix D: Scenario 2: Inpatient Care – Cardiac Patient; Physician	91
Appendix E: Scenario 3 Critical Care – Cardiac Patient; Nurse	94
Appendix F: Recruitment Screener Adapted from NISTIR 7742.....	95
Appendix G: Example Tester’s Guide.....	98
Appendix H: Informed Consent and Non-Disclosure Forms Adapted from NISTIR 7742 ...	105
Appendix I: Usability Ratings.....	108
Appendix J: Incentive Receipt and Acknowledgment Form.....	109
Glossary of Acronyms	110
Further Reading.....	111

1 Executive Summary

This document summarizes the rationale for an Electronic Health Record (EHR) Usability Protocol (EUP) and outlines procedures for design evaluation and human user performance testing of EHR systems. The procedures include general steps and guidance for evaluating an EHR user interface from clinical and human factors perspectives, and for conducting a validation study (i.e., summative usability test) of EHR user interfaces with representative user groups performing realistic tasks.

The document begins with a brief overview of the context for this guidance: Why is EHR usability critical? Usable EHRs have the potential to reduce “use errors” and improve patient care.¹ The International Organization for Standardization’s definition of usability is “The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use.”² The purpose of this proposed usability protocol is to encourage user-centered development processes focused on safety by facilitating the design of EHR interfaces with good usability. The authors of this document seek to make EHRs safer by providing methods to measure and validate user performance prior to deployment. Moreover, the authors hope to encourage system developers to apply human factors best practices and incorporate user-centered design processes into the development and deployment of EHR systems. Such practices and processes have a proven record in industries such as aviation, military systems, transportation, and nuclear power.

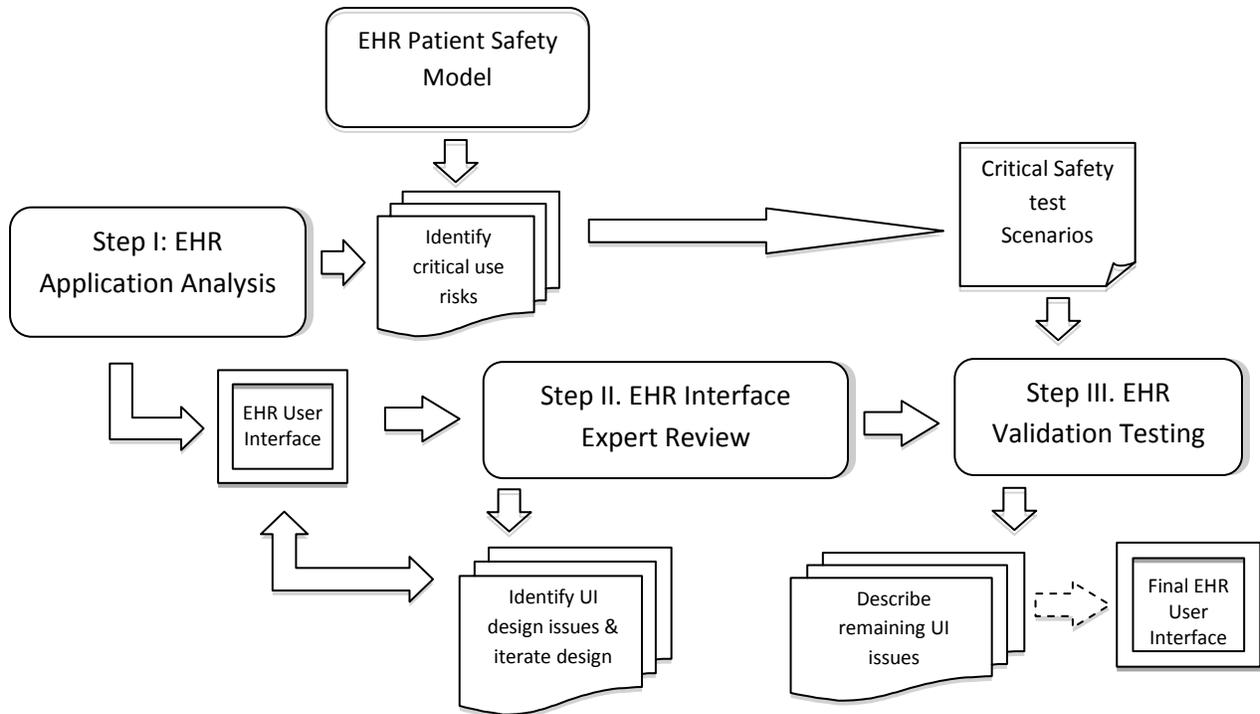
We include a detailed description of research findings relating to usability issues and their relationship to patient safety. This research has resulted in the development of a model for understanding the relationship between usability and patient safety outcomes, presented in Section 2 of this document. This model provides the foundation for the evaluation and testing process outlined in the EUP.

The EUP is a three-step process represented in Figure 1: (1) EHR Application Analysis, (2) EHR User Interface Expert Review, and (3) EHR User Interface Validation Testing (See Figure 1).

¹ “Use error” is defined by ANSI/AAMI HE75 as “an act or omission of an act that has a different result to that intended by the manufacturer or expected by the operator.” It is a term used very specifically to refer to cases where poorly designed user interfaces cause users to make errors of commission, where actions are erroneously taken, or errors of omission, where actions are erroneously omitted. It is true that human error can and does occur, but many errors are attributable not to the user *per se* but rather to designs that are flawed, e.g., poorly written messaging, misuse of color-coding conventions, omission of relevant information, etc.

² ISO/IEC. (1998). *Ergonomic requirements for office work with visual display terminals (VDT)s - Part 14 Menu dialogue* (ISO/IEC 9241-14).

Figure 1. EUP Three-Step Process



Step 1: EHR Application Analysis: The Application Analysis is both a key component of user-centered development processes³ and the foundation for the EUP and all subsequent analysis and testing activities. The Application Analysis can rely heavily, where applicable, on the EHR application developer’s user requirements and system requirements analysis. These elements include a description of the application’s basic functions, analysis of the user characteristics, task analysis describing the interactions between users and the application, analysis of the anticipated environment of use related to interactions with the application, and the identification of critical user tasks related to aspects of patient safety. A basic description of the application’s user interface should also be included in this step to facilitate the process of the subsequent EHR Interface Expert Review (Step 2). The Application Analysis should provide a description of the design of the application’s user interface and how the design has been optimized via iterative formative and exploratory usability assessments during development.

³ “NIST Guide to the Processes Approach for Improving the Usability of Electronic Health Records” (NISTIR 7741) provides NIST guidance for those developing electronic health record (EHR) applications following user-centered design (UCD) processes. An established UCD process ensures that EHRs are efficient, effective, and satisfying to the user. The main purpose of this guide is to provide practical guidance on methods relating to UCD and usability testing. The intended audiences of this document are those with a role in determining the features and functions contained in the EHR and how those are represented in the user interface.

Step 2: EHR User Interface Expert Review: The Expert Review is conducted by a combination of the vendor’s development team and a dedicated team of clinical safety and usability experts. The evaluators compare the EHR’s user interface design to scientific design principles and standards, and identify design issues that could lead to safety risks. Subsequent to this review, the application developer may choose to modify aspects of the application’s user interface to eliminate problems or departures from accepted best practice that are related to patient safety issues.

Step 3: EHR User Interface Validation Test: The Validation Test evaluates actual user performance on critical patient safety-related tasks identified in the previous steps, including a validation test conducted by qualified usability/human factors professionals prior to EHR implementation/deployment. Performance is examined by collecting user performance data that are relevant indicators of the presence of safety risks. These measures may include, but are not limited to, objective measures of successful task completion, number of errors and corrected errors, performance difficulties, and failures to complete the task successfully or in proper sequence. Performance is also evaluated by conducting post-test interviews focused on what users identify as risks based on confusion or misunderstanding when carrying out directed scenarios of use. The goal of the validation test is to make sure that critical interface design issues are not causing patient safety-related use error; in other words, that the application’s user interface supports error-free user interaction. Sample forms for test scenarios are provided in the appendices as examples only; the development and test teams will develop other scenarios and modify these examples as necessary for their medical context.

The balance of this document summarizes research findings on the relationship between usability and patient safety applicable to EHRs, and describes the overall EUP in detail (with examples of supporting documents provided in appendices). *It is our expectation that the potential for use errors can be identified and mitigated by using the EUP.*

2 Concept for an EHR Usability Protocol (EUP)

EHRs offer great promise for improving healthcare processes and outcomes, including increased patient safety. Emerging evidence suggests that the use of health information technology (HIT) may help address significant challenges related to healthcare delivery and patient outcomes.⁴ For example, three recent reports suggest that the use of HIT may improve healthcare outcomes⁵ and reduce patient mortality.⁶ In addition, the use of HIT is a key component of a national strategy to improve healthcare quality and patient safety.⁷ We anticipate that over the next few years, experience with meaningful use of EHRs in hospital and outpatient settings will grow rapidly. Given the estimate that one in three patients will potentially be harmed during hospitalization,⁸ the potential for using EHRs to improve patient safety may be significant.

On the other hand, a prior study found that patient mortality unexpectedly increased following the introduction of an EHR in a pediatric hospital.⁹ Additional research is needed to better understand how EHR usability can impact patient outcomes.¹⁰ As with any health information technology, EHR usability problems that can adversely impact patient safety can be assessed, understood, and controlled.

Usability Affecting Safety is the Key Focus of the EUP. Compromised EHR system usability can have a number of significant, negative implications in a clinical setting. Two key impacts that are of concern to the healthcare IT community are (1) use errors that can potentially cause patient harm; and (2) attenuating EHR adoption rates.

Experts have identified shortcomings in the usability¹¹ of current EHR systems as one of the barriers to adoption and meaningful use of these systems. The President's Council of Advisors on Science and Technology in December of 2010¹² framed the issue this way:

⁴ Buntin, M., Burke, M., Hoaglin, M., & Blumenthal, D. (2011). The benefits of health information technology: a review of the recent literature shows predominantly positive results. *Health Affairs*, 30 (3), 464-471.

⁵ See Gibbons, M.C., et al. (2009). Impact of consumer health informatics applications. *Evidence Report/Technology Assessment*, Oct (188), 1-546; and Cebul, R., Love, T., Jain, A.K., & Hebert, C.J. (2011). Electronic health records and quality of diabetes care. *New England Journal of Medicine*, 365, 825-833.

⁶ Longhurst, C., Parast, L., Sandborg, C., et al. (2010). Decrease in hospital-wide mortality rate after implementation of a commercially sold computerized physician order entry system. *Pediatrics*, 126, 14-21.

⁷ Institute of Medicine. (2001). *Crossing the quality chasm: a new health system for the twenty-first century*. Washington, DC: National Academy Press.

⁸ Classen, D.C., et al. (2011). Global trigger tool shows that adverse events in hospitals may be ten times greater than previously measured. *Health Affairs*, 3(4), 581-589.

⁹ Han, Y.Y., Carcillo, J., Venkataraman, S., et al. (2005). Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics*, 116(6), 1506–1512. Correction in *Pediatrics*, 117(2), 594.

¹⁰ Institute of Medicine. (2012). *Health IT and patient safety: building safer systems for better care*. Washington, D.C.: The National Academies Press.

¹¹ Defined in ISO 9241-11 as “effectiveness, efficiency, and satisfaction with which the intended users can achieve their tasks in the intended context of product use.”

¹² President's Council of Advisors on Science and Technology. (2010). *Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward*. Washington, DC.

“...the current structure of health IT systems makes it difficult to extract the full value of the data generated in the process of healthcare. Most electronic health records resemble digital renditions of paper records. This means that physicians can have trouble finding the information they need, and patients often wind up with poor access to their own health data and little ability to use it for their own purposes...*market innovation has not yet adequately addressed these challenges to the usability of electronic health records.*” (p.10, emphasis added)

Poor usability of EHR applications is widely believed to have a substantial negative effect on clinical efficiency and data quality. One step toward improving the overall usability of EHRs was the recent publication of NISTIRs 7741¹³ and 7742, which provide practical guidance to the vendor community on how to do user-centered design and diagnostic usability testing to improve the usability of systems under development.

The authors recognize that the factors preventing more widespread adoption of EHR systems are multidimensional and systemic, and they should be addressed and handled throughout the product development life cycle, as discussed in NIST 7741.

For the purposes of this document, the authors expressly focus on identifying and mitigating usability issues that, if left unaddressed, could result in errors of omission or commission that could potentially lead to patient harm. A white paper from the Healthcare Information and Management Systems Society (HIMSS) Usability Task Force underscored the importance of improving usability because of its “strong, often direct relationship” with error rates and user fatigue.¹⁴ Examples of safety-related usability issues that have been reported by healthcare workers include poorly designed EHR screens that slow down the user and might sometimes endanger patients, warning and error messages that are confusing and often conflicting, and alert fatigue (both visual and audio) from too many messages, leading users to ignore potentially critical messages. The proposed EUP is therefore concerned with helping vendors, hospitals, and/or other stakeholders to ensure that use errors with EHR systems are minimized, and providing technical guidance for summative usability evaluations prior to deployment or implementation of an EHR.

¹³ National Institute of Standards and Technology. (2010). *NIST Guide to the Processes Approach for Improving the Usability of Electronic Health Records* (NISTIR 7741). Gaithersburg, MD: http://www.nist.gov/manuscript-publication-search.cfm?pub_id=907313.

¹⁴ Health Information Management Systems Society. (2009). *Defining and Testing EMR Usability: Principles and Proposed Methods of EMR Usability Evaluation and Rating*. (http://www.himss.org/content/files/HIMSS_DefiningandTestingEMRUsability.pdf): Belden, J., Grayson, R., & Barnes, J.

2.1 General Approach

Starting assumption: Application designs have been optimized for general usability. In carrying out the proposed EUP, the priority is ensuring that necessary and sufficient usability validation and remediation has been conducted on the given application so that use error is minimized. Thus, this protocol focuses on identifying and minimizing critical patient safety aspects of usability issues associated with EHR interface design. The intent of the EUP is to validate, by way of systematic steps, that sufficient analysis and testing has taken place to mitigate the prevalence and severity of critical usability issues with the given EHR application user interface.

This protocol is not meant to provide a comprehensive review or incorporation of all of the factors that contribute to the usability of an EHR application. The authors endorse and encourage vendors to conduct formative and exploratory user research and usability testing early and often in the design process to isolate points of confusion and difficulty and to correct usability issues in the system related to workflow, navigation, screen layout, interaction model, visual design, etc. As more issues are discovered and corrected, the authors suggest that summative testing representing the anticipated environment of use is conducted to validate the application with real users prior to deployment. (For further technical guidance regarding the appropriate application of formative and summative testing methods, see NISTIR 7741.)

Protocol is founded on historical industry success. The proposed EUP builds on best practices, current procedures, and guidance from various government agencies for systematic application of human factors in the development process. We briefly summarize the history and effectiveness of a few of these programs in *Appendix A* in order to provide examples of human factors and usability evaluation and validation processes that resulted in positive impacts on safety and effective use of systems.

Intended audience. The intended audience for this document is any stakeholder interested in ensuring that use errors with EHR systems with the potential for patient harm are minimized by providing technical guidance for summative usability evaluations prior to deployment or implementation of an EHR. Examples of stakeholders include system developers, healthcare delivery organizations, consumers, government entities, and the Office of the National Coordinator's Regional Extension Centers.

2.2 EUP Foundation: Research Findings Defining EHR Usability Issues and Their Impact on Medical Error (EHR Patient Safety Model)

This section provides an in-depth look at research findings on critical EHR usability issues and the relationship between usability and potential adverse events. This section provides the technical foundation for evaluations of EHR user interface design. This section does not cover other factors that may affect patient safety such as clinical expertise, work environment factors, adherence to policies and procedures, etc.

Medical Device Usability and Safety Incidents. Usability issues have significant consequences when risk is introduced due to user confusion or inability to gain access to accurate information during clinical decision making. According to one source, more than one-third of medical device incident reports have been found to involve use error, and more than half of the recalls can be traced to user interface design problems.¹⁵ As a result, the FDA has placed increased emphasis on testing the user interfaces of devices in pre-market approval, as evidenced by recent publication of the agency's human factors guidance.¹⁶

Usability as a safety issue is not new to industries where human error can have severe consequences. Lessons learned from decades of experience using human factors and usability methods from industries such as nuclear power, military, and commercial aviation are relevant, as described in *Appendix A*. As healthcare becomes increasingly patient-driven and delivered in ambulatory settings (home, community outpatient centers, surgical centers, etc.), technological tools that facilitate healthcare delivery like EHRs will offer great potential and are expected to be widely used in these environments. Usability issues, and particularly those associated with patient safety, will be no less important in these nonhospital settings where the environments and the variability of patient health issues will present additional challenges to EHR interface design.¹⁷

In the remainder of this section, we will:

1. Discuss a model for understanding EHR patient safety risks;
2. Define usability and associated measures; and
3. Define categories of medical errors.

2.2.1 A Model for Understanding EHR Patient Safety Risks

Emanuel and colleagues¹⁸ have defined patient safety as:

¹⁵ Comments from FDA spokesperson at Association for the Advancement of Medical Instrumentation (AAMI) conference, June 25, 2011, Washington DC.

¹⁶ Food and Drug Administration. (2011). *Draft guidance: applying human factors and usability engineering to optimize medical device design*.

<http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM259760.pdf>

¹⁷ National Research Council. (2011). *Health care comes home: the human factors*. Washington, DC: The National Academies Press.

¹⁸ Emanuel, L., Berwick, D., Conway, J., Combes, J., Hatlie, M., Leape, L., Reason, J., Schyve, P., Vincent, C., & Walton. (2008). *Advances in patient safety: new directions and alternative approaches (vol. 1: assessment)*. Rockville, MD: Agency for Healthcare Research and Quality.

“A discipline in the healthcare sector that applies safety science methods toward the goal of achieving a trustworthy system of healthcare delivery. Patient safety is also an attribute of healthcare systems; it minimizes the incidence and impact of, and maximizes recovery from, adverse events.”

Emanuel and colleagues defined mechanisms for achieving patient safety as:

High-reliability design. A central component of high-reliability design is that artifacts are designed to be resilient (fail-safe) to error traps, which are situations in which errors are highly likely.

Safety sciences. A central method for safety sciences is performing system adjustments based on an analysis of contributors to adverse events of artifacts in use (that have already been designed and tested and can be observed in the field). An integrated taxonomy of contributors to medical error events created by Mokkarala, Zhang, and colleagues¹⁹ after reviewing a number of similar taxonomies is provided in Figure 2.

Methods for causing change. Improving patient safety in an organization typically requires reducing gaps between acknowledged guidelines, standards, or protocols and practice through multiple strategies, including standardization, monitoring relevant measures, and collaboration across organizations.

¹⁹ Mokkarala, P., Brixey, J., Johnson, T., Patel, V., Zhang, J., & Turley, J. (2008). *Development of comprehensive medical error ontology: advances in patient safety, new directions and alternative approaches*. (AHRQ Publication Nos. 080034, 1-4). Rockville, MD: Agency for Healthcare Research and Quality.

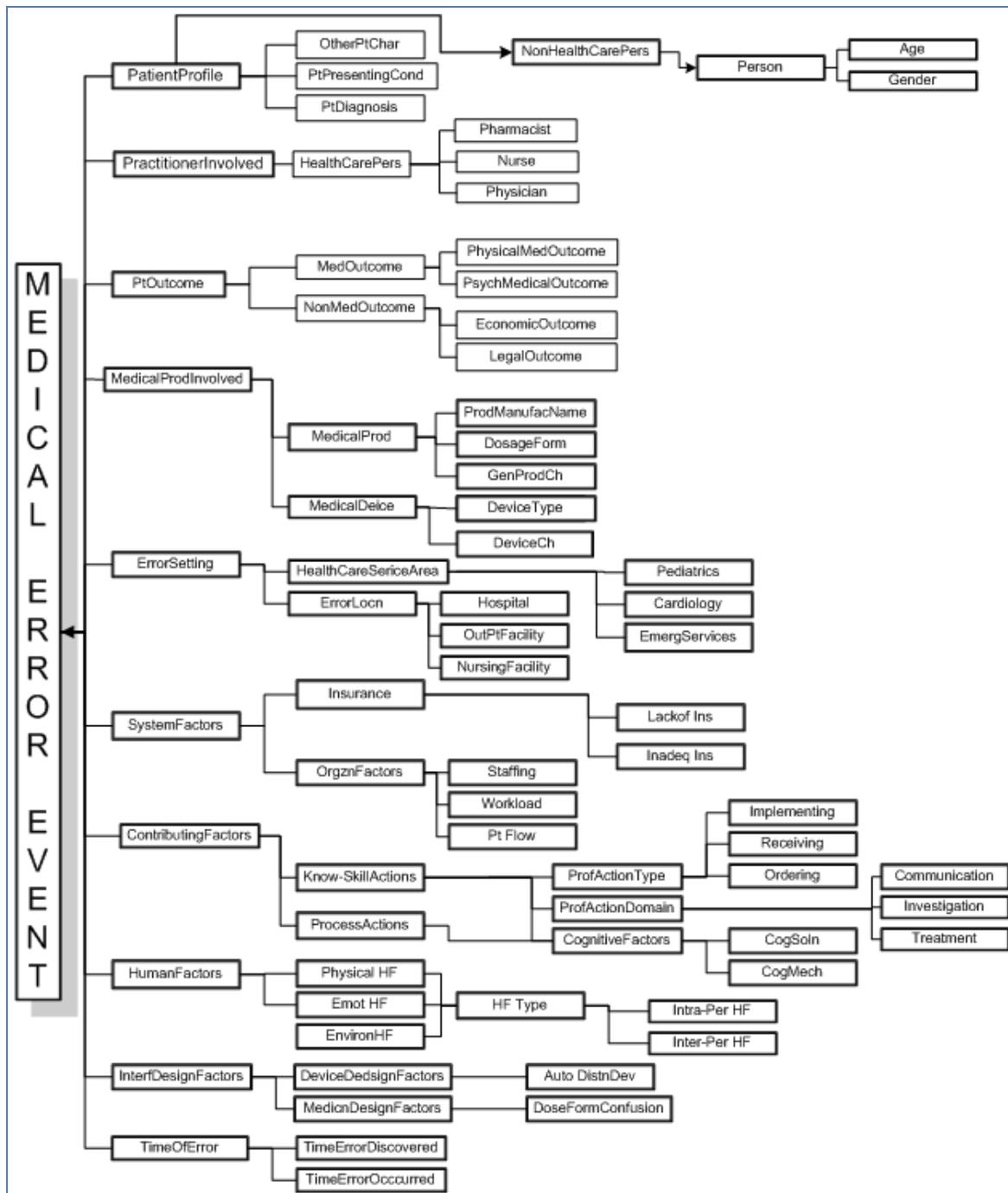


Figure 2. Taxonomy of contributors to medical error events (Mokkarala et al., [2008]; used with permission)

Figure 3 illustrates a summary of research findings on critical use errors, and shows how the potential for patient safety is identifiable based on risk factors and evaluative indicators.

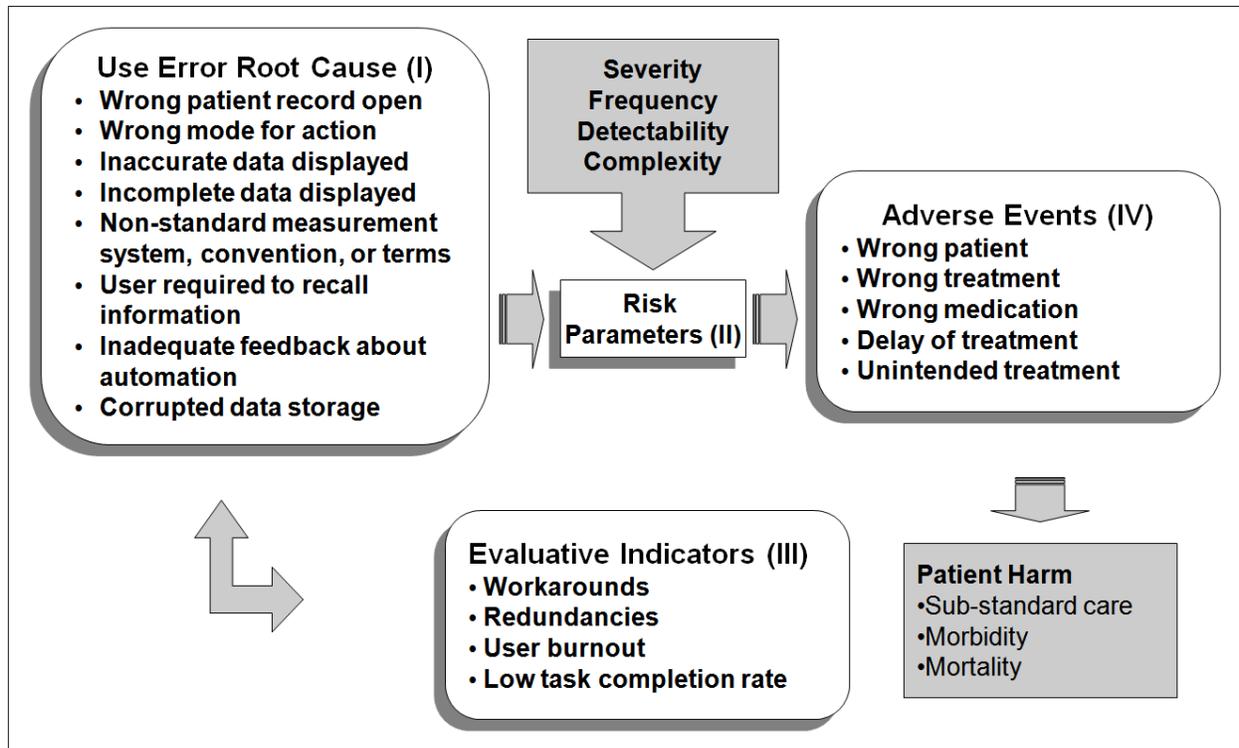


Figure 3. A model for analysis and understanding of use-related risks of EHR systems.

There are four main components in Figure 3. These are:

- I. **Use Error Root Causes**—Aspects of the user interface design that induce use errors when interacting with the system.
- II. **Risk Parameters**—These are attributes regarding particular use errors, i.e., their severity, frequency, ability to be detected, and complexity.
- III. **Evaluative Indicators**—Indications that users are having problems with the system. These are identified through direct observations of the system in use in situ, or through interviews with users.
- IV. **Adverse Events**—A description of the outcome of the use error, and standard classification of patient harm.

Use Error Root Causes (I) can be defined as attributes of the interface that produce an act or omission of an act that has a different result than that intended by the manufacturer or expected by the operator. Preliminary use error categories are listed below. The **hypothetical** examples only have one use error category for illustrative purposes, whereas it is likely that actual use errors could involve multiple root causes.

- **Patient identification error:** Actions are performed for one patient or documented in one patient’s record that were intended for another patient. For example, a surgeon removed the wrong limb because he or she was unable to access the display with the patient’s identifying information. In this example, it was standard to initiate surgical procedures before the entire

surgical team had entered the room, a practice that was intended to minimize costs and maximize efficiency.

- **Mode error:** Actions are performed in one mode that were intended for another mode. An example is *direct dose vs. weight dose*: a 100 kilogram patient received a 100-fold overdose of a vasoactive drug because weight dosing (mcg/kg/min) was selected instead of direct dosing (mcg/min). The error occurred due to lack of feedback about an unusual mode choice, e.g., there was no warning about an unusually high dose, and also due to parallax issues when looking down on the display, making it appear that the incorrect mode was on the same horizontal line as the appropriate button. The latter mode error was related to *test mode vs. production mode*: actions intended to be done in test mode to debug new software functionality for medication ordering were inadvertently done in production mode, partly because there were no differences in the displays between the test account and production account after the login procedure. The overall result was that a pediatric patient was nearly administered a medication with a dangerously high dose.
- **Data accuracy error:** Displayed data are not accurate. For example, a physician ordered the wrong dose of a medication because the amount of the medication dose was truncated in the pick list menu display.
- **Data availability error:** Decisions are based on incomplete information because related information is not updated within a reasonable time or requires additional navigation, access to another provider's note, or taking actions to update the status. For example, a patient received four times the intended dose of a medication because the provider did not see the comments field, which was not visible without being opened, and which explained that there were progressive dose reductions (taper dosing) over several days to wean the patient off the medication.
- **Interpretation error:** Differences in measurement systems, conventions, and terms contribute to erroneous assumptions about the meaning of information. For example, a patient received a larger dose of a medication than was intended because most displays used the English system but the pediatric dose calculation feature used the metric system.
- **Recall error:** Decisions are based on incorrect assumptions because appropriate actions require users to remember information rather than recognize it. For example, the wrong dose of a medication is ordered because, during the ordering process for an outpatient medication, when a one-time schedule is initially selected, the user must enter the appropriate quantity manually, whereas for recurring orders the user can select the dose from a list.
- **Feedback error:** Decisions are based on insufficient information because a lack of system feedback about automated actions makes it difficult to identify when the actions are not appropriate for the context. For example, a patient received eight times the dose of a medication for several weeks when a physician did not realize that a twice-a-day order for *1/4 tablet* was automatically changed to *1 tablet* when batch converting all 13 inpatient medications to outpatient medications. The conversion was intended to be a patient safety measure so that patients wouldn't need to understand how to administer partial tablets at home, but no feedback was provided to the user that the dose had been changed.

- **Data integrity error:** Decisions are based on stored data that are corrupted or deleted. For example, a patient received more than one pneumococcal vaccine because documentation that the vaccine had been given previously was automatically deleted by the EHR. Specifically, the physician selected the clinical reminder dialog box option “Order Pneumovax vaccine to be administered to patient,” and clicked on “Next” to process the next clinical reminder for a diabetic foot exam. The text documenting the vaccine order disappeared from the progress note. If the physician had pressed “Finish” instead of “Next,” the text documenting the vaccine order would have been correctly saved in the chart.

Risk Parameters (II) can be defined as controllable or uncontrollable factors that affect variation in the magnitude of the potential risk due to a use error. Risk parameters²⁰ are:

- **Severity:** Magnitude of potential harm.
- **Frequency:** Probability of harm occurring.^{21,22}
- **Detectability:** Ease of recognizing use error that could lead to potential safety issues.
- **Complexity:** Presence of factors that increase patient complexity for special patient populations, such as pediatric patients, patients with co-morbidities for which the risk of harm is higher, or patients with compromised immune systems. Complexity is known to be associated with increased opportunities for error, and thus increases the risk of patient harm.²³

Evaluative Indicators (III) can be defined as recurring themes in reports of system use that can serve as early indicators about systems issues in general, some of which might stem from usability problems. By gathering data through interviews, focus groups, ethnographic research, formative usability tests, and observations of the system in use, gaps in optimal user interaction design can be identified. In addition, evaluative indicators can be used to develop use cases and scenarios for usability evaluations that are more likely to detect system flaws that create use error hazards or traps proactively.

Preliminary evaluative indicator categories are:

- **Workarounds:** User-identified differences between the system’s design and their locally adopted workflow.

²⁰ Note that these factors collapse six theoretical dimensions of risk (Identity; Permanence; Timing; Probability; Victims; and Severity) identified in: Perrow, C. (1984). *Normal accidents: living with high-risk technologies*. New York, NY: Basic Books. Although theoretically clinicians or others could be harmed by interface design decisions, such as when nurses are erroneously blamed for stealing narcotic medications when the person logging in to return an unused narcotic is not the same as the nurse in a prior shift who removed the narcotic, this framework is restricted to situations where the harm is to a patient.

²¹ Note that frequent events are often associated with strategies to mitigate risk after systems have been implemented for a sufficient time to have stable work processes. Thus, infrequent events may actually be associated with a higher risk of patient harm for stable systems. Therefore, the probability is of patient harm, not of an adverse event.

²² There is a difference between the frequency of the triggering fault, or, given the triggering fault, the frequency of an adverse event following it. Following the ISO draft standard currently adopted by the National Health Service in the United Kingdom, we assume the latter definition.

²³ Institute of Medicine. (2006). *Preventing medication errors: quality chasm series*. Washington, DC: National Academies Press. PDF available at: <http://www.nap.edu/catalog/11623.html>.

- **Redundancies:** Actions that users must repeat or re-document because system components are poorly integrated.
- **Burnout:** Noticeable increase in clinician perception of long-term exhaustion and diminished engagement in providing care, possibly contributing to loss of staff or early retirement decisions.
- **Low task completion rate:** Frequent initiation—but not completion—of tasks.
- **Potential patient safety risk:** Accidental or preventable injuries attributable wholly or partially to the design or implementation of an EHR.

Adverse Events (IV) can be defined as sentinel events attributable wholly or partially to an EHR's user interface design defects. These defects create error traps that make it easy for use errors of commission, where actions are erroneously taken, or errors of omission, where actions are erroneously omitted.²⁴ These event outcomes are similar to the patient safety checklist items for EHRs developed by HIMSS.²⁵ The proposed categories of outcomes produced by use-related errors are:

- **Wrong patient action of commission:** Actions with potentially harmful consequences are performed for one patient that were intended for another patient primarily due to inadequate selection mechanisms or displays of patient identifiers.
- **Wrong patient action of omission:** A patient is not informed of the need for treatment primarily due to inadequate selection mechanisms or displays of patient identifiers.
- **Wrong treatment action of commission:** Treatments that were not intended for a patient are provided primarily due to inadequate selection mechanisms or displays of treatment options.
- **Wrong treatment action of omission:** Treatments that were intended for a patient are not provided primarily because of inadequate selection mechanisms or displays of patient identifiers.
- **Wrong medication:** A patient receives the wrong medication type, dose, or route primarily due to inadequate selection mechanisms or displays of medication data.
- **Delay of treatment:** A patient receives a significant delay in the provision of care activities due to design decisions made to satisfy billing, security, or quality improvement objectives.
- **Unintended or improper treatment:** A patient receives care that was not intended due to inadequate selection mechanisms, inadequate displays of care provision options, or indistinguishable modes that are alternatively used for regular operation versus for software testing, training, and/or software demonstration. This category does not include the receipt of treatments intended for other patients.

There are three levels of potential patient harm attached to these outcomes in the list above:

²⁴ Wickens, C., Gordon, S., & Liu, Y. (1998). *An introduction to human factors engineering*. New York, NY: Addison-Wesley Educational Publishers Inc.

²⁵ Health Information Management Systems Society. (2011). *Defining and testing EMR usability: principles and proposed methods of EMR usability evaluation and rating*. Accessed September 10, 2011 at http://www.himss.org/content/files/HIMSS_DefiningandTestingEMRUsability.pdf.

- **Substandard care:** Lack of preventive care, wrong care, or unnecessary care, including tests or decreased comfort attributable to design choices with an EHR. These events are likely to reduce patient satisfaction or increase costs, and therefore are important aspects of patient harm, but are defined as the lowest level of patient harm.
- **Morbidity:** The rate of incidence of a disease, condition, or state. In a clinical setting, morbidity may be defined as the frequency of the appearance of complications following a surgical procedure or other treatment.
- **Mortality:** The rate of incidence of death. In a clinical setting, this would mean a fatal outcome.

2.2.2 Definition of Usability and Associated Measures

Given our working definition of Usability, “The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use.”²⁶ These terms are further defined as:²⁷

- **Effectiveness:** The accuracy and completeness with which users achieve specified goals.
- **Efficiency:** The resources expended in relation to the accuracy and completeness with which users achieve goals.
- **Satisfaction:** Freedom from discomfort and positive attitude toward use of the product.
- **Context of use:** Characteristics of the users, their tasks, and their organizational and physical environments.

In the International Organization for Standardization’s (ISO) most recent standards (i.e., ISO 25010), usability is included as one of eight attributes of software quality.²⁸ The sub-characteristics of usability are:

- **Appropriateness recognizability:** The degree to which the software product enables users to recognize whether the software is appropriate for their needs.
- **Learnability:** The degree to which the software product enables users to learn its application.
- **Operability:** The degree to which users find the product easy to operate and control.
- **User error protection:** The degree to which the system protects users against making errors.
- **User interface aesthetics:** The degree to which the user interface enables pleasing and satisfying interaction for the user.
- **Accessibility:** Usability and safety for users with specified disabilities.

The other quality attributes are: functional suitability, reliability, performance efficiency, security, compatibility, maintainability, and portability.

²⁶ ISO/IEC. (1998). *Ergonomic requirements for office work with visual display terminals (VDT)s - Part 14 Menu dialogue* (ISO/IEC 9241-14).

²⁷ *Ibid.*

²⁸ ISO/IEC. (2011). *Systems and software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Software product quality and system quality in use models*. (CD 25010.3), First Edition March 1, 2011.

Usability is correlated with how useful a system is perceived to be.^{29,30} Based on the rationale that usefulness cannot be meaningfully separated from usability, Zhang and Walji³¹ describe how they have expanded the ISO definition for usability, resulting in three dimensions for usability:

1. **Useful:** A system supports the work domain where the users accomplish the goals for their work, independent of how the system is implemented.
2. **Usable:** The system is easy to learn, easy to use, and error-tolerant.
3. **Satisfying:** Users of the system have a good subjective impression of how useful, usable, and likable the system is.³²

Objective, Measurable Dimensions of Usability. Systems that have better usability on these dimensions can be expected to have fewer or no negative patient consequences:

Efficiency, measured objectively as:

- Time to accomplish a task (average, standard deviation)
- Number (average, standard deviation) of clicks and keyboard/input device interactions to accomplish a task

Consistency, measured via a subjective rating scale for the sub-dimensions of:

- Optimal placement of design elements (e.g., “cancel” buttons/ controls),
- Consistent labeling of design elements (e.g., dialog box has the same label as the button that was pressed to open it),
- Consistent use of keyboard shortcuts (e.g., Ctrl-C for copy),
- Appropriate color coding of information (e.g., red reserved for errors, yellow for alerts and warnings),
- Appropriate font size and type, and
- Appropriate and consistent use of measurement units (e.g., kilograms).

Unintended Action or Inaction can Lead to Use Errors. In healthcare, use errors have been identified as a particularly important aspect of usability due to their potential consequences for patients, as well as the associated liability concerns of healthcare organizations, providers, and manufacturers. The definition of a **use error** is “an act or omission of an act that has a different result to that intended by the manufacturer or expected by the operator.”³³ Figure 4, from the *2007 American National Standard*

²⁹ Keil, M., Beranek, P. M., & Konsynski, B. R. (1995). Usefulness and ease of use: field study evidence regarding task considerations. *Decision Support Systems*, 13, 75-91.

³⁰ Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13 (3), 319-340.

³¹ Zhang, J., Walji, M. (2011). TURF: Toward a unified framework of EHR usability. *Journal of Biomedical Informatics*, 44 (6), 1056-67.

³² *Ibid.*

³³ ANSI/AAMI. (2009). *Human factors engineering — Design of medical devices (ANSI/AAMI HE75)*.

for Medical Devices on Application of usability engineering to medical devices³⁴ depicts how user actions and inactions are related to use errors. User actions (or inactions) that are unintended can lead to use errors due to attention failure, memory failure, rule-based error, knowledge-based error, or nescient error, defined as a lack of awareness of the adverse consequences of a skill-based action.

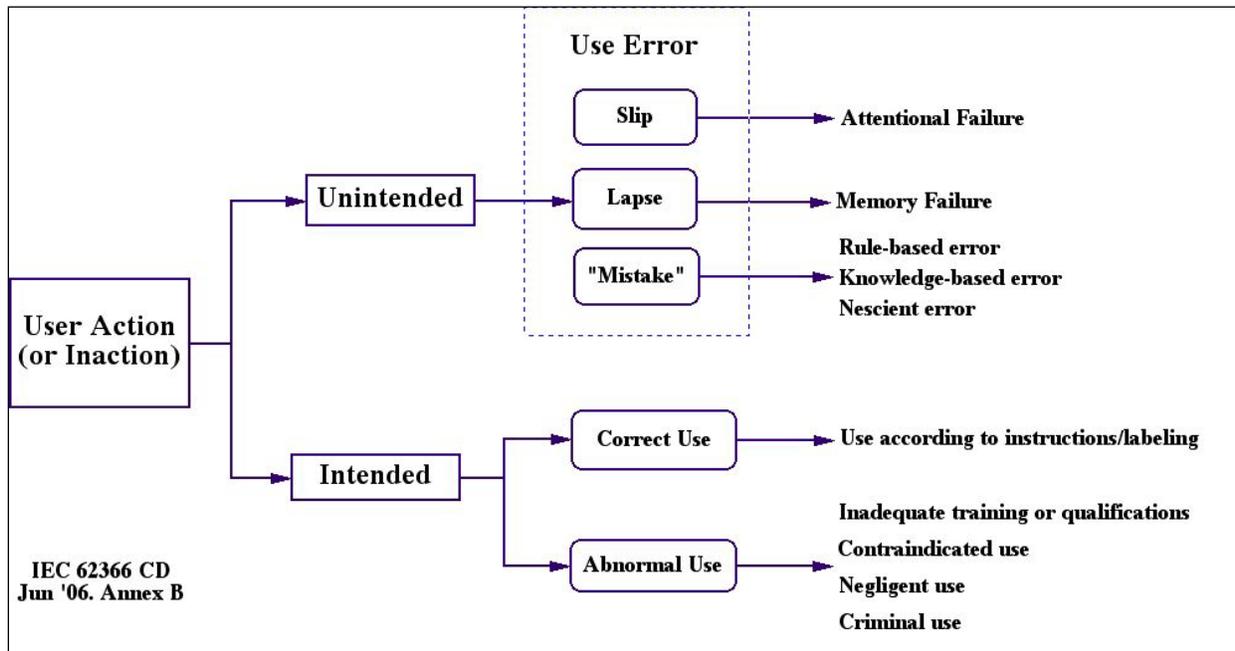


Figure 4. Relationship of user actions and use errors (from ANSI/AAMI/IEC 62366)

Use error can also stem from inappropriate use. Although intended abnormal uses are not included in the definition of use errors in Figure 4, the recently released FDA human factors draft guidance titled “Draft Guidance for Industry and Food and Drug Administration Staff - Applying Human Factors and Usability Engineering to Optimize Medical Device Design” does recommend proactively protecting against anticipated inappropriate use. Specifically, the FDA defines use-related hazards as “occurring for one or more of the following reasons”:³⁵

- Device use requires physical, perceptual, or cognitive abilities that exceed the abilities of the user.
- The use environment affects the user’s physical, perceptual, or cognitive capabilities when using the device to an extent that negatively affects the user’s interactions with the device.
- Device use is inconsistent with user’s expectations or intuition about device operation.
- Devices are used in ways that were not anticipated by the device manufacturer.

³⁴ ANSI/AAMI/IEC. (2007). *Medical devices - Application of usability engineering to medical devices* (ANSI/AAMI/IEC 62366).

³⁵ Food and Drug Administration. (2011). *Draft guidance: Applying human factors and usability engineering to medical device design*. <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm259748.htm> .

- Devices are used in ways that were anticipated but inappropriate and for which adequate controls were not applied.

Thus, the FDA looks at usability problems as a process of understanding how human capabilities and limitations and the environment of use may cause errors during device use. This requires additional investigative methodology in testing that goes beyond conventional measures such as time and number of errors.

2.2.3 Categorization of Medical Errors

This last section focuses on critical use errors with EHR user interfaces that may lead to medical error. It also focuses on the usability of user interface characteristics as defined in ISO 25010 as “the degree to which the system protects users against making errors.”

The effect of medical errors can be categorized by potential for patient harm. The National Coordinating Council for Medication Error and Reporting and Prevention (NCC MERP) index was originally developed for this purpose in 1996, and subsequently updated in 2001, to provide a taxonomy for medication error classification (see Figure 5).³⁶ Although developed for medication errors, the concept of collapsing the concepts of severity and level of harm in patient safety outcome definitions is applicable to the EHR case: 1) no error; 2) error, no harm; 3) error, harm; and 4) error, death.

³⁶ Forrey, R., Pedersen, C., & Schneider, P. (2007). Interrater agreement with a standard scheme for classifying medication errors. *American Journal of Health-System Pharmacy*, 64(2), 175-181.

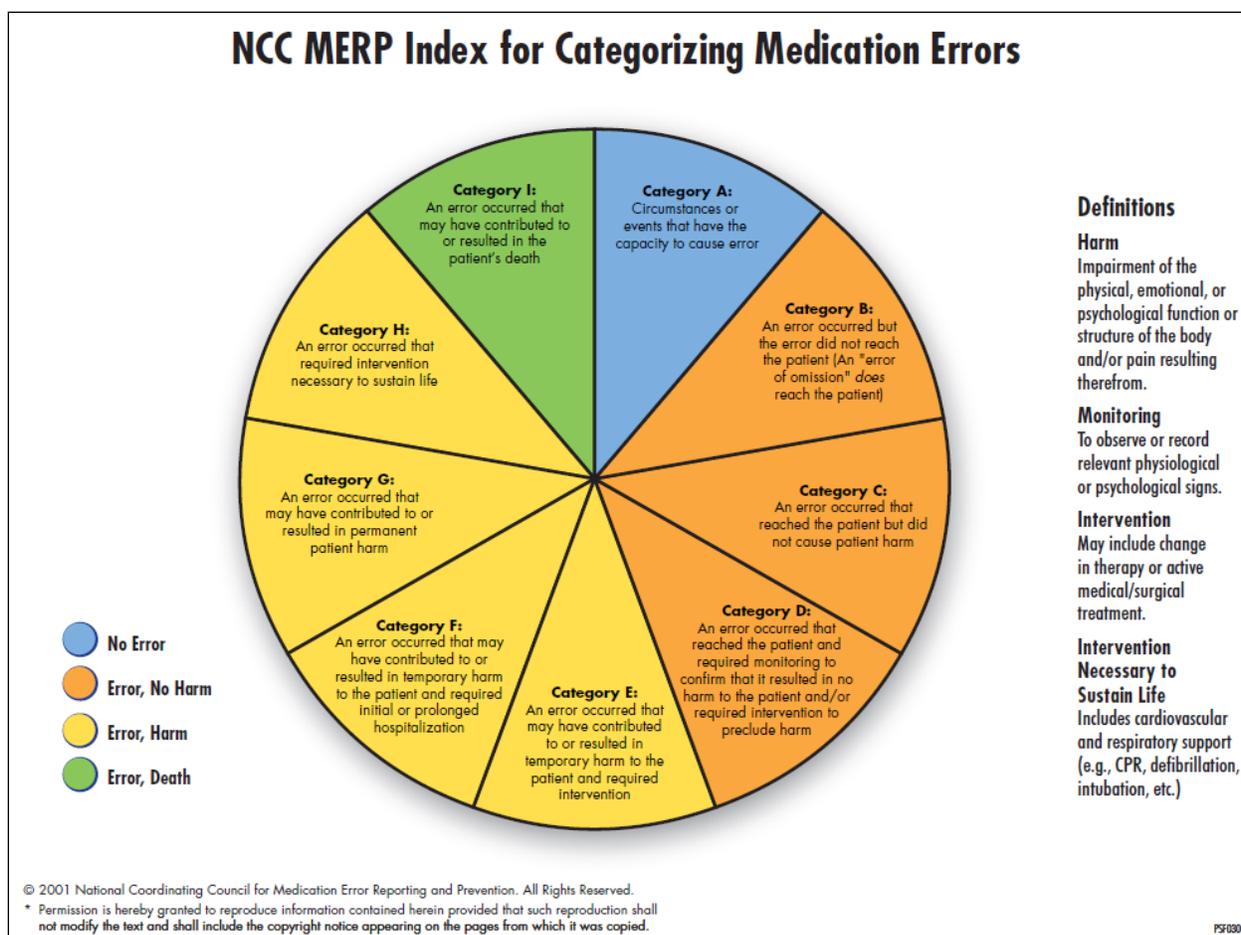


Figure 5. Levels of Patient Harm (National Coordinating Council, [2001]; used with permission)

VHA Categorization Scheme for Patient Safety Issues. The Veterans Health Administration (VHA) has been a recognized leader since 2000 in reducing the potential for patient harm due to health information technology design and implementation. The VHA’s Office of Health Information (OHI) created an innovative system to track patient safety issues, and has used this system proactively to address hundreds of voluntary reports. Data are reported as National Online Information Sharing (NOIS) events and are entered through the *Patient Safety Internal Notification Form*, which includes fields for a description of the problem, the specific incidents that occurred, the harm to patient(s), the site where the problem was identified, and contact information to obtain more information. Each patient safety issue is then distributed to a patient safety workgroup through a secure email distribution list and discussed ad hoc through email exchanges and conference calls. The Patient Safety Workgroup then generates a plan to address the issue. Each issue is triaged within a week of reporting based upon severity, frequency, and detectability. The issues are rated on a scale of 1 to 4 for each of these dimensions, where severity ranges from minor to catastrophic on the basis of level of potential patient harm, frequency ranges from not likely to happen in the lifetime of the software use to likely to happen several times a week, and detectability ranges from obvious to not detectable within the use of a single software package. Since the scope of the EHR Usability Protocol is limited to pre-implementation, no

specific guidance is provided regarding how the VHA approach might be tailored for use in a complementary fashion following system implementation.

There exists a lack of standardization in procedures and terminology used in the health IT industry to classify, report, and track usability-related incidents.³⁷ An example of how resolutions of reported Patient Safety Issues are tracked in the VHA is provided in Table 1 below. Color codes are used with the following coding: clear is a closed issue, green is an issue that is awaiting a fix, yellow is an issue where analysis is ongoing, and red is a newly reported issue that is awaiting analysis by a designated ad hoc committee of clinical experts, usability experts, and vendor representatives that has been formed to address the issue.

No.	Date Found	Date Fixed	Date Fix Released	Reported?	Contact	Resolution	Related Issues	Priority
2011-PSI-03	5/31/11	6/2/11	6/6/11	Yes	Smith, John	Medication doses truncated in pick list menu makes it easy to pick the wrong dose	2009-PSI-08	High

Table 1. Veteran’s Health Administration’s system for tracking resolution of reported issues

Synopsis

In Section 2.2, we synthesized research findings to define categories of critical usability issues that could potentially induce professional healthcare personnel to make use errors with EHRs that could directly contribute to a patient incurring harm. Though they are not intended as a prescriptive framework, it is our hope that these research-based categories will help evaluation teams to structure their approach to the Application Analysis and Expert Review outlined in Steps 1 and 2 of the EUP. A key overall EHR design objective is proactive detection and elimination of potential use errors well before summative testing and system deployment. These findings provide the technical, risk-based foundation for determining what to include in the summative test, or Step 3 of the EUP.

³⁷ McDonnell, C, Werner, K, & Wendel, L. (2010). *Electronic Health Record Usability: Vendor Practices and Perspectives*. (AHRQ Publication No. 09(10)-0091-3-EF). Rockville, MD: Agency for Healthcare Research and Quality.

2.3 Overview of the Protocol Steps

Different usability evaluation methods are appropriate for different stages of the software development cycle, the EUP includes both expert evaluation by human factors experts and summative usability testing (validation) with real users. The EUP consists of three major steps:

- (I) **EHR Application Analysis:** Identifies the characteristics of the system’s anticipated users, use environments, scenarios of use, and use-related usability risks that may induce medical errors. The analysis is ideally conducted by a multidimensional team including the application development team, clinical experts, human factors practitioners, and others providing direct application expertise. The analysis provides a description of the design of the application’s user interface and how the design has been optimized via iterative formative and exploratory usability assessments during development.
- (II) **EHR Interface Expert Review:** An evaluation of the critical components of the user interface in the context of executing various use case scenarios and comparing the design against user interface design best practices and principles.
- (III) **EHR Interface Validation Testing:** A summative usability test with representative users in a simulated use environment that identifies remaining application interface features that are related to critical use errors.

Step I: EHR Application Analysis. The first step is for the application development team to conduct and provide an overall assessment of the EHR application describing the EHR anticipated user groups and their characteristics, the environments of use for the system, the various scenarios of system use, the envisioned user interface, and the critical use-related risks for each use scenario. We have included a more detailed outline of the Application Analysis in Section 3, but a brief summary of this analysis includes the following:

- **Who are the users?** Descriptions of all of the different EHR users including their job classifications (physician, registered nurse, admin, trainee, etc.), knowledge, training on the system, and user characteristics.
- **What is the user’s work area like?** Ambient conditions of use including lighting, noise, vibration, distraction, workstation layout, etc.
- **What do the users do?** Description of the major actions that users perform in the form of a narrative of typical sequences of interaction, also known as a “use scenario.” (This does not include the “keypressing” level of detail, but simply describes the step-by-step process of how a user interacts with the system to enter data, view patient data, or retrieve information.) A task analysis approach captures the hierarchy of steps during user interaction and is recommended for documenting use scenarios. Given the variability of workflows in healthcare, it is important to map out the complete flow of information and activities, including who is doing what activities, collecting what information, documenting what health data, etc.
- **What does the user interface look like and how does it operate?** Description of the major elements of the user interface including layout, displays, controls, means of interaction (cursor

control, mouse, touch screen, etc.) This could be conveyed and provided through storyboards, flash simulations, or working models of the user interface, etc.

- **What mistakes might users make?** For each of the scenarios, identification of any potential errors a user might make during interaction with the EHR system and description of the potential healthcare delivery consequence resulting from that error. This analysis includes errors of omission as well as commission. For example, in a patient room, a caregiver not seeing or hearing vital information conveyed by an EHR that requires immediate intervention with a patient, would be an error of omission, while selecting the wrong drug from a formulary due to ambiguous or confusing abbreviations would be an error of commission. The analysis also includes indirect results like problematic workarounds and contextual issues that contribute to errors, e.g., interruptions.
- **How has the design been iterated to mitigate mistakes and optimize overall usability?** Any formative usability assessments conducted in the process of developing the application user interface can be described in terms of findings that led to iterative design changes.

Step II: EHR User Interface Expert Review. The second step is an expert review of the application's user interface. Usability/Human Factors (HF) and clinical subject matter experts conduct the expert review to determine the application's human factors deficiencies and its adherence to best design principles and usability standards. Potential design modifications addressing these deficiencies may result from this review.

NIST research conducted by a combination of human factors and clinical experts has resulted in user interface design review protocols for EHR user interfaces.³⁸ These reviews are focused on identifying potential user interface design-induced use errors including features that do not represent known best practice in interface design. For instance, an EHR that displays a dialog box whose width is not sufficient to show the full name of a drug from the formulary is a potential cause for alarm. If the full name of the drug cannot be shown, the doctor may select the wrong drug. Other examples exist, such as inconsistent representation of the patient's name, incomplete allergy information, etc.

Other input to this process will result from the development of consistent and reliable techniques for user interface expert reviews. Current research in this area is being funded by the Office of the National Coordinator for Health Information Technology. This program, known as Strategic Health IT Advanced Research Projects – Cognitive (SHARPC), is developing a Rapid Usability Assessment Protocol based on known user interface design best practices, e.g., consistency.³⁹ Specifically, this method involves the use

³⁸ National Institute of Standards and Technology. (2010). Customized common industry format template for electronic health record usability testing (NISTIR 7742). Gaithersburg, MD: http://www.nist.gov/manuscript-publication-search.cfm?pub_id=907312.

³⁹ Zhang, J., Walji, M. (2011). TURF: Toward a unified framework of EHR usability. *Journal of Biomedical Informatics*, 44 (6), 1056-67.

of two expert evaluators applying design review heuristics (i.e., “design rules of thumb”) concerning user interface design to determine if there are departures from good design principles and the level of severity for the departures.⁴⁰ The violations range from “advisory” (moderate concern) to “catastrophic” (critical to safety of interaction). In a post rating analysis session, the independent evaluations of the user interfaces against the design heuristics are aggregated and differences between reviewers are reconciled.

The findings of the reviewers are summarized and documented. These findings are prioritized with respect to risk of medical error to identify the most critical user interface problems that are then collaboratively addressed with the development team. Design modifications may result from this review. The resulting modified user interface will represent the system version to be tested in the next step, User Testing.

Step III: EHR User Interface Validation Test. The final step is a test of the application’s user interface with representative users. In this test, often referred to as a summative usability test, users perform representative tasks with the EHR and their performance is observed, recorded, and categorized as successful, successful with issues or problems, or unsuccessful based on certain criteria that define success.

NIST recommends conducting usability validation testing as outlined below and reporting those studies using the “Customized Common Industry Format Template for Electronic Health Record Usability Testing” by NISTIR 7742.

The critical analysis in a validation study is in the examination, through post-testing interview techniques, of any observed errors or task execution difficulties, confusions, hesitations, corrected mistakes, etc., with the intention of identifying “root cause” of any errors observed.

The series of methods, reviews, and tests in this EUP reflect a due diligence where the knowledge and outcomes produced by the entire protocol is greater than the sum of its parts. Different methods will reveal different bodies of evidence; in particular, the expert review is likely to uncover more instances of heuristic rule breaking whereas the user test is more likely to reveal real user behavior with the system and associated usability issues. Multiple reviewers and different methods are used together in order to increase the validity of the methods. Overall, the EUP provides a process for incorporating user-centered design into the development and deployment of EHR systems with the purpose of mitigating unnecessary patient risk due to interfaces with poor usability.

⁴⁰ Nielsen, J. (1994). *Usability Inspection Methods*. New York, NY: John Wiley and Sons.

2.4 What the EHR Usability Protocol Is Not

The proposed EUP is primarily about ensuring safe use of EHRs. It outlines methods focused on identifying the potential for use errors with the potential for patient harm and validating that potential errors identified in Step 1 and Step 2 of the protocol are not made in Step 3. The summative usability testing evaluation is meant to be independent from factors that engender creativity, innovation, or competitive features of the system. The EUP does not question an “innovative feature” being introduced by a designer, but could proactively identify and reduce risks to patient safety when introducing the feature.

The EUP is not intended to result in a relative or comparable score of usability. Thus, each case involves the development of metrics that are unique in their level of detail to the specific context of use, user characteristics, potential risks, and tasks of the system against which they are tested. The EUP is a roadmap for stakeholders to validate that the core of an EHR system does not contribute to use error with specified categories of intended users, doing representative tasks, in a defined context of use. It is intended to validate that systems designers have used best practices to proactively mitigate concerns from predicted use errors prior to patient harm throughout the design cycle and to communicate summative testing results in a standard way. Following the guidelines in this protocol does not guarantee that all risks will be completely avoided. Following the protocol is expected to help reduce known, predictable, and unnecessary risks to patients.

3 Step I: EHR Application Analysis

The EHR Application Analysis is typically the responsibility of the application developer. The EHR Application Analysis is the foundation for the conduct of the entire EUP processes embodied in the final two steps of the EUP – the EHR Interface Expert Review and the subsequent EHR Interface Validation Test. This section provides a description of the suggested outline for an EHR Application Analysis.

3.1 Describe Application Scenarios of Use, User Tasks, and Application Interface Interaction

Provide a description of the various scenarios representing how the application is used to deliver healthcare or provide clinical information. Scenarios of use are narratives of the basic sequence of interaction with the application by the user. These scenarios can be further decomposed into a list of tasks or steps that the user performs to complete each scenario with reference to aspects of the user interface that are used to complete these steps. The use of screen shots provides an invaluable and recommended tool for the subsequent interface expert review process.

3.2 Define and Describe Application User Groups

Provide a description of each distinct user group that interacts with the application in carrying out the scenarios of use in Section 3.1. These descriptions can include any population characteristics that would affect user interaction performance such as age demographics, visual, auditory, or physical impairments, repetitive stress injury factors, education, training, experience, computer literacy, and other factors potentially affecting interaction with the application.

3.3 Describe Application Environment of Use

Provide a description of the application's environment(s) of use including aspects that may affect user performance such as lighting, noise, distraction, interruptions, mobility factors, ergonomics, e.g., sitting vs. standing at a keyboard, etc.

3.4 Identify Critical Safety-Related Tasks

Within each scenario of use, identify the critical steps or tasks that may have patient safety implications. For each of these critical steps, the analysis will describe certain errors (both errors of commission and errors of omission) that could propagate unsafe situations related to the factors identified in the patient safety categorization found in Section 2 of this document. In particular, the potential adverse outcomes and/or root causes are identified in this analysis as listed in Section 2, Figure 3, Root Causes (I) and Adverse Events (IV).

3.5 Formative and/or Preliminary Usability Assessment Overview

The Application Analysis concludes with a description (or tabular summary) of the EHR application development team's conduct of formative usability evaluations, design reviews, cognitive walkthroughs, focus groups, etc., that were instrumental in shaping the design of the user interface, and particularly in eliminating usability problems related to safety or general usability factors. Each study is briefly described identifying the users tested, interface design changes that resulted, and what problems were addressed.

4 Step II: EHR Interface Expert Review

Here we discuss briefly, the interface expert review process and design principles used in reviews.

4.1 Overview of Interface Expert Review Process

The EHR Interface Expert Review process consists of four major phases: (1) Review of the EHR Application Analysis, (2) Selection of the Expert Review team, (3) Conduct of the Expert Review, and (4) Conclusions and Recommendations provided by the review team to the application developer.

4.2 Overview of Design Principles Used in Reviews

Principles of good user interface design are well defined in the literature. Nielsen and Molich (1990),⁴¹ Norman (1986),⁴² and others have offered sets of principles of good design that can guide designers and developers, but also provide a basis to evaluate existing designs. Nielsen (1994)⁴³ refined this early work and developed a set of principles that are still widely circulated today, including:

- User should easily be able to view system status.
- System should match the real world.
- User should have control and freedom.
- System should maintain consistency and standards.
- System should prevent errors.
- User should be able to recognize rather than recall.
- System should support flexibility and efficiency of use.
- System should have aesthetic and minimalist design.
- System Help should allow users to recognize, diagnose, and recover from errors.
- System should be supplied with Help and documentation.

These more general usability design principles will be affected by the characteristics of the healthcare environment including the users, the tasks they perform, and their work environments. For this reason, advanced work on shaping these design heuristics is being conducted by Zhang and his collaborators under the SHARPC grant at the National Center for Cognitive Informatics & Decision Making at the University of Texas-Houston, School of Biomedical Informatics. This program will develop structured expert evaluation / analyses for EHR user interfaces that are more specific to the EHR and healthcare environments. The EHR Usability Protocol advocates the use of expert reviews as a method to identify and rate a design's degree of departure from user interface best practices.

⁴¹Nielsen, J., & Molich, R. (1990). *Heuristic evaluation of user interfaces*. Proceedings from ACM CHI'90 Conference. Seattle, WA, 249-256.

⁴²Norman, D. (1986). Cognitive engineering. In Norman, D. and Draper, S. (Eds.). *User centered system design: new perspectives on human-computer interaction* (pp. 31—61). Mahwah, NJ: Lawrence Erlbaum Associates.

⁴³Nielsen, J. (1994). *Usability Engineering*. San Diego: Academic Press.

4.3 Selection and Composition of the Review Team

The review is conducted with a multidisciplinary team that must include usability/human factors expertise and clinical expertise. Such professionals are selected based on criteria of education, experience, and expertise in human factors, user interface design and health IT. It is recommended that the professional leading the review have (a) a degree in a human factors discipline, and (b) extensive experience with testing EHRs and other health information technologies. Examples of degrees that pertain to human interaction with computing technologies include:

- Applied Experimental Psychology
 - Applied Psychology, Human Factors, or Engineering Psychology
 - Applied Science with an Emphasis in Human Factors or Usability
 - Human Computer Interaction
 - Human Factors
 - Human Factors and Applied Psychology
 - Human Factors and Ergonomics
 - Human Factors in Information Design
 - Industrial Engineering with Specialization in Human Factors or Usability
 - Industrial and Operations Engineering
 - Usability Engineering
- Professional experience is also important. Experts are expected to also have some of the following demonstrated capabilities:
 - Assessing human performance against usability requirements;
 - Performing expert reviews based on usability principles that can be applied to human interaction with health information technology applications;
 - Leading sessions where two or more expert reviewers discuss findings;
 - Explaining rationale behind judged failure to comply with guidelines and requirements;
 - Reporting expert review findings;
 - Assessing conformance with common accepted standards of user interface design;
 - Evaluating usability testing results including summative usability studies; and
 - Reporting impacts of findings on the user experience.
- Experience and expertise in user interface design and testing of EHRs and/or other health information technologies.

Clinical experts participating in the review are expected to be medical professionals (i.e., nurses, doctors) with some knowledge of medical informatics and are recommended to have at least some experience with or awareness of usability issues.

4.4 Review Process Overview

The process is similar to that used in the SHARPC program and is adapted from Nielsen⁴⁴ as follows:

1. Based on the submitted Application Analysis from the EHR application developer, an expert reviews the interface in total and then inspects each user interface element with a set of design practices in mind (e.g., those from Nielsen, above). Nielsen recommends experts go through each principal page (or screen) of the interface at least twice: the first time to understand the interaction flow and general scope; the second time to focus on specific interface elements in the context of the whole interface. The interface inspection is conducted with regard to general usability principles and also on the use errors identified in Figure 3. *Appendix B* provides a detailed form for evaluating the user interface based upon the research synthesis from Section 2.2 and applicable best practice design heuristics for usable interface design. This evaluation form (particularly Section 1) is an adaptation and extension of usability guidelines to evaluating EHRs based upon the authors' guidance. The steps are as follows:
 - Gain access to the EHR.
 - Decide on (1) the scope of the review (which pages, flows, processes, etc., are to be reviewed), and (2) the scenarios and data sets to be used during the review. We recommend using sample cases similar to those outlined in *Appendices C, D, and E* as the data.
 - Each expert reviewer proceeds to review each page and evaluate that page with the relevant items from the checklist from *Appendix B*.
 - When issues are identified, in the comment field, the expert documents (a) the page, (b) the context, (c) the severity, and comments on why this problem is an issue. The expert may want to annotate the findings with images of the screen. The experts should specifically *list each problem separately* along with a severity rating. Severity is judged on an ordinal scale from 1-4, with 1 being low and 4 being catastrophic.
2. In order to assure independence in the review process, each expert reviews the user interface on his/her own and produces a written report. In general, each review should be a thorough review of the major elements of the user interface. The written report should include a review of each page with any violations noted with comments and justification about how and why good practices are considered violated. Ultimately, the expert provides a completed expert review form that must be consolidated with the reviews done by other experts.
3. Once all experts have completed the evaluations, a lead expert consolidates all the findings into a single report. When the aggregation of the findings takes place, the severity ratings can help to identify the frequency (did multiple experts see it?), impact (will the users overcome the problem easily?), and persistence (is it a one-time issue or recurring problem?). This combined

⁴⁴ Nielsen, J. *How to Conduct a Heuristic Evaluation*. http://www.useit.com/papers/heuristic/heuristic_evaluation.html. Accessed on Sep 10, 2011.

report allows for comparison and discussion among the experts. The expert reviewers will review the findings and reach consensus on the problem list, and a final report is produced.

Expert evaluation and review is a method for finding problems, not necessarily coming up with ways of fixing those problems. However, since the screens/designs are viewed through the lens of usability best practices, there are often fixes that are obvious to those responsible for making the change. The final report of the expert evaluation should be used with the appropriate team members to determine actionable recommendations for mitigating the issues identified. In turn, the recommendations should be implemented before Step 3, so that the summative test does not merely re-identify issues identified in the expert evaluation. To be most effective and identify the greatest number of problems, more than one expert evaluator is used.⁴⁵

⁴⁵ Nielsen, J., & Landauer, T. K. (1993). Proceedings ACM/IFIP INTERCHI'93 Conference: *A mathematical model of the finding of usability problems*. Amsterdam, 206-213.

5 Step III: EHR User Interface Validation Test

5.1 Purpose of Summative Validation Testing and General Approach

This section sets out general testing conditions for the conduct of the EHR User Interface Validation Test. *To distinguish conditions and steps required to provide an accessible test environment for participants with disabilities, those conditions and steps are set out in italics.*⁴⁶

Because performance difficulties and errors that occur when using EHRs are so highly variable and contextual, simply counting “failures” is insufficient to understand the usability and safety of the system. A necessary (though somewhat more challenging) test process involves describing the failures as well as their outcomes and whether patterns of similar or identical failures occur. At a late stage of design, residual problems with system usability need to be understood so that a determination can be made as to whether they can be eliminated or reduced in frequency by further modifications to the system, or if they are acceptable (i.e. not critical). Doing this kind of data-driven analysis requires a combination of objective performance observation and post-test inquiry with participants to identify the source or “root cause” of any observed problems.

Successful Validation Requires Formative Work. The authors understand that meaningful validation testing must be coupled with a user-centered design approach, and fully support and encourage formative user research be incorporated into the development and design of new EHR products. The role of formative testing is to support design innovations that would increase the usefulness of a system and provide feedback on the utility of an application to an intended user, e.g., would the user use it? The role of summative research is to validate the usability of an application in the hands of the intended users, in the context of use for which it was designed, performing the typical or most frequent tasks that it was designed to perform. A well-designed application that has benefited from careful, user-centered design and formative research should, in theory, result in a successful validation test, as outlined in this section. The summative test is therefore a tool to help HIT developers confirm the validity of their assumptions about the use cases, and verify that the design principles applied in development of the application succeed in assisting the user with task completion without critical error.

Validation Testing Environment. For the purpose of providing detailed technical guidance, the study design outlined in this section assumes certain approaches for which there are alternatives, each of which involves its own trade-offs. Perhaps the most significant of these is the trade-off between generating cleaner, more generalized data sets in a controlled lab environment and stressing user performance in a simulated clinical environment. The EUP recommendation is that validation testing be carried out in a controlled lab environment, rather than in a simulated clinical setting. EHR systems are highly configurable and provider roles, workflows, and work environments can be highly diverse in ways that are difficult to replicate and control in a lab environment. However, there is a clear rationale for

⁴⁶ Note that while we do point out key areas that should be considered for testing people with disabilities, this document is not intended to be used as guidance for testing an EHR for accessibility.

using a more closely managed test environment for summative testing. Testing in a simulated environment will reveal the most usability issues—both simple issues and more of the problems that are highly contingent on context. However, any patient-critical errors that can be observed during a validation test in a lab setting would only be amplified in a clinical environment. Therefore, lab testing *will* reveal issues that would also be observable in a clinical environment, but cannot guarantee that a tested interface is 100 percent risk-free—no volume or method of testing can make that guarantee. The primary goal is to mitigate simple, unnecessary risk, knowing that not every possible risk can or will emerge in testing.

Validation Test Participants. Another important parameter of the summative test study design outlined in this section is the makeup of participants tested and how participants might be trained during the study. EHR applications are not intended to be walk-up-and-use applications. Typically, EHRs require an investment of time for the user to be proficient at day-to-day use. For the purposes of the EUP, this fact requires some means of accounting for recruiting participants with predefined levels of competency, or getting participants to attain a certain level of competency through some means of training prior to the evaluation.

Participants should be a representative sample of all user groups of the EHR application being tested as defined in the EHR Application Analysis (Step I), which may include physicians, physician assistants, nurse practitioners, nurses and/or all ancillary, technical, and trainee roles in multidisciplinary teams across diverse settings. The training and knowledge of different users according to their individual roles is accounted for in the study design, either by giving different user groups different training conditions and/or tasks to perform, or by identifying participants and analyzing their test performance by their pretest level of competency.

Training Prior to Testing—Important Trade-offs. If training will be provided during the test session, the test team should take care to ensure that training is administered consistently and correctly, in a way that simulates realistic training conditions, as defined prior to the test. While training variability for real users is inevitable, systematically controlling the training protocol in a summative test will help the team understand the impacts of that variability.

In selecting participants, the following participant backgrounds might be considered:

1. Participants with no experience on EHRs;
2. Participants with experience on an earlier version of an EHR made by the same vendor; and
3. Participants with prior experience on a different EHR, but with no experience on EHRs made by the same vendor.

Each of these participant types presents its own considerations, so the test team is advised to select the approach that is most practical. Finding participants with no EHR experience is probably out of the question as such people will be increasingly hard to find, and will likely take too long to get to an adequate level of proficiency in a reasonable period of time. Participants who are existing users of the vendor EHR present a couple of considerations. First, current users often have highly customized

interfaces and may have developed shortcuts that could result, in some cases, in worse performance than novices who are unfamiliar with the system. For example, a study by Frensch and Sternberg showed that expert bridge players initially did worse than novices, when some of the rules were changed.⁴⁷ Second, they might be more likely to come to the session with a sense that they should voice their opinions during testing, which is not the preferred technique for validation testing. Participants with experience on a different vendor's EHR have already made the modal jump from paper to electronics, and generated a mental model of how these applications work. The main concern is that such users will have different levels of experience and differential knowledge based upon any given system used.

On balance, each of these participant backgrounds has issues and carries with it certain trade-offs that require management. If a completely new system is being tested, it is reasonable to teach participants in a way that would be consistent with real-world training. If the system being tested is a new version or release, it is reasonable to test participants who are already users of the existing interface, and perhaps compare their performance against that of new users, especially if completely new users would be adopting the system in the future. However, for the purposes of the EUP, it is recommended that participants have at least one year of consistent clinical use of an EHR. The application development team should develop and provide the test team with any self-guided training that is made available to end users, such as manuals, help systems, or tip sheets. If in-person training is provided during the test session, training should present as realistic a program as possible given the time and resources available for testing. It is understood that testing immediately following training biases performance, but as a practical matter, it may be a compromise that is made given the time, cost, and complexity of this type of testing. Potential suggestions for managing this include consulting with statistical experts as to whether to stratify the experiment, control for experience, or statistically manage variation during analysis.

If participants require significant training, then consideration has to be given to the length of the testing session and how the session is managed. One option may be to bring participants in for a training session and ask them to come back at a later date (usually a couple of days to a week) to perform the test.

Moderator Guidance. As with the expert review, a qualified usability/human factors professional oversees the overall validation testing. Such professionals are selected based on criteria of education, experience, and expertise in human factors and health IT. It is recommended that the professional have (a) a degree in a human factors discipline, and (b) extensive experience with testing EHRs and other health information technologies. These recommended qualifications do not extend to all members of the testing team ("testers"), which may also include other experienced moderators and test team

⁴⁷ Frensch, P. & Sternberg, R. (1989). Expertise and intelligent thinking: When is it worse to know better? In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*: Vol. 5 (pp. 157–188). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

support staff such as note takers, additional analytical observers, and technical support for the lab set up.

At least two testers are needed to conduct the sessions. These two testers are:

1. An expert/test administrator who facilitates the testing and is in charge of interacting with the participant during the test session. The expert/test administrator must be skilled in usability testing and human factors; be familiar with the test plan and procedures; and understand how to operate the EHR Application Under Test (EHRUT). *In the case of testing participants with disabilities, the expert/test administrator must also be an expert in accessibility.*
2. An assistant/data logger who is responsible for all aspects of data collection. The assistant/data logger must be an expert in usability, usability testing, and human factors; understand the space required and configurations required for usability testing; be familiar with the test plan and procedures; and understand how to operate the EHRUT. *In the case of testing participants with disabilities, the assistant/data logger must also have some expertise in accessibility.* The data logging role can be fulfilled with data capture software where appropriate; however, a two-person test team is the minimum recommendation.

Other Testing Personnel Roles. In addition to the core expert/test administrator and the assistant/data logger responsibilities, there may be other roles to fill over the course of the preparation for and administration of the test. The expert/test administrator and assistant/data logger may fill these additional roles, or they may alternatively be filled by additional testers. For example, the expert/test administrator and assistant/data logger may serve as recruiter, test facility selector, greeter, moderator, or systems administrator.

1. Testers who act as recruiters must be skilled in interviewing. *In the case of recruiting participants with disabilities, they must be skilled in interacting with people with disabilities.*
2. Testers who ascertain the appropriateness of the test facility must be skilled in usability testing. *In case testing will involve participants with disabilities, they must be familiar with the Americans with Disabilities Act (ADA), as well as Occupational Safety and Health Act (OSHA) guidelines and other guidelines for accessible buildings.*
3. Testers who set up the testing environment must be skilled in usability testing and must know how to set up and operate the EHRUT. *In case testing will involve participants with disabilities, they must have an understanding of ADA, as well as OSHA guidelines and other guidelines for accessible buildings; they must have an understanding of accessibility devices and the requirements for accommodating them in setting up the testing environment.*
4. Testers who act as systems administrators must be skilled in setting up and closing down the testing environment. They must understand how to operate and assemble all equipment to be used in the test including the EHRUT, and if they are to be used, cameras, microphones, and computers. *In case testing will involve participants with disabilities, they must also be expert in the operation and assembly of assistive devices.*

Additional experts, whose roles are not defined in this document, may participate in data analysis and reporting together with the expert/test administrator and the data collector.

5.2 Overview of Validation Testing Steps

We strongly urge the test team to create a systematic test plan protocol before beginning any validation test. Elements of testing can be found in many published volumes and are also outlined in Section 9 of NISTIR 7741 - [NIST Guide to the Processes Approach for Improving the Usability of Electronic Health Records](#).

Plan Elements. The validation test plan defines:

1. **Participant Composition.** Describe the selection criteria for the appropriate test participants corresponding to the user groups to be tested as outlined in the Application Analysis.
2. **Task Selection.** Provide a rationale for what user tasks are being tested, and how these test tasks relate to identifying user interaction that could lead to potential medical error. Identify what Scenarios of Use include these critical safety-related tasks.
3. **Testing Environment.** Describe the environment of testing and how it is representative of real-world application use in terms of lighting, noise, distraction, vibration, and other conditions in the workplace.
4. **Performance Assessment of Failures and Difficulties.** For each critical task identified (item 2 above), describe the technique for classifying observed performance as being successful, successful with difficulties or effort, and failures.
5. **Facility and Equipment.** Describe the facility and equipment required to represent final or near-final design of the user interface, including operation on the appropriate application platform(s) that will be used (e.g., desktop computer, smart phone, touch screen, tablet PC, etc.) Where appropriate, the application should be tested on multiple platforms, as an EHR can have different appearances on different platforms. For example, an EHR that is available on a desktop PC and also on a tablet PC might be considered to be separate and distinct software products, even though they are distributed by the same company.
6. **Training Plan.** Describe the realistic training scenario that will be administered to participants prior to validation testing and how the potential training-to-performance learning decay will be represented in the test.

Once the test plan has been written, testers can follow the major steps of the protocol sequentially:

1. Develop screener, recruit, and schedule participants.
2. Set up the EHRUT system(s) (and any other relevant materials that the participants will interact with) and write moderator's guide.

3. Set up test environment with equipment and furnishings to simulate a real clinical environment and capture audio and video recordings of the session.
4. On testing days, greet, orient, and instruct participants prior to each session.
5. Conduct the test.
6. After each session, debrief participants and ready the test environment for the next participant.
7. Store and analyze data, and report the results.

Step 1. Recruit and Schedule Participants

The test team will write a recruitment screener (see *Appendix F* for an example) that operationalizes the participant criteria in a series of interview-style questions. The collective responses of the prospective participants to the screener interview determine whether or not individuals are suitable for the study.

To be recruited for the study, prospective participants are expected to be:

- Currently practicing as a professional healthcare provider; and
- Capable of proficiently reading, conversing, and expressing themselves in English.

Prospective participants are expected NOT to be:

- Commercially connected to any EHR vendor, e.g., no close relative as an employee or owner;
- Have a background in computer science;
- A member of the application development team or related IT development committees; and/or
- A super user that has special or additional knowledge beyond that of a typical user.

All other relevant inclusion criteria are specified in the screener, including age, gender, role, years of experience, current or prospective user, etc.⁴⁸ Typically, the test team will have to over-recruit (schedule more than that target number of test sessions) to allow for subjects who do not make the appointment on time, who are discovered to be ineligible upon arrival, or who do not complete the session. An additional strategy to ensure that a sufficient sample size is reached is to recruit stand-by “floaters” who are called on to participate in case any primarily scheduled participant fails to make the appointment.

The testing should meet all federal and state legal requirements for the use of human subjects.

⁴⁸ For additional guidance on physician population demographics, see for example *AMA Physician Statistics, AMA Physician Characteristics and Distribution in the US 2008, 2009*.

For additional guidance on mid-level provider population demographics, see [2008 Health Resources and Services Administration \(HRSA\), Bureau of Health Professions, Division of Nursing](#) or [American Academy of Physician Assistants Census National Report 2009](#).

For additional guidance on nurse participants population demographics, see [2008 Health Resources and Services Administration \(HRSA\), Bureau of Health Professions, Division of Nursing](#).

Step 1.1 Sample Size

The summative validation study should employ an appropriate sample size that will identify residual user interface problems. This determination follows paradigms established from several decades of usability testing.

Generally, the larger the number of users tested, the more design flaws detected. Faulkner (2003) reports that with ten participants, 80 percent of the problems are found whereas 95 percent of the problems are found with twenty participants. The FDA's recommendation for summative validation tests of medical devices is to test a minimum of fifteen users per distinct user group.⁴⁹ Sauro (2009) presents a very readable discussion on the topic; he, too, settles on a sample of about twenty representative participants per user group to capture most of the variance at a practical level of effort. Therefore, the EUP recommends that fifteen to twenty participants per distinct user group (as defined in the Application Analysis) be tested. The test team should clearly document the rationale for any determinations made about the number of user groups and the sample size, which can have serious implications on experimental design and the validity of the research.

Step 1.2 Recruitment Screener

The purpose of a screening questionnaire ("screener") is to identify and select participants who match the profile of the target or representative user group. The screener provides a set of inclusive and exclusive criteria for the various users that are going to be tested. Once a valid participant has passed the screener, they are invited to participate. The number of user groups is laid out in the screener as well as the total number of recruits per user group. A sample recruitment screener can be found in *Appendix F*, as copied and adapted from NISTIR 7742 *Appendix 1* of [Customized Common Industry Format Template for Electronic Health Record Usability Testing](#).

Checklist: Recruiting

- A recruiter screens potential participants.
 - A recruiter ascertains that persons being screened fit the requirements for participants.
 - A recruiter ascertains that persons being screened fit the requirements for any form of disability that needs to be accommodated during the study (if study will include accessibility testing).*
- A recruiter recruits participants.

⁴⁹ A user group here refers to a distinct healthcare provider role, e.g., nurses and physicians are distinct user groups. Emergency Room (ER) and Neonatal Intensive Care Unit (NICU) nurses may or may not be distinct user groups, depending on the level of difference between their respective tasks and environments. The recommendation is to test at least fifteen participants *per user group*, and should not be interpreted as a "per product" or "per study" guideline. See also *Appendix B* of FDA's draft guidance *Applying Human Factors and Usability Engineering to Optimize Medical Device Design* (June 22, 2011).

- A recruiter invites person(s) fitting the requirements for testing to participate in a test session.
- A recruiter schedules participants.
 - A recruiter assigns a test time to participants at intervals appropriate to the test protocol for the specific test.
 - In scheduling, a recruiter allows the appropriate number of minutes for each test session as appropriate to the test protocol for the specific test.
 - In scheduling, a recruiter sets aside the appropriate number of minutes between sessions for resetting the EHRUT, reconfiguring the testing environment as necessary, and giving testers time to refresh themselves between sessions.
 - A recruiter informs the participant of the scheduled test time.
 - A recruiter informs the participant of the test location.
 - A recruiter obtains participant contact information.
 - A recruiter may also want to accept contact information about a care-giver who will accompany the participant to the test location.*
 - A recruiter inquires whether a participant will require an escort from the entrance of the test facility to the testing environment.*
 - A recruiter informs the participant that he or she will be compensated.
 - A recruiter informs the participant of the amount of the compensation.

Step 2. Set up EHR Systems and Materials, Write Moderator’s Guide

Step 2.1 Configuring the EHR System for testing

Most EHR applications are intended to be customized to the site/practice/hospital and often personalized to the user. For the purposes of testing, the application development team must verify the application in the configuration that it judges is the most typical. Systems must be tested on the *minimum* hardware, software, and network configurations recommended by the application development team.

While not ideal, in some cases, application development teams may find it very difficult or impossible to install an instance of the application locally on a machine/network in a test lab. In these situations, application development teams may opt to allow the user to operate the application on a remote desktop using a screen sharing application (many such freeware and commercial products are in use today). In such a case, the system response times will be mildly inflated; a note must be made in the results describing this qualification.

The EHRs must have patient records and histories loaded as data sets. The scenarios can be similar to the ones provided in *Appendices C, D, and E*, but must be specific to the application being tested and to the associated risks previously identified for that application.

Ultimately, the application development team is responsible for setting up and validating that the technical set up is to their specifications and satisfaction, with appropriate caveats noted.

Step 2.2 Moderator's Guide and Test Tasks

In order to maintain consistency during the evaluation, testers follow a scripted moderator's guide with the participants during each session. An example of such a script can be found in *Appendix G*.

The EUP recommendation is to focus the tasks in the moderator's guide (at least initially) on functions that potentially could have impact on patient safety related to Meaningful Use (MU) criteria.⁵⁰ Many of the MU criteria have a significant human factors or usability component. For instance, the e-prescribing criteria for Stage 1 requires that more than 40 percent of all permissible prescriptions written by the eligible providers are transmitted electronically using certified EHR technology. The objective is clear and the functional requirements are seemingly straightforward. However, from a usability standpoint, there is potential for error because of the implementation of the functionality. For instance, if the medication formulary list is truncated in the user interface and does not show the dosage when it is displayed, there is the potential that the patient might get an improper drug or an improper dosage. In order to test whether or not this potential risk results in use error, the moderator's guide would include a scenario that involves e-prescribing. Having participants do the tasks that are necessary to meet MU criteria would reveal whether or not they are potentially harmful to patients. If participants fail the e-prescribing task (or any task in the moderator's guide), it would not be because of inadequate or limited software functionality that made the task impossible. Rather, it would show that the functions do not provide sufficient and usable information to guide the user in successfully completing the task.

Sample test tasks are presented in *Appendices C, D, and E*. We recommend using test cases that are developed so that they can be evaluated on both a clinical and usability level and include challenging scenario elements such as nonstandard formularies, dosing, and scheduling, and imprecise, incomplete, and inaccurate data. The Application Analysis (Step I of the overall protocol) identifies various applicable Scenarios of Use for the given application.

Step 3. Set up Test Environment

It is not easy to recreate a test environment that simulates the clinical environment. As recommended above, we suggest that testing take place in a controlled lab environment. The key is to create safe, comfortable, and distraction-free environments to reduce the likelihood that performance issues cannot be attributed to exogenous environmental factors. There must be sufficient space and environmental comfort in which to carry out the testing. The test facility should have the following characteristics:

- Test room should be consistent and compliant with OSHA standards for an office environment.

⁵⁰ Meaningful Use Stage 1 Final Rule can be found at: <http://edocket.access.gpo.gov/2010/pdf/2010-17207.pdf>

- Minimum dimensions of 10' x 10'
- Ambient lighting in the range of 400-600 lx. If possible, use indirect lighting rather than overhead fixtures or direct sunlight so as to reduce screen glare
- Ambient noise levels below 40dB
- Adequate ventilation such that the participants do not report a stuffy or drafty feeling
- Comfortable temperature, between 68 and 76 Fahrenheit, such that participants are not too chilled or too drowsy
- Comfortable relative humidity levels between 20 percent and 60 percent
- Table and two chairs (the participant's chair stationary so as not to encourage rolling out of camera view)
- (Optional) One-way mirror or camera and microphone allowing observers to remotely view the participant who is using the application

The application development team may choose to perform the testing at an internal (i.e., in-house) lab or at a third-party facility as long as it meets the key characteristics specified above. See [this OSHA guideline](http://www.osha.gov/dts/osta/otm/otm_iii/otm_iii_2.html) (http://www.osha.gov/dts/osta/otm/otm_iii/otm_iii_2.html) for more detailed recommendations.

Checklist: A Tester Ascertains ADA compliance of testing facility

- A tester ascertains that a test facility is accessible for parking, sidewalks, pathways, hallways, doorways, corridors, and bathrooms per ADA guidelines, available at <http://www.ada.gov/stdspdf.htm>. Questions about test facility accessibility can be answered by the United States Access Board, info@access-board.gov, phone toll free: (800) 872-2253.
- A tester assures that the evacuation procedure in place for the test facility provides for evacuation of participants with disabilities.
 - In multistory facilities, where the testing environment may be located above or below ground-level exits, a tester assures the availability of evacuation chairs to enable emergency evacuation of individuals with mobility impairments on stairs.
- A tester assures that there is access to an ADA-compliant drinking fountain available.
- A tester assures the availability of clear floor space in the room where the EHRUT will be tested.
 - A tester assures that this available space is level with no slope exceeding 1:48.
 - A tester assures that this available space is positioned for a forward approach or a parallel approach.

Checklist: Testers Set up the Testing Environment

- A tester may wish to make snacks and drinks available to participants and observers.

- A tester assures that the testing environment has room for all testers and observers to work without disturbing the participant, interfering with data collection, or interfering with participants' interaction with the EHRUT; *including testers and observers with disabilities.*
- A tester assures that furniture in the testing environment accommodates participants, testers and any observers, *including people with disabilities. For example, a tester assures that work surfaces are adjustable to a height that accommodates participants using manual dexterity assistive devices and that the proper toe and knee clearances have been provided.*
- A tester assures that seats for testers are placed so that testers can observe the participant, but in a configuration whereby testers do not block a camera or distract the participant.
- A tester sets up a place for a data collector to record observation data.
- A tester assures that the test facility and the testing environment are kept free of obstacles, including video recording equipment, throughout all test sessions.
- A tester assures that the test environment and the route(s) to and from the test environment is/are free of all obstacles, e.g.,
 - obstacles on the floor
 - obstacles that hang from a ceiling
 - protruding obstacles

Checklist: Set up the EHR in the test environment

- A systems administrator is responsible for setting up the EHRUT in the test environment.
 - If the EHRUT must be installed by the development team, the system administrator must provide the correct hardware and software environment.
 - If the EHRUT can be installed by the systems administrator, the systems administrator follows the instructions provided by the EHRUT development team.
 - If the EHRUT is to be used remotely, the systems administrator must coordinate with the development team to ensure all screen sharing functions are operating correctly.
 - A systems administrator assures that the EHRUT displays and responds correctly in the test environment.
- If assistive devices are provided by the testers, a systems administrator sets up the assistive device(s) to be used with the EHRUT according to the assistive device manufacturer's instructions and according to any relevant instructions provided by the EHRUT manufacturer.*
- A systems administrator designs a recording solution for the test environment and sets up the recording equipment in the test environment.
 - If a camera is used, a systems administrator sets up the camera so that it will not interfere with the participant's interaction with the EHRUT.
 - A systems administrator sets up the audio recording environment both to capture and record the participant's comments, and to pipe audio from the test session into an observation room where observers and the data logger can hear.
 - A systems administrator is responsible for ensuring that the data capture and logging software is properly capturing data on the EHRUT.

- A systems administrator ensures that assistive devices are available to meet the needs of any participants with disabilities who have been recruited for the study.

Step 4. Greeting and Registering Participants

On test days, the tester who acts as greeter is responsible for greeting participants upon arrival, and verifying that they are there for the appropriate purpose. *If, during recruitment, it has been established that the participant requires an escort to the testing environment, the greeter meets the participant at the entrance to the facility and escorts the participant to the test environment.* The greeter then provides each participant with an informed consent form to sign (if required, see *Appendix H* for an example of a consent form) and escorts the participant to the test room at the appropriate time. The greeter should take care not to discuss with the participant any information that might bias them in any way about the upcoming user testing. The goal is to minimize, if not eliminate entirely, any “tester effect,” or influence on the participant prior to testing.

Checklist: A Greeter Welcomes and Orients the Participant

- A greeter welcomes the participant to the testing environment.
 - A greeter welcomes the participant when he or she enters the testing environment.
 - A greeter verifies that the participant has come to participate in the EHRUT test.
 - A greeter verifies (by means of ID) that this is the participant scheduled for this session.
- A greeter shows the participant to the area where he or she will complete paperwork prior to the test.

Checklist: The Participant Completes the Consent Form

- A greeter gives the participant two copies of the consent form and directs the participant to read the consent forms.
 - If, for any reason, the participant requests that the consent form be read aloud, the greeter reads the consent form aloud for the participant.
 - The greeter asks the participant if they have any questions about the consent form, and answers any of the participant’s questions about the consent form.
- The participant signs both copies of the consent form.
- A greeter witnesses the participant’s signature on both copies of the consent form.
 - If the greeter has witnessed the participant signing the consent form, the greeter signs the consent form as a witness.
- If, for any reason, the participant cannot sign the consent form, the greeter asks the participant if he or she consents.

- If the participant consents, the greeter notes this on the consent form.
- If the participant does not consent, the test is terminated.
- A greeter offers one copy of the consent form to the participant.
- A greeter retains the other copy of the consent form and makes it part of the records of the test session.

Step 5. An Expert/Test Administrator Conducts the Test

Step 5.1 Give Participant General Instructions

Once the participant has been seated in the test room, the expert/test administrator (“tester”) reads the instructions as presented in NISTIR 7742 guidance on participant Instructions. These instructions provide the participant with context for the balance of the testing session:

(1a) *If applicable*: The participant will receive some training on a new EHR interface.

(1) The participant will be asked to perform a number of typical tasks on a fictitious patient record. The tester is not able to help.

(2) Each task will be read to the participant by the tester, and then the participant will receive a card with the task written on it for reference.

(3) The participant is instructed to do the task as if they would at work, and attempt to be as accurate as possible.

(4) At the end of each task, the participant will be asked a few follow up and rating questions.

(5) Once all tasks have completed, there will be final follow up questions and overall ratings.

(6) The participant should be reminded of the total test session time.

Step 5.2 Collect Data

Both the tester and assistant data logger who is observing (in a separate observation room) do data collection. A description of the performance and rating data to be collected (as well as example data sheets) can be found in NISTIR 7742.

Additionally, and critically, since the focus of this protocol is detection and correction of use errors that affect patient safety, documentation of use errors observed should be full and complete. Using the framework from Figure 3 as the foundation, the data collection must include a plan for identification of potential use errors.

There are commercial software tools that will aid in the collection of performance data. However, identification and documentation of use-related errors or user interface issues that might engender errors requires knowledge, experience, and skill. Testers must be familiar with the EHR issues that might relate to patient safety and be able to identify areas of concern whether the user committed an error or not. In other words, the facilitator must behave as a skilled observer and use his or her experience to identify and report on areas of significant concern, whether or not an error was observed. It is possible as well that some errors may only be detected after the user has completed the process (for example, a participant may report in the debrief that he or she *almost* pressed the wrong button), or during secondary review of the test results.

Checklist: An expert/ test administrator conducts the test session

- An expert/test administrator escorts the participant to the test environment with the EHRUT.
- A tester welcomes the participant and directs the participant to sit in front of the EHRUT.
- If a camera is to be used, once the participant is seated in front of the EHRUT, without disturbing the participant, a systems administrator checks the camera position to assure that:
 - The camera will capture the EHRUT screen and the participant.
 - The microphone is close enough to capture the participant's comments.
 - The camera is turned on.
 - Audio and video recording is actively taking place.
- An expert/test administrator reads general instructions to the participant.
 - An expert/test administrator hands the general instructions to participants with auditory disabilities.*
- An expert/test administrator verifies that the participant understood the general instructions.
 - The participant reviews the general instructions.
 - An expert/test administrator asks the participant if there are any questions about the general instructions.
 - An expert/test administrator answers any participant questions about the general instructions.
- An expert/test administrator presents task instructions to the participant.

- An expert/test administrator hands cards with the written task instructions to a participant who has no disability *and to a participant with a non-visual disability who is capable of holding the task instructions.*

or

- An expert/test administrator offers to read the task instructions to a participant with blindness or a low vision disability.*

or

- An expert/test administrator places the task instructions on a table or cart for participants with a mobility or manual dexterity disability.*
- An expert/test administrator verifies that the participant can comfortably consult the instructions.
- The participant reviews the task instructions.
- The participant begins the task.
- Testers do not interact with participants while participants complete test tasks. Exceptions are:
 - An expert/test administrator may read task instructions to participants with blindness or low vision disabilities.*
 - An expert/test administrator may perform certain functions for participants with manual dexterity disabilities.*
 - An expert/test administrator may respond to a request for help when the participant cannot proceed further with a task, but this will result in that task receiving a fail rating.
 - An expert/test administrator may announce the end of a task if the participant is unable to complete it in sufficient time to accommodate all other test tasks in the guide.
- The participant attempts to perform all test tasks.
- The participant completes the post-test questionnaire (System Usability Scale)⁵¹ found in *Appendix I*.

Checklist: An Assistant/Data Logger Captures Data during Testing

- During test sessions, an assistant/data logger observes participants and collects data while participants perform test tasks and interact with the EHRUT.
- An assistant/data logger captures observation data.
 - Prior to the arrival of the first participant, the assistant/data logger prepares the observation data collection form that he or she will use.
 - If a computer is used for observation data collection, an assistant/data logger resets the computer to an unpopulated version of the data collection form, identifying the session, for

⁵¹ The System Usability Scale provides a composite measure of usability that is used to benchmark the usability of a system to other systems. For example, if a system receives an aggregate score of 60 on the scale of 0-100, the system score may be compared to a meta-analysis of studies across many types of devices/systems to see where the SUS score for the system falls relative to those systems (e.g., average? better than average?)

example, by data session number, date and time, and stating the participant's identification code.

- If a paper form is used for observation data collection, an assistant/data logger enters information into an unpopulated observation data collection paper form for the session identifying the session, for example, by data session number, date and time, and stating the participant's identification code.
- All assistants/data loggers use the same observation data collection form across all participant sessions.
- An assistant/data logger collects observation data of the following types:
 - An assistant/data logger notes whether each individual task has been completed.
 - An assistant/data logger notes whether each completed task has been completed without personal assistive technology.*
 - If assistive technology has been used, an assistant/data logger notes which technology has been used.*
 - An assistant/data logger notes whether each task has been completed within the expected time.
 - An assistant/data logger notes if all tasks have been completed.
 - An assistant/data logger notes critical incidents, based upon the Application Analysis from Step 1.
 - An assistant/data logger takes note of the participant's comments that reflect the participant's experience with the EHRUT.

Step 5.3 Subjective Evaluation: Investigation of Causes of Failures and Performance Issues

Once test task performance and rating data have been recorded, the tester may conclude with a qualitative debrief with the participant. Unlike in formative testing, there is no debrief until after all the tasks have been attempted or completed. In a validation test, extra care must be taken to avoid the appearance or reality of tester influence.

The focus of the debrief is on uncovering any residual usability problems. If the tester observed failures on critical tasks, struggled successes, or self-corrected errors, these instances should be investigated with the participant with probing questions posed so as to determine the root cause of the observed behavior. The objective of this debrief is to determine whether observed problems or failures can be attributed to user interface design, such as poor visibility of information, confusion surrounding how particular features work, difficulty of manipulating controls, etc. Contextual errors, including fear, lack of appropriate knowledge, or incorrect assumptions may not necessarily be attributable to interface design. The goal of the tester during the subjective evaluation is to make a determination as to the root cause of the problematic behavior observed during the test so that attribution to interface design may be differentiated from attribution to other contextual factors.

Step 6. Concluding the session

Once the testing is complete, the tester provides the participant with his/her incentive compensation and thanks the participant for his/her time. The test environment is reset for the next session. After all sessions are complete, the test environment is dismantled and the data is merged and securely stored.

Checklist: An Expert /Test Administrator Presents Compensation

- An expert/test administrator presents compensation to the participant.
- An expert/test administrator requests that the participant sign a receipt for the compensation. An example can be found in *Appendix J*.
 - In the case where a participant with a disability cannot sign or mark the consent form, the expert/test administrator notes on the receipt that the participant has received the compensation.*
- An expert/test administrator gives the participant a duplicate of the receipt.
- An expert/test administrator retains the receipt as part of the documentation of the test session.

Checklist: A Greeter Escorts the Participant from the Test Facility

- A greeter escorts the participant from the testing environment if the participant desires.*
- A greeter escorts the participant from the test facility if the participant desires.*
- A greeter escorts the participant to the parking area if the participant desires.*

Checklist: The Testing Environment Is Reset for the Next Participant

- An assistant/data logger prepares the observation data collection form.
 - If a computer is used for observation data collection, an assistant/data logger resets the computer to an unpopulated version of the data collection form, identifying the session, for example, by data session number, date and time, and stating the participant's identification code.
 - If a paper form is used for observation data collection, an assistant/data logger enters information into an unpopulated observation data collection form for the next session identifying the session, for example, by data session number, date and time, and stating the participant's identification code.
- An expert/ test administrator (with help from a systems administrator if needed) resets the EHRUT to the state in which a new user would find it when approaching the EHRUT for daily use.
 - An expert/ test administrator resets all adjustable aspects of the EHRUT that may have been changed by a participant during the previous sessions, e.g., a systems administrator resets the font size to the standard default value.
- If an electronic questionnaire is used, an assistant/data logger clears all interactions with the electronic questionnaire made by the prior participant.

- An assistant/data logger enters information into an unpopulated version of the electronic or paper questionnaire identifying the next session and stating the user identification code.
- If a camera is to be used, an assistant/data logger prepares the camera to record the next session by inserting a new memory card, cartridge, etc. or creating a new file name which is marked with the user identification and session identification.

Checklist: A tester closes down the EHRUT and test equipment at the end of each testing day.

- If a computer was used for observation data collection, a systems administrator shuts the computer down.
- A systems administrator closes down the EHRUT according to the vendor instructions.
- If a computer was used for a questionnaire, a systems administrator shuts down the computer used for the questionnaire.
- If a camera has been used, a systems administrator turns off the camera.
- If necessary, a systems administrator packs away all testing equipment including, but not limited to the camera and computer.
- A system administrator backs up the data that has been collected since the last time the system was recording.

Checklist: An Assistant/Data Logger Merges and Stores Data

- An assistant/data logger backs up or stores observation data.
 - If a computer was used to capture observation data, an assistant/data logger backs up all observation data collected during the session.
 - If observation data is collected by hand, an assistant/data logger stores this data with the records of the session.
- If the EHRUT has log files, an assistant/data logger stores data from the EHRUT according to the vendor instructions.
- An assistant/data logger removes/resets data from the EHRUT according to the vendor instructions.
 - An assistant/data logger stores all data produced on the EHRUT and assures that it is marked with the participant identification and the session identification.
 - An assistant/data logger assures that all interactions with the EHRUT made by the participant have been cleared.
- If a questionnaire has been used, an assistant/data logger aggregates and secures the completed questionnaire data.
 - If a computer was used to capture questionnaire data, an assistant/data logger backs up all questionnaire data collected during the session.

- If questionnaire data is collected on paper, an assistant/data logger stores this data with the records of the session.
- If a questionnaire has been used, an assistant/data logger assures that the questionnaire data is identified with the participant's identification number and the session identification.
- If a camera has been used, an assistant/data logger removes all memory cards, cartridges, etc. from the camera.
 - An assistant/data logger assures that all memory cards, cartridges, etc. are marked with the participant's identification number and the session identification.
- An assistant/data logger stores all data.
 - An assistant/data logger assures that the data is stored in a way that assures that data integrity will not be compromised.
 - An assistant/data logger assures that the privacy of the participant will not be violated by the way that the data is identified or stored.
 - An assistant/data logger assures that all data is properly identified, identifying the session, for example, by data session number, date and time, and stating the participant's identification code.

Step 7. Analyze Data and Report the Results

NIST Common Industry Format for Usability Test Reports is the Foundation for Reporting Results.

Documentation of the performance results must be provided according to the layout of the Common Industry Format (CIF) for Usability Test Reports (NISTIR 7742). This document provides an excellent starting point for the data analysis team's reporting of the application validation test with augmentation and additions described in the following sections.

Additional Analysis Focused on Failures and Performance Issues. Beyond results reporting outlined in the CIF, a thorough classification, reporting, and analysis of observed errors committed by the participants and those identified by the expert/ test administrator is required. Two essential elements of this reporting are identifying the priority of the safety issue that surfaced and the mitigation plan for correcting that issue. The analysis team should focus on identifying the causes of failures and task difficulties on the critical safety-related steps or tasks. Ultimately, this data will be used to decide what remaining patterns of errors need to be addressed through any modifications of the interface.

Checklist: Testers Analyze Data

- Testers analyze data
 - Each participant's performance on each task is recorded in metrics that are relevant indicators of risk for use error. These metrics may include but are not limited to task time, task success, efficiency (i.e., steps taken/optimal steps), near-misses, success with

corrections of errors during task performance, actual errors, verbalizations concerning performance difficulty, and task ratings concerning task difficulties.

- Usability ratings (such as the System Usability Scale or SUS) are also aggregated to compute an overall score, however, *these ratings alone should not be considered as validation metrics*. Usability ratings should be used to drive post-test questioning of the participants to focus on interface aspects that persist in causing potential task failures.

Checklist: Testers Report Test Findings

- Testers identify the EHRUT by name and version/release number.
- Testers report test findings in the Common Industry Format with the following augmentations.
 - Basic Performance Reporting.** Testers report results based on user performance relative to pre-determined definitions of task success, unexpected behaviors, near-misses, success with error correction, appropriate time to completion, actual errors, etc., as well as summary statistics of task difficulty ratings and/or usability ratings. Table 3 below illustrates the summarization method contained in the NIST Common Industry Format for usability results reporting
 - Analysis of Critical Failure Patterns.** Testers should also provide a detailed summary of all critical safety-related issues. This includes an explanation of the cause for each observed failure and exhibited task difficulty (confusions, hesitations, error corrections) by participants. The analysis should treat each of these instances as a potential “adverse event” worthy of systematic inquiry to determine the cause.
 - Conclusions.** The analysis team should convene a multidisciplinary team including members with clinical risk knowledge to determine if there are consistent patterns of failures on critical safety-related steps or tasks. If these patterns exist, conclusions should provide rationale as to why these issues are not worthy of mitigation through redesign or other methods of risk removal, or, if warranted, provide a plan for further mitigation and retesting of their effectiveness.

We provide an annotated table (Table 2) derived from the NIST Common Industry Format for Usability Test Reports document (Note that special emphasis is given to the explanation of observed failures and other usability issues that surface in the validation test.)

Table 2. Example of tabular data reporting of validation study results. (From NISTIR 7742, Common Industry Format for Usability Test Reports.)

Use Scenario: *[Name of Use Scenario containing the analyzed task.]*

Task: *[Name of task, or individual step level if there are multiple sub-steps.]*

Critical Use Errors: *[List the behaviors that would potentially lead to adverse outcomes.]*

User	Unassisted Task Effectiveness (% complete)	Assisted Task Effectiveness (% complete)	Time	Effectiveness (Mean time on task)	Errors, Failures, Omissions, and Task Difficulties	Assists
1 2 3 ... N	<i>This data is an overall indicator of success in study but not the main objective</i>	<i>Testers should not be “assisting” in validation unless it is needed to continue test. Assists are considered task failures</i>	<i>Time is not of relevance in validation testing unless time is critical to the task. Time measures can point to task difficulties and lead test team to inquire about causes of lengthier times on task.</i>		<i>Each instance of failure to complete the task, omissions, errors corrected, difficulties or uncorrected errors should be listed here for the purpose of post-test inquiry regarding root cause of problems.</i>	<i>Simply document instances when tester had to assist.</i>

6 Conclusion

This document summarizes the rationale for an EUP, which is improving user performance with EHRs through application of human factors best principles, thereby mitigating use errors that could have potential negative implications in patient care.

Within this document there is a detailed description of research findings relating to usability issues and their potential impact on patient care. These findings resulted in the development of a model for understanding usability and patient safety outcomes. Based on this model, the EUP encompasses procedures for (1) EHR Application Analysis, (2) EHR User Interface Expert Review, and (3) EHR User Interface Validation Testing. The EUP is therefore an evaluation paradigm that calls for human factors and clinical expert analysis of EHR user interfaces to identify and mitigate potential patient safety issues.

We describe in detail the overall steps recommended, provide some examples of scripts, data sheets, scenarios, and checklists, and offer some discussion about discretionary topics such as the fidelity of the test environment, participant makeup, and managing variable training conditions. The samples provided in the appendices are meant to be examples only; the test and development teams will need to modify them as necessary for their specific systems, use cases, and context.

It is our expectation that the potential for use errors can be identified and mitigated based on a summative usability test conducted by qualified usability/human factors professionals prior to EHR implementation/deployment.

Appendix A: Government Best Practices of Human-System Integration

Throughout the past four decades, the U.S. Government has systematically increased incorporation of human factors analysis, evaluation and testing requirements for complex systems. Consistent application of human factors and usability validation exists for commercial aviation and nuclear power industry systems; perhaps the most sustained of these efforts has been directed towards military system development and procurement requirements. This process has been labeled Human-System Integration (HSI) and covers several individual program efforts by the armed services. We briefly summarize the history and effectiveness of these programs below to provide examples of human factors and usability evaluation and validation processes that resulted in positive impacts on safety and effective use of systems.

According to Department of Defense (DOD) Directive 5000.2, HSI in defense system procurement is concerned with both the application of Human Factors Engineering (HFE) during weapon system acquisition and modification, and the prediction of HFE consequences on manpower, personnel, training, safety and health/biomedical requirements. The DOD Human Factors Technical Advisory Group (TAG) is responsible for implementation of this directive. This TAG explores how policies, procedures and practice can best facilitate Human-System Integration (HSI) with system development teams. Its emphasis is more on management and communication than on technology, more on acquisition than research and development, and more on the application of HSI and HFE tools than on the tools themselves. Typical topics of interest include RFP preparation, source selection, design analysis, design reviews, interactions among staffs of different services/represented organizations, interactions among human factors engineers and other system engineers, review of contractor data submissions, test planning, evaluation or research products in the application environment, and coordinated research and development request activity.

U.S. Army Manpower and Personnel Integration (MANPRINT) Best Practices

MANPRINT is the U.S. Army's Human Systems Integration Directorate, with headquarters at the Office of the Deputy Chief of Staff, G-1. Its mission is to establish policies and procedures for Army Regulation (AR) 602-2, Human Systems Integration in the System Acquisition Process for new system procurements or revisions to existing systems. MANPRINT's mission is to optimize total system performance, reduce life cycle costs and minimize risk of soldier loss or injury by ensuring a systematic consideration of the impact of materiel design on soldiers throughout the system development process. MANPRINT sets development team requirements and enforces policy via *human system interface assessments*, as appropriate, delineating issues in acquisition programs for acquisition executives that pertain to system design risks related to soldier-system interaction.

The rationale for the MANPRINT initiative began in the 60s, 70s and early 80s, as the Army introduced hundreds of new weapons and equipment into the force. This force modernization was designed to increase Army capability and readiness. The Army turned to technology to generate greater combat

power.

The Army encountered two persistent problems. First, when a new system was put into the hands of soldiers, field performance did not always meet the standards predicted during the system's development. For example, a system designed for a 90 percent chance of a first-round hit achieved only 30 to 50 percent when fired by soldiers. Second, the replacement of an existing system with a technologically complex system generated requirements for more highly skilled soldiers and a higher ratio of soldiers per system for operators, maintainers and support personnel.

These systemic problems were *not* solved by putting more systems in the field, recruiting more highly skilled soldiers, expanding training (as well as increasing training dollars), and increasing the size of the Army. In the 1960s, Dr. John Weisz, Director of the U.S. Army Human Engineering Laboratory, pointed out that we can no longer afford to develop equipment and merely hope that the necessary manpower can be found to operate and maintain it in a relatively short time, especially in wartime.

In 1980, Army commanders concluded that human-system performance assessments were not integrated and were conducted too late to influence the design stages of the system acquisition process. Supporting their conclusion, in the 1980s the General Accounting Office (GAO) published reports attributing 50 percent of equipment failures to human error and stressed the need to integrate Manpower, Personnel and Training (MPT) considerations into the system acquisition process.

In 1982, an Army study showed that the integration of MPT considerations early in the design process could have made a difference in reducing error and preventing accidents and incidents related to user interface design. At this point, General Thurman directed that MANPRINT, focused on manpower and personnel integration, be initiated. Starting as a Special Assistant Office in 1986, it became an official Directorate in the Office of the Deputy Chief of Staff for Personnel (ODCSPER) in 1987.

MANPRINT assessments are conducted by Army system program management directorates and focus on the seven areas of MANPRINT concern: (1) Manpower required, (2) Personnel aptitudes, (3) Training requirements, (4) Application of human factors engineering principles design, (5) System safety and prevention of human error, (6) Health hazards, and (7) Soldier survivability. **Of these, System Safety and Human Factors Engineering process criteria present the most synergy with the EHR process goals.**

Assessments take the form of both written analysis by domain experts and human factors experts within the Army labs, and validation testing conducted in simulated environments. A number of dedicated field labs (Battle Labs) were commissioned in the 1990s to serve the test and evaluation needs for MANPRINT requirements. System development teams must conduct MANPRINT studies throughout the systems development process culminating in a fieldable and testable system evaluated by the appropriate Army field lab.

Nuclear Regulatory Commission (NRC) Human Factors Best Practices

As a result of a history of human system interaction as root cause for nuclear incidents and accidents, the Human Factors staff of the Nuclear Regulatory Commission (NRC) has begun conducting nuclear power plant design certification reviews based on a design process plan that describes the HFE program elements that are necessary and sufficient to develop an acceptable detailed design specification and an

acceptable implemented design to mitigate or eliminate sources of human error in plant operation. The need for this developed as (1) design certification applications submitted to NRC did not include detailed design information, and (2) human performance literature and industry experiences showed that many significant human factors issues arise early in the design process, however, certification documents submitted by the operator did not address the criteria for user interface design process evaluation and testing.

The result was the HFE Program Review Model (HFE PRM, NUREG-0711, Rev.2). It was developed as a basis for performing user interface design certification reviews that include design process evaluations as well as review of the final design. A central tenet of the HFE PRM is that the HFE aspects of the plant should be developed, designed and evaluated on the basis of a structured top-down system analysis using accepted HFE principles. The HFE PRM consists of ten elements: (1) HFE program management, (2) operating experience review, (3) functional requirements and allocation analysis, (4) task analysis, (5) staffing, (6) human reliability analysis, (7) human-system interface design, (8) procedures development, (9) training program development, and (10) verification and validation. Each element is divided into four sections: (1) background, (2) objective, (3) applicant submittals and (4) review criteria. This design review approach has been used in several advanced reactor HFE reviews over the past decade.

Federal Aviation Administration (FAA) Flight deck Certification Best Practice

The Federal Aviation Administration (FAA) flightdeck systems certification program includes rigorous human factors test and evaluation prior to compliance and certification of pilot user interfaces. The process includes both evaluation and simulation or flight testing with the production-level system.

System Evaluations are an assessment of the design conducted by the applicant, who then provides a report of the results to the FAA. Evaluations typically use a display design model that is more representative of an actual system than drawings. Evaluations have two defining characteristics that distinguish them from tests: (1) the representation of the display design does not necessarily conform to the final documentation, and (2) the FAA may or may not be present. Evaluations may contribute to a finding of compliance, but they generally do not constitute a finding of compliance by themselves.

Evaluations begin early in the certification program. They may involve static assessments of the basic design and layout of the display, part-task evaluations and/or full-task evaluations in an operationally representative environment (environment may be simulated). A variety of development tools may be used for evaluations, from mockups to full installation representations of the product or flight deck.

The manufacturer should fully document the process used to select test participants, the type of data collected, and the method(s) used to collect the data. The resulting information should be provided as early as possible to obtain agreement between the applicant and the FAA on the extent to which the evaluations are valid and relevant for certification credit. Credit will depend on the extent to which the equipment and facilities represent the flight deck configuration and realism of the flight crew tasks.

Flight or Simulation Testing is the final step in certification, and is conducted in a manner very similar to the system evaluations above, but is performed on more final production level systems in accordance with an approved test plan, with either the FAA or its designated representative present. A test can be conducted on a test bench, in a simulator, and/or on the actual airplane, and is often more formal,

structured, and rigorous than an evaluation. Bench or simulator tests that are conducted to show compliance should be performed in an environment that adequately represents the airplane environment, for the purpose of those tests.

Flight tests should be used to validate and verify data collected from other means of compliance such as analyses, evaluations and simulations. During the testing process, the flightcrew workload assessments and observed or critical failure classification validations should be addressed in a flight simulator or an actual airplane, although the assessments may be supported by appropriate analysis.

Results of evaluation, testing and analysis are presented to FAA human factors and systems certification experts, appropriately formed and convened by the appropriate FAA certification offices located throughout the U.S. Each instance of convening this body may be unique, depending on the expertise needed from the agency.

Food and Drug Administration (FDA) Pre-Market Approval of Medical Devices Best Practices

The Food and Drug Administration (FDA) process for pre-market approval of medical devices has established an effective process for human factors application in optimizing device use safety [see FDA guidance documents: *Medical Device Use Safety: Incorporating Human Factors in the Risk Management Process* (2000)⁵² and *Applying Human Factors and Usability Engineering to Optimize Medical Device Design* (draft, 2011)⁵³].

The FDA's Center for Devices and Radiological Health (CDRH) Office of Device Evaluation (ODE) has established a Human Factors Premarket Evaluation Team (HFPMET@fda.hhs.gov) that reviews human factors information in device premarket applications and notifications and provides recommendations on whether or not this material indicates that the device can be safely and effectively used by the intended users.

The Agency determines whether a new device submission will be approved or cleared or not based on its regulatory review of device performance data as well as, in some cases, human factors evaluation data. Human factors evaluation is often a critical consideration and consists of a systematic use-related risk analysis that forms the basis for subsequent human factors formative studies and usability validation testing. The validation testing should involve representative users performing simulated-use scenarios that focus on the highest-priority (the most safety-critical and all essential) user tasks. The test data should include a summary of users' subjective assessments and findings with respect to the safety of the use of the device. The test results should demonstrate that the device has been optimized with

⁵² Available online at:

<http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm094460.htm>

⁵³ Available online at:

<http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm259748.htm>

respect to safety and effectiveness. To the extent that use errors occur during validation testing, it is critical to analyze those errors and determine whether the root cause of the errors are associated with the design of the device, its labeling (including the instructions for use), or the content or format of training. The validation testing data should support the conclusion that the intended users use the device safely and effectively.

The human factors evaluation process includes both analysis and testing of the user interface with anticipated users. The process has three elements:

- *Preliminary Analysis*—
 - Identification all prospective device user groups and demonstrate an understanding of their potential capabilities and limitations.
 - Development of scenarios of use—high-level descriptions of user interactions involved when performing specific tasks.
 - Analysis of all prospective environments and conditions of use for factors that may affect user performance.
 - Analysis of the required perceptual, cognitive and physical demands associated with device interactions.
 - Analysis of previous use-related hazards with similar devices, and identify critical design shortcomings that could affect patient/user safety.
- *Formative Evaluation*—Systematic and iterative evaluation of the user interface and instructions for use through usability assessment methods such as expert reviews and usability testing, specifically focused on removal of use-related problems and retesting of design modifications to address these problems.
- *Validation Testing*—Formal usability tests conducted with representative users and production-level user interfaces designed to identify any use-related problems that could negatively affect patient safety or healthcare outcomes. This testing involves an analysis of any use-related problems that were observed and post-test identification of the root causes of the problems. User interface-related causes should be mitigated and solutions should be retested for their effectiveness and to ensure that no new use errors or problems were introduced.

The three-tiered process listed above follows the general expectations of the Federal Code of Regulations Quality Systems Regulation’s Design Controls Process (21 CF reg. Section 820.30), instituted in 1996 by the FDA to ensure that device manufacturers follow a prescribed design process. Human Factors methods that are applicable to the three elements above map onto the design controls process:

- Preliminary analysis element applies human factors methods to the design concept and design inputs stage of design controls, creating user profiles, use scenarios, and environmental analysis, task analysis and use error analysis.
- Formative evaluations, such as expert reviews and cognitive walkthrough testing, apply to the design outputs and verification stages of design controls.
- Validation testing requirement matches the need to test the device in simulated or actual use environments per the Validation stage of Design Controls.

Appendix B: Form for Expert Review

This review form is adapted from the form available at the [Usability Toolkit](http://www.stcsig.org/usability/resources/toolkit/toolkit.html) (<http://www.stcsig.org/usability/resources/toolkit/toolkit.html>). The original form is copyrighted by Usability Analysis & Design, Xerox Corporation, 1995 and adapted from:

Weiss, E. (1993). *Making Computers-People Literate*. ISBN: 0-471-01877-5

Nielsen, J, & Mack, R. (1994). *Usability Inspection Methods*. ISBN: 1-55542-622-0

Expert Review of EHR

System Title: _____

Release #: _____

Evaluator: _____

Date: _____

Each violation of a usability principle is categorized as follows:

Rating	Severity
4	Catastrophic: Potential for patient mortality.
3	Major: Potential for patient morbidity.
2	Moderate: Potential for workarounds that create patient safety risks.
1	Minor: Potential for lower quality of clinical care due to decreased efficiency, increased frustration, or increased documentation burden or workload burden.
0	No Issue / Not applicable

The system should protect the user and patient from potential use errors. Items 1A through 1H are principles of good design that help identify areas that might engender use error in EHRs. Items 2-13 are general principles of good user interface design.

1A. Patient identification error

Actions are performed for one patient or documented in one patient's record that were intended for another patient.

#	Review Checklist	Severity Rating	Comments
1A.1	Does every display have a title or header with two patient identifiers?		
1A.2	When a second patient's record is open on the same workstation from the same user login at the same time, is an alert to the increased risk of wrong patient errors (with the ability to override an automatic close to the first record) provided to the user?		
1A.3	When a second user opens a patient chart, are protections in place to protect data integrity for simultaneous data entry? If a lockout feature is employed (so only one user can change data at one time), can users view which user is locking out other users at that time?		
1A.4	When an integrated application (e.g., imaging) is opened from within the EHR, does the display have a title or header with an accurate unique patient identifier (i.e., displaying the previous patient's identifier information is avoided when there is a broken link or inability to access the correct information)?		
1A.5	When an integrated application (e.g., imaging) opened from within the EHR remains open, and a new patient record is opened, does the patient identifier and associated data update accurately?		
1A.6	If an action will cause data to be destructively overwritten with another patient's data, is the user alerted?		
1A.7	If there are other patient records with highly similar identities (e.g., Jr., multiple birth patient, same first and last name) that increase the risk of wrong patient errors or having two files for the same patient (e.g., due to data entry errors at registration, name changes, variations in names such as Jr.) where critical information is not included in both files, are the similar patients highlighted for the user just prior to final		

	selection of the record?		
1A.8	If multiple records for a patient are being merged, are users provided clear information about what the impacts of the pending merge will be to the merged patient data (e.g., permanently overwritten data for one patient)? Is there a way to “undo” the potentially destructive operation immediately following? Is there a way to trace back what a record was previously labeled as following a merge?		
1A.9	If information is copied from the record of one patient and pasted into another, is feedback provided to anyone viewing the record what specific information was pasted from the record of a different patient (e.g., by having a subtle background color around copied text) in order to aid detection of erroneous data entry?		

1B. Mode error

Actions are performed in one mode that were intended for another mode.

#	Review Checklist	Severity Rating	Comments
1B.1	When an unusual mode choice is selected, is the user alerted?		
1B.2	When a medication dose mode is selected, is clear feedback given about the units associated with the mode (e.g., mcg/kg/min or mcg/min)?		
1B.3	When an unusually high or low dose is selected, is the user provided with a warning and a usual range?		
1B.4	Are dose range warnings appropriate for patient populations (e.g., pediatric patients with low weights)?		
1B.5	Is the display designed to reduce the risk of selecting the wrong mode based on parallax issues (e.g., sufficient spacing, offsetting row coloring, clear grouping of what is on the same row)?		
1B.6	Is the same default mode used consistently throughout the interface (e.g., direct dose vs. weight dose, same units, same measurement system)?		
1B.7	Are test actions separated from production actions (e.g., test accounts used rather than test modes on patient records for patients currently being treated for testing new functionality)?		
1B.8	Are special modes (e.g., view only, demonstration, training) clearly displayed?		

1C. Data accuracy error

Displayed data are not accurate.

#	Review Checklist	Severity Rating	Comments
1C.1	Is information not truncated on the display (e.g., medication names and doses in pick list menu displays are accurate and complete and distinguishable from other items in the pick list)?		
1C.2	Does accurate information automatically display (e.g., without requiring an active refresh command by the user)?		
1C.3	Can inaccurate information be easily changed (e.g., allergies)?		
1C.4	When a medication is renewed and then the dose is changed before signing, is the correct information displayed?		
1C.5	Do changes in status (e.g., STAT to NOW) display accurately?		
1C.6	If a medication schedule is changed, does the quantity correctly update?		
1C.7	If a medication order is discontinued, is the information updated on all displays about the change?		
1C.8	Is truncation of numbers such that the numeric value entered is different than the one saved avoided (e.g., user types in 10000 and 100 is saved in the field since it is limited to 3 characters)?		
1C.9	If the precision of an entered value is adjusted by the system, is this adjustment appropriate, and if so, is it shown to the user before the information is saved? Are precision modifications for special populations (e.g., morphine units for pediatric patients) taken into account?		
1C.10	Is automatic removal of outdated orders without alerting the user and allowing an override avoided?		

1C.11	Are dates checked to ensure that they are reasonable values for the situation, and if not, is the user alerted (e.g., entering the patient's birthdate for the current date would be reasonable for labor and delivery, but not for most clinical settings)?		
-------	--	--	--

1D. Data availability error

Decisions are based on incomplete information because related information requires additional navigation, access to another provider's note, taking actions to update the status, or is not updated within a reasonable time.

#	Review Checklist	Severity Rating	Comments
1D.1	Is all the information needed to understand regular doses, complex doses, and non-standard doses easily accessible and is it easy for the user to see without additional navigation (e.g., additional clicks) any additional information? (e.g., do not use comment fields that have to be clicked on individually to read information about what the dosage should be on that day, such as "Taper Dose 80 mg day 1 and 2, 60 mg day 3 and 4, 40 mg day 5 and 6, 20 mg day 7 and 8")		
1D.2	Are complex doses displayed in ways that users can easily understand what is intended on a particular day without additional navigation beyond what is required for regular dose schedules?		
1D.3	Are the contents of unsigned notes clearly identified as being notes in progress, and accessible to designated users (e.g., avoid hiding unsigned notes from all but the user who initiated them)?		
1D.4	Is information accurately updated in one place efficiently and accurately updated in other areas or in integrated software systems (e.g., avoid having a discharge summary display an outdated medication dose)?		

1E. Interpretation error

Differences in measurement systems, conventions and terms contribute to erroneous assumptions about the meaning of information.

#	Review Checklist	Severity Rating	Comments
1E.1	Is the same measurement system used consistently?		
1E.2	Are the same measurement units used consistently?		
1E.3	Are accepted domain conventions used consistently (e.g., axes of a pediatric growth chart)?		
1E.4	Does the system provide support for generic or brand names of medications to be used and displayed consistently?		
1E.5	Does the system provide support for organizations to use standardized terminology which is organized consistently (e.g., a clinical reminder building template with a consistent structure like What/When/Who is provided for organizations to optionally employ)?		
1E.6	Are negative structures avoided (e.g., "Do you not want to quit")?		
1E.7	Are areas of the interface that are intended for use by only certain categories of users displayed only for those users and either not displayed or displayed as grayed out/unavailable for other users?		

1F. Recall error

Decisions are based on incorrect assumptions because appropriate actions require users to remember information rather than recognize it.

#	Review Checklist	Severity Rating	Comments
1F.1	Does the interface enable recognition of information, rather than requiring users to remember information (e.g., one-time medication orders linked with a scheduled order should not require providers to remember the dose and type it in)?		
1F.2	Are frequently used and/or evidence-based options clearly distinguished from other options?		
1F.3	Is auto-fill avoided where there is more than one auto-fill option that closely matches in order to reduce the risk of picking the wrong medication?		
1F.4	Is identical information from another part of the system automatically filled in to avoid errors in redundant data entry?		
1F.5	Are STAT medications easy to recognize from summary displays?		
1F.6	When creating a new patient record, are predictable errors from workarounds that are based on manipulating existing records that could result in the destruction of patient data prohibited and/or users alerted to the risks?		

1G. Feedback error

Decisions are based on insufficient information because lack of system feedback about automated actions makes it difficult to identify when the actions are not appropriate for the context.

#	Review Checklist	Severity Rating	Comments
1G.1	Are user-entered fields (e.g., medication types, doses, and routes, test and procedure orders, diagnoses, dates, etc.) changed by the system and, if so, is normalization of field values appropriate, and does the user have the opportunity to see the changes before the information is saved (e.g., do not automatically change partial tablets to full tablets without alerting the user)?		
1G.2	Are changes to displays easy to detect and track?		
1G.3	Are automated merges of patient record data (e.g., automated algorithms that identify and merge multiple similar records based upon similar field entries) minimized? If used, are they done with sufficient feedback, active confirmation from the user, and the ability to track what actions were taken?		

1H. Data integrity error

Decisions are based on stored data that are corrupted or deleted.

#	Review Checklist	Severity Rating	Comments
1H.1	Do view-only software modes avoid changing stored data?		
1H.2	Is it possible to know who is blocking access to a data element or record when multiple users are accessing the same record simultaneously?		
1H.3	Are predictable scenarios where corrupted backup data would permanently destroy patients' records, and possibly all data for the entire organization, protected against through design measures and alerts?		
1H.4	Can activities performed during down times be easily entered into the record?		
1H.5	Can critical information (e.g., important pathology reports, images, or information about ineffective HIV anti-retroviral medications) be proactively tagged to avoid deletion during purges (due to policies implemented to reduce storage overhead)?		
1H.6	Can inappropriate clinical reminders and alerts be easily removed (e.g., clicking a "does not apply" option that is always last on the interface)?		
1H.7	When a system goes down unexpectedly and is restarted, do modifications for special populations avoid getting defaulted to standard settings (e.g., are alert settings for standard doses for pediatric patients maintained after the system is restarted, rather than defaulted to alert settings for adult populations)?		

2. Visibility of System Status

The system should always keep the user informed about what is going on, through appropriate feedback within reasonable time.

#	Review Checklist	Severity Rating	Comments
2.1	Does every display begin with a title or header that describes screen contents?		
2.2	Is there a consistent icon design scheme and stylistic treatment across the system?		
2.3	In multipage data entry screens, is each page labeled to show its relation to others?		
2.4	If pop-up windows are used to display error messages, do they allow the user to see the field in error?		
2.5	Is there some form of system feedback for every operator action?		
2.6	After the user completes an action (or group of actions), does the feedback indicate that the next group of actions can be started?		
2.7	Is there visual feedback in menus or dialog boxes about which choices are selectable?		
2.8	Is there visual feedback in menus or dialog boxes about which choice the cursor is on now?		
2.9	If multiple options can be selected in a menu or dialog box, is there visual feedback about which options are already selected?		
2.10	Is there visual feedback when objects are selected or moved?		
2.11	Is the current status of an icon clearly indicated?		
2.12	Do Graphic User Interface (GUI) menus make obvious which item has been selected?		
2.13	Do GUI menus make obvious whether deselection is possible?		
2.14	If users must navigate between multiple screens, does the system use context labels, menu maps, and place markers as navigational aids?		

3. Match between System and the Real World

The system should follow the user's language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

#	Review Checklist	Severity Rating	Comments
3.1	Are menu choices ordered in the most logical way, given the user, the item names, and the task variables?		
3.2	Do related items appear on the same display?		
3.3	Do the selected colors correspond to common expectations about color codes?		
3.4	When prompts imply a necessary action, are the words in the message consistent with that action?		
3.5	Do keystroke references in prompts match actual key names?		
3.6	On data entry screens, are tasks described in terminology familiar to users?		
3.7	Are field-level prompts provided for data entry screens?		
3.8	For question and answer interfaces, are questions stated in clear, simple language?		
3.9	Does the system automatically enter leading or trailing spaces to align decimal points?		
3.10	Does the system automatically enter commas in numeric values greater than 9999?		
3.11	Do GUI menus offer activation: that is, make obvious how to say "now do it"?		

4. User Control and Freedom

Users should be free to select and sequence tasks (when appropriate), rather than having the system do this for them. Users often choose system functions by mistake and will need a clearly marked “emergency exit” to leave the unwanted state without having to go through an extended dialogue. Users should make their own decisions (with clear information) regarding the costs of exiting current work. The system should support undo and redo.

#	Review Checklist	Severity Rating	Comments
4.1	In systems that use overlapping windows, is it easy for users to rearrange windows on the screen?		
4.2	In systems that use overlapping windows, is it easy for users to switch between windows?		
4.3	Are users prompted to confirm commands that have drastic, destructive consequences?		
4.4	Is there an "undo" function at the level of a single action, a data entry, and a complete group of actions?		
4.5	Can users cancel out of operations in progress?		
4.6	If users can reduce data entry time by copying and pasting existing data, is there a way to track what was copied and what was modified in order to make it easier to detect erroneously copied information (e.g., light background color behind copied text)?		
4.7	If menu lists are long (more than seven items), can users select an item either by moving the cursor or by typing a mnemonic code?		
4.8	If the system uses a pointing device, do users have the option of either clicking on menu items or using a keyboard shortcut?		
4.9	Are menus broad (many items on a menu) rather than deep (many menu levels)?		
4.10	If the system has multipage data entry screens, can users move backward and forward among the pages in the set?		
4.11	If the system uses a question and answer interface, can users go back to previous questions or skip forward to later questions?		

4.12	If users can set their own system, session, file and screen defaults, are there protections against predictable use errors for likely defaults?		
------	---	--	--

5. Consistency and Standards

Users should not have to wonder whether different words, situations or actions mean the same thing. Follow platform conventions.

#	Review Checklist	Severity Rating	Comments
5.1	Has a heavy use of all uppercase letters on a screen been avoided?		
5.2	Do abbreviations not include punctuation?		
5.3	Are integers right-justified and real numbers decimal-aligned?		
5.4	Are icons easy to interpret and is there a redundant way to interpret them (e.g., labels, mouseover labels)?		
5.5	Are there no more than twelve to twenty icon types?		
5.6	Are there salient visual cues to identify the active window?		
5.7	Does the menu structure match the task structure?		
5.8	If "exit" (or its equivalent, such as "quit" or "close") is a menu choice, does it always appear at the bottom of the list?		
5.9	Are menu titles either centered or left-justified?		
5.10	Are field labels consistent from one data entry screen to another?		
5.11	Are high-value, high-chroma colors used to attract attention?		

6. Help Users Recognize, Diagnose and Recover From Errors

Error messages should be expressed in plain language (NO CODES).

#	Review Checklist	Severity Rating	Comments
6.1	Are prompts brief and unambiguous?		
6.2	Are error messages grammatically correct?		
6.3	Do error messages avoid the use of exclamation points?		
6.4	Do error messages avoid the use of violent or hostile words?		
6.5	Do all error messages in the system use consistent grammatical style, form, terminology and abbreviations?		
6.6	Do messages place users in control of the system?		
6.7	Do error messages inform the user of the error's severity?		
6.8	Do error messages suggest the cause of problem?		
6.9	Do error messages provide sufficiently detailed information that makes it easy to do the intended behavior?		
6.10	Do error messages indicate what action the user needs to take to correct the error?		

7. Error Prevention

Even better than good error messages is a careful design that prevents a problem from occurring in the first place.

#	Review Checklist	Severity Rating	Comments
7.1	Is the menu choice name on a higher-level menu used as the menu title of the lower-level menu?		
7.2	Has the use of qualifier keys (e.g., shift, control, command, alt) been minimized?		
7.3	If the system uses qualifier keys, are they used consistently throughout the system?		
7.4	Does the system prevent users from making errors whenever possible?		
7.5	Does the system warn users if they are about to make a potentially serious error?		
7.6	Do data entry screens and dialog boxes indicate the number of character spaces available in a field?		
7.7	Do fields in data entry screens and dialog boxes contain default values when appropriate?		

8. Recognition Rather Than Recall

Make objects, actions and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

#	Review Checklist	Severity Rating	Comments
8.1	Does the data display start in the upper-left corner of the screen?		
8.2	Are all data that a user needs on display at each step in a transaction sequence?		
8.3	Have prompts been formatted using white space, justification and visual cues for easy scanning?		
8.4	Have zones been separated by spaces, lines, color, letters, bold titles, rules lines, or shaded areas?		
8.5	Are field labels close to fields, but separated by at least one space?		
8.6	Are optional data entry fields clearly marked?		
8.7	Are meaningful groups clearly demarcated (e.g., borders used)?		
8.8	Is color coding consistent throughout the system?		
8.9	Is color used in conjunction with another redundant cue?		
8.10	Is the first word of each menu choice the most important?		
8.11	Are inactive menu items grayed or omitted?		
8.12	Are there menu selection defaults?		
8.13	Do data entry screens and dialog boxes indicate when fields are optional?		
8.14	On data entry screens and dialog boxes, are dependent fields displayed only when necessary?		

9. Flexibility and Minimalist Design

Accelerators-unseen by the novice user-may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions. Provide alternative means of access and operation for users who differ from the “average” user (e.g., physical or cognitive ability, culture, language, etc.)

#	Review Checklist	Severity Rating	Comments
9.1	If the system supports both novice and expert users, are multiple levels of error message detail available?		
9.2	Does the system allow novice users to enter the simplest, most common form of each command, and allow expert users to add parameters?		
9.3	Does the system provide function keys for high-frequency commands?		
9.4	For data entry screens with many fields or in which source documents may be incomplete, can users save a partially filled screen?		
9.5	If menu lists are short (seven items or fewer), can users select an item by moving the cursor?		
9.6	If the system uses a pointing device, do users have the option of either clicking on fields or using a keyboard shortcut?		
9.7	Does the system offer "find next" and "find previous" shortcuts for database searches?		
9.8	In dialog boxes, do users have the option of either clicking directly on a dialog box option or using a keyboard shortcut?		
9.9	Can expert users bypass nested dialog boxes with either type-ahead, user-defined macros, or keyboard shortcuts?		

10. Aesthetic and Minimalist Design

Dialogues should not contain information that is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

#	Review Checklist	Severity Rating	Comments
10.1	Is only (and all) information essential to decision making displayed on the screen?		
10.2	Are all icons in a set visually and conceptually distinct?		
10.3	Have large objects, bold lines and simple areas been used to distinguish icons?		
10.4	Does each icon stand out from its background?		
10.5	If the system uses a standard GUI where menu sequence has already been specified, do menus adhere to the specification whenever possible?		
10.6	Are meaningful groups of items separated (e.g., by white space)?		
10.7	Does each data entry screen have a short, simple, clear, distinctive title?		
10.8	Are field labels brief, familiar and descriptive?		
10.9	Are prompts expressed in the affirmative, and do they use the active voice?		
10.10	Is each lower-level menu choice associated with only one higher-level menu?		
10.11	Are menu titles brief, yet long enough to communicate?		
10.12	Are there pop-up or pull-down menus within data entry fields that have many, but well-defined, entry options?		

11. Help and Documentation

Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

#	Review Checklist	Severity Rating	Comments
11.1	If menu choices are ambiguous, does the system provide additional explanatory information when an item is selected?		
11.2	Are data entry screens and dialog boxes supported by navigation and completion instructions?		
11.3	Are there memory aids for commands, either through on-line quick reference or prompting?		
11.4	Is the help function visible (e.g., by a key labeled HELP or a special menu)?		
11.5	Is the help system interface (navigation, presentation and conversation) consistent with the navigation, presentation and conversation interfaces of the application it supports?		
11.6	Navigation: Is information easy to find?		
11.7	Presentation: Is the visual layout well designed?		
11.8	Conversation: Is the information accurate, complete and understandable?		
11.9	Is there context-sensitive help?		
11.10	Is it easy to access and return from the help system?		
11.11	Can users resume work where they left off after accessing help?		

12. Pleasurable and Respectful Interaction with the User

The user's interactions with the system should enhance the quality of her or his work-life. The user should be treated with respect. The design should be aesthetically pleasing- with artistic as well as functional value.

#	Review Checklist	Severity Rating	Comments
12.1	Is each individual icon a harmonious member of a family of icons?		
12.2	Has excessive detail in icon design been avoided?		
12.3	Have flashing text and icons been avoided?		
12.4	Has color been used with discretion?		
12.5	Has color been used specifically to draw attention, communicate organization, indicate status changes, and establish relationships?		
12.6	Are typing requirements minimal for question and answer interfaces?		
12.7	If the system supports multiple input devices, has hand and eye movement between input devices been minimized?		

13. Privacy

The system should help the user to protect personal or private information belonging to the user or his/her patients.

#	Review Checklist	Severity Rating	Comments
13.1	Are protected areas inaccessible under normal circumstances?		
13.2	Can protected or confidential areas be accessed when necessary by following relevant security protocols (e.g., password protection)?		

Appendix C: Scenario 1: Ambulatory Care – Chronic Complex Patient; Mid-Level Provider

Includes NIST Test Procedures (V1.1):

§170.302.a Drug-Drug, Drug Allergy, Formulary Checks

§170.302.c Maintain Up-to-Date Problem List

§170.302.d Maintain Active Medication List

§170.302.h Incorporate Lab Test Results

§170.304.h Clinical Summaries

§170.306.a Computerized Provider Order Entry

§170.306.g Reportable Lab Results

§170.302.g Smoking Status

A mid-level provider (Nurse Practitioner or Physician Assistant) is providing care.

The patient is a 45-year-old African-American female living in an urban center. She has hypertension (HTN), obesity, mild congestive heart failure (CHF), type 2 diabetes, elevated cholesterol (LDL), and asthma. She started smoking when she was 17 years old and is actively trying to quit.

The patient comes in for a recheck of her weight and diabetes.

At the end of the prior visit, the plan was to get a fasting blood sugar (BS) in the office, collect a blood specimen to do a lipid panel & HbA1c (sent out), have an intake nurse get vital signs, including weight, do a diabetic foot exam, and talk with the patient to get an intervening history.

Task 1: Review active patient medications and medication history to identify if prescription refills are needed and ensure that discontinued medications do not need to be renewed

Removing a medication patch can be difficult since there is usually no order associated with it, therefore it relies upon **recall**.

The patient is currently on these active medications:

Diabeta (glyburide) 2.5 mg tablet by mouth every morning

Lipitor (atorvastatin calcium) 10 mg tablet by mouth daily

Lasix (furosemide) 20 mg tablet by mouth 2 times per day

Klor-Con (potassium chloride) 10 mEq tablet by mouth 2 times per day

Erythromycin (erythromycin ethylsuccinate) oral suspension 400 mg by mouth every 6 hours

Nicoderm patch as needed

Albuterol inhaler (albuterol sulfate) aerosol 2 puffs by oral inhalation every 4 hours prn

Lasix is sometimes administered earlier than ordered for patient comfort sleeping at night, which **deviates from routine workflow**.

Ordering, administering, and updating PRN medications can be challenging, because they are often treated as a different **mode**, and therefore viewed separately.

Task 2: Review patient labs to determine if changes are needed for care plan.

Lab data reveal the need to start the patient on Coumadin 2.5 mg and increase the dose of Lipitor.

Task 3: Modify active medications

Increasing the dose of an existing medication can be difficult, because **editing an order** can sometimes be harder than writing a new one.

Increase dose of Lipitor to 20 mg tablet by mouth daily. [If drug interaction warning occurs, review warning and respond appropriately.]

Task 4: Order new medications

A **high false alarm rate** for drug-drug interactions is challenging: 705 drugs interact with Coumadin.

It has been decided that the patient needs to get Coumadin.

Order Coumadin 2.5 mg with first dose now and next and subsequent doses in the morning. (If drug interaction warning occurs, review warning and respond appropriately.)

Ordering “first dose now” followed by a scheduled regular administration time can be challenging, particularly if **recall** is required for the dose amount, information about a medication order is **truncated**, the **dosing mode** varies, and **feedback** about what was ordered can be difficult to **interpret**.

During the patient visit, it is learned that the patient has had to use the inhaler more frequently due to the high pollen in the air. It is decided that the patient should use steroids to deal with asthma concerns that likely have escalated to bronchitis.

Order a “taper dose” of oral methylprednisolone. (If drug interaction warning occurs, review warning and respond appropriately.)

Complex doses can be challenging to order, administer, and **interpret**. Taper doses, where the dose is reduced over time, are particularly challenging when **data are not available** without additional navigation.

Task 5: Update problem list

The patient reveals that she lost her job, and has started abusing drugs. Add “substance abuse” to problem list.

Sensitive diagnoses are sometimes handled differently than other diagnoses by providers, particularly if the **patient views the record**.

Task 6: Order a consult

In order to address the drug abuse, request a consult with a social worker.

Task 7: Document progress note

Incorporate available test results in the documentation.

Individual strategies for incorporating test results in documentation can sometimes have unintended consequences, particularly if **outdated data** are copied and pasted from previous notes.

Appendix D: Scenario 2: Inpatient Care – Cardiac Patient; Physician

Includes NIST Test Procedures (V1.1):

§170.304.h Clinical Summaries *

§170.306.a Computerized Provider Order Entry

§170.302.q Automatic Log-off

§170.304.b Electronic Prescribing *

§170.304.j Calculate & Submit Quality Measures

§170.306.e Electronic Copy of Discharge Information

§170.306.h Advance Directives

Using the same patient across scenarios makes the evaluation more efficient. On the other hand, more complex patient populations, such as **pediatric patients**, can be tested with different patients.

A physician is providing care.

The patient is a 45-year-old African-American female living in an urban center. She has hypertension (HTN), obesity, Type 2 Diabetes, elevated cholesterol (LDL), and asthma. She started smoking when she was 17 years old and is actively trying to quit.

The patient is brought to the Emergency Room by Emergency Medical Services. She called 911 when she was having chest pain.

Upon admission, the patient reports that she is currently on these active medications:

Diabeta (glyburide) 2.5 mg tablet by mouth every morning

Lipitor (atorvastatin calcium) 20 mg tablet by mouth daily

Lasix (furosemide) 20 mg tablet by mouth 2 times per day

Klor-Con (potassium chloride) 10 mEq tablet by mouth 2 times per day

Erythromycin (erythromycin ethylsuccinate) oral suspension 400 mg by mouth every 6 hours

Coumadin 2.5 mg by mouth daily

Nicoderm patch as needed

Verbal orders sometimes need to be performed for time-critical situations, even when it deviates from policy. Documenting verbal orders and medication administration can be challenging, particularly if a different provider is documenting than did the order.

Albuterol inhaler (albuterol sulfate) aerosol 2 puffs by oral inhalation every 4 hours prn

Methadone – the patient is not sure of the dose

Task 1: Document nitroglycerin under the tongue given in the ER by a nurse per verbal order 3 hours after admission (Note that no order has been made by the physician or verified by the pharmacist for this medication).

Tasks that are usually **performed by others** can be challenging to do the first time.

Task 2: Enter vital signs [Blood pressure (BP) 172/95, heart rate 90]

Task 3: Order labs

Order labs to determine if patient is having a heart attack. [Up to user to determine which labs. If user requests which labs to order, say Creatine Kinase – Total and MB, Troponin I/T, Electrocardiogram (EKG).]

Task 4: Modify active medications

Change Lasix from PO to IV at the same dose.

Changing the route can be challenging, particularly when users typically use default order sets that do not require **recall**.

Task 5: Review labs.

[They indicate patient is having a mild heart attack. BP remains elevated. Patient will be admitted.]

Do documentation for handoff from ER to coronary care unit or inpatient unit [Could be progress note, could be dedicated handoff documentation].

Documentation for **patient handoffs** can be challenging, particularly when there is high variability in policies and/or redundant data entry.

Task 6: Document DNR status

Task 7: Determine status of STAT medication that was ordered a few hours before

Having accurate data pulls from the database when multiple records are open can be challenging, including **patient identifiers**.

In the middle of the documentation, interrupted. Must leave software open and open a new copy of the electronic health record to answer a question (from surrogate over the phone) about why a pediatric patient has not yet received a STAT chemotherapy medication. [Requires going to screens that show that the order was done correctly by the physician, but that the pharmacist has not yet verified the medication.]

Task 8: Return to finish the documentation for the handoff

Interruptions increase the risk of forgetting information or to complete tasks.

Task 9: Day 2. Review morning labs and vital signs.

[They show that the labs have stabilized and vitals including blood pressure have returned to normal. The patient can be discharged.]

Task 10: Transfer all inpatient medications to outpatient medications.

Batch processing of medications from the inpatient to outpatient setting can be challenging, particularly if the system **feedback** does not include automated changes to orders, such as from partial tablets to full tablets.

Task 11: Print discharge summary

Verifying that information is accurate can be challenging for complex patients, particularly when free text comment fields are used to **communicate** between providers as well as with patients.

Task 12: Print a report for a hospital administrator that shows how the organization is doing on the quality measure about how soon nitroglycerine is given to patients with chest pain in the emergency department.

Appendix E: Scenario 3 Critical Care – Cardiac Patient; Nurse

Includes NIST Test Procedures (V1.1):

- §170.306.h Advance Directives

A Registered Nurse is providing care in the Medical Intensive Care Unit.

The patient is a 68-year old African American female living in an urban center. She has hypertension (HTN), obesity, type 2 diabetes, elevated cholesterol (LDL), and asthma. She started smoking at 17 years old and is actively trying to quit.

The patient was admitted to the Emergency Room by Emergency Medical Services. She called 911 when she was having crushing chest pain, sweating and significant difficulty breathing. Nitroglycerin under the tongue was given in the ER. Her initial vital signs in the ER were blood pressure 168/95 and heart rate 112, and lab tests (CK-MB, troponen, LDH, and EKG) resulted in immediate medical intensive care placement, intubation and placement on a ventilator, a catheter placement, and a cardiology consult.

Task 1: Document change in DNR status: Remove DNR

In the ER, a request was made for DNR status, which was documented. Now the patient's family has arrived, and brought along a Living Will, which specifies that the DNR status is incorrect. In fact, the patient wishes to be resuscitated in all circumstances.

Task 2: Document intake and outtake record

Document the intake and outtake record for the last 12 hours. Over the last six hours, the patient has had 1000 mL of D5 W infusing IV at 30 mL/hour.

In the ER prior to arriving in the ICU the nurse had previously documented the following amounts voided: 400 cc at 7:00 am; 100cc at 10:00 am; 200cc at 12 noon; and 150 cc at 2:00 pm. Now the nurse needs to add voiding 400cc at 6:00 pm.

Task 3: Document medication administration

The patient is indicating continued chest pain. The physician order calls for Morphine Sulfate 2-4 mg IV q2-4 PRN. Document giving the patient three doses of 3mg, each three hours apart.

Appendix F: Recruitment Screener Adapted from NISTIR 7742

This is provided as a sample only. You will need to develop the specifics based upon your requirements.

Hello, my name is *[insert name]*, calling from *[Insert name of recruiting firm]*. We are recruiting individuals to participate in a usability study for an electronic health record. We would like to ask you a few questions to see if you qualify and if you would like to participate. This should only take a few minutes of your time. This is strictly for research purposes. If you are interested and qualify for the study, you will be paid to participate. Can I ask you a few questions?

Customize this by dropping or adding questions so that it reflects your EHR's primary audience.

1. *[If not obvious] Are you male or female? [Recruit a mix of participants]*
2. Have you participated in a focus group or usability test in the past xx months? *[If yes, Terminate]*
3. Do you, or does anyone in your home, work in marketing research, usability research, web design [...etc.]? *[If yes, Terminate]*
4. Do you, or does anyone in your home, have a commercial or research interest in an electronic health record software or consulting company? *[If yes, Terminate]*
5. Which of the following best describes your age? *[23 to 39; 40 to 59; 60 to 74; 75 and older]*
[Recruit Mix]
6. Do you require any assistive technologies to use a computer? *[If so, please describe]*

Professional Demographics *Customize this list to reflect the primary user groups expected for your EHR.*

7. What is your current position and title? (Must be healthcare provider)
 - RN: Specialty _____
 - Physician: Specialty _____
 - Resident: Specialty _____
 - Administrative Staff
 - Other *[Terminate]*
8. How long have you worked as? _____ *[Record]*

9. Describe your work location (or affiliation) and environment? (Recruit according to the intended users of the application) *[e.g., private practice, health system, government clinic, etc.]*
10. Which of the following describes your highest level of education? *[e.g., high school graduate/GED, some college, college graduate (RN, BSN), postgraduate (MD/PhD), other (explain)]*

Computer Expertise *Customize this to reflect what you know about your EHR's audience.*

11. Besides reading email, what professional activities do you do on the computer? *[e.g., access EHR, research; reading news; shopping/banking; digital pictures; programming/word processing, etc.] [If no computer use at all, Terminate]*
12. About how many hours per week do you spend on the computer? *[Recruit according to the demographics of the intended users, e.g., 0 to 10, 11 to 25, 26+ hours per week]*
13. What computer platform do you usually use? *[e.g., Mac, Windows, etc.]*
14. What Internet browser(s) do you usually use? *[e.g., Firefox, IE, AOL, etc.]*
15. In the last month, how often have you used an electronic health record? *[Record]*
16. How many years have you used an electronic health record? *[Record]*
17. How many EHRs do you use or are you familiar with?
18. How does your work environment record patient records? *[Recruit according to the demographics of the intended users]*
 - On paper
 - Some paper, some electronic
 - All electronic

Contact Information *If the person matches your qualifications, ask*

Those are all the questions I have for you. Your background matches the people we're looking for. *[If you are paying participants or offering some form of compensation, mention here]* For your participation, you will be paid *[amount]*.

Would you be able to participate on *[date, time]*? *[If so collect contact information]*

May I get your contact information?

- Name of participant:

- Address:
- City, State, Zip:
- Daytime phone number:
- Evening phone number:
- Alternate [cell] phone number:
- Email address:

Before your session starts, we will ask you to sign a release form allowing us to videotape your session. The videotape will only be used internally for further study if needed. Will you consent to be videotaped?

This study will take place at *[location]*. I will confirm your appointment a couple of days before your session and provide you with directions to our office. If you need any assistance, please let us know.

Appendix G: Example Tester's Guide

Only three tasks are presented here for illustration.

EHRUT Usability Test

Tester's Guide

Tester (Lead) Name _____

Data Logger Name _____

Date of session _____ Time _____

Participant # _____

Location _____

Prior to testing

- Confirm schedule with Participants
- Ensure EHRUT lab environment is running properly
- Ensure lab and data recording equipment is running properly

Prior to each participant:

- Reset application
- Start session recordings with *tool*

Prior to each task:

- Reset application to starting point for next task

After each participant:

- End session recordings with *tool*

After all testing

- Back up all video and data files

All information to be read to the participants in the Tasks is underlined; tester's notes are in *italics*.

Orientation (X minutes)

Thank you for participating in this study. Our session today will last **XX minutes**. During that time I'll ask you to interact with an electronic health record system.

I will ask you to complete some tasks using this system. We are testing the system, not you or your abilities. Our goal in this testing is to understand how easy (or how difficult) this system is to use, what steps you use to accomplish the goals, and your subjective impressions. Please complete the tasks as quickly as you can without sacrificing accuracy. Please only complete the assigned tasks. While we would like to know when you are having difficulty with the system, I may not be able to address your questions immediately during the task. We will discuss any issues at length following the procedure. Should you have extreme difficulty in completing a subtask, I may move us on without finishing that portion. Please save your detailed comments until the end of the session as a whole when we can discuss freely.

The product you will be using today is *[describe the state of the application, i.e., production version, early prototype, etc., taking care not to include information that the participant does not need for the test or that might prime them in unexpected ways]*. Some of the data may not make sense as it is placeholder data.

We are recording the session today by *[describe specifics of the recording methods implemented by the system administrator and follow consent procedures]*. All of the information that you provide will be kept confidential and your name will not be associated with your comments at any time. Recording the session allows me to focus more on talking with you and less on taking notes because I can review the tape if necessary. My colleague is in another room watching the video and audio projection of this session helping me to take notes.

Do you have any questions or concerns?

Warming up: Preliminary Questions (X minutes)

What is your job title / appointment?

How long have you been working as X? How long have you been at this hospital/clinic?

What are some of your main responsibilities?

Tell me about your experience with electronic health records.

How long have you been using EHRs?

What do you like about them?

What do you dislike about them?

Task 1: Patient Summary Screen (XXX Seconds)

Take the participant to the starting point for the task. Read the task aloud, and then reveal a task card with the task on it. Begin timer.

Before going into the exam room, review the Patient's chief complaint, history and vitals by finding this information.

Record Success:

- Completed according to proper steps
- Completed with difficulty or help. Describe below
- Not completed

Comments:

Task Time: _____ Seconds

Optimal Path: Screen A → Screen B → Drop Down B¹ → "OK" Button → Screen X...

- Correct
- Minor Deviations / Cycle. Describe below
- Major Deviations. Describe below

Comments:

Observed Errors and Verbalizations:

Comments:

Rating:

Overall, how would you rate this task: _____

Show participant written scale: "Very Easy" (1) to "Very Difficult" (5)

Administrator / Note-taker Comments:

Task 2: Find Lab Results (XXX Seconds)

Take the participant to the starting point for the task. Read the task aloud, and then reveal a task card with the task on it. Begin timer.

On her last visit, you sent Patient to get a colonoscopy. Locate these results and review the notes from the specialist.

Record Success:

- Completed according to proper steps
- Completed with difficulty or help. Describe below
- Not completed

Comments:

Task Time: _____ Seconds

Optimal Path: Screen A → Screen B → Drop Down B¹ → “OK” Button → Screen X...

- Correct
- Minor Deviations / Cycles. Describe below
- Major Deviations. Describe below

Comments:

Observed Errors and Verbalizations:

Comments:

Rating:

Overall, how would you rate this task: _____

Show participant written scale: “Very Easy” (1) to “Very Difficult” (5)

Administrator / Note-taker Comments:

Task 3: Prescribe medication (XXX Seconds)

Take the participant to the starting point for the task. Ensure that this patient has a drug-drug and a drug-food allergy to the drug chosen. This will put force the participant to find other drugs and use other elements of the application. Read the task aloud, and then reveal a task card with the task on it. Begin timer.

After examining *Patient*, you have decided to put this patient on a statin – *drug name*. Check for any interactions and place an order for this medication.

Record Success:

- Completed according to proper steps
- Completed with difficulty or help. Describe below
- Not completed

Comments:

Task Time: _____ Seconds

Optimal Path: *Screen A → Screen B → Drop Down B¹ → “OK” Button → Screen X...*

- Correct
- Minor Deviations / Cycles. Describe below
- Major Deviations. Describe below

Comments:

Observed Errors and Verbalizations:

Comments:

Rating:

Overall, how would you rate this task: _____

Show participant written scale: “Very Easy” (1) to “Very Difficult” (5)

Administrator / Note-taker Comments:

Final Questions (X Minutes)

What was your overall impression of this system?

What aspects of the system did you like most?

What aspects of the system did you like least?

Were there any features that you were surprised to see?

What features did you expect to encounter but did not see? That is, is there anything that is missing in this application?

Compare this system to other systems you have used.

Would you recommend this system to your colleagues?

Administer the Usability Ratings

We have one final task for you. Could you please complete the following 10 questions about your experience with this application? (*Appendix I.*)

Thank participant once completed.

Appendix H: Informed Consent and Non-Disclosure Forms Adapted from NISTIR 7742

These are sample forms. The non-disclosure agreement is discretionary. Other examples may be found at www.usability.gov.

Informed Consent

[Test Company] would like to thank you for participating in this study. The purpose of this study is to evaluate an electronic health records system. If you decide to participate, you will be asked to perform several tasks using the prototype and give your feedback. The study will last about *[xxx]* minutes. At the conclusion of the test, you will be compensated for your time.

Agreement

I understand and agree that as a voluntary participant in the present study conducted by *[Test Company]* I am free to withdraw consent or discontinue participation at any time. I understand and agree to participate in the study conducted and videotaped by *[Test Company]*.

I understand and consent to the use and release of the videotape by *[Test Company]*. I understand that the information and videotape is for research purposes only and that my name and image will not be used for any purpose other than research. I relinquish any rights to the videotape and understand the videotape may be copied and used by *[Test Company]* without further permission.

I understand and agree that the purpose of this study is to make software applications more useful and usable in the future.

I understand and agree that the data collected from this study may be shared outside of *[Test Company]* and *[Test Company]* client. I understand and agree that data confidentiality is assured, because only de-identified data – i.e., identification numbers not names – will be used in analysis and reporting of the results.

I agree to immediately raise any concerns or areas of discomfort with the study administrator. I understand that I can leave at any time.

Please check one of the following:

YES, I have read the above statement and agree to be a participant.

NO, I choose not to participate in this study.

Signature: _____ **Date:** _____

Printed Name: _____

Witness: _____ **Date:** _____

Printed Name & Affiliation: _____

Non-Disclosure Agreement

THIS AGREEMENT is entered into as of _____, 201x, between _____ (“the Participant”) and the testing organization [*Insert Test Company Name*] located at [*Address*].

The Participant acknowledges his or her voluntary participation in today’s usability study may bring the Participant into possession of Confidential Information. The term "Confidential Information" means all technical and commercial information of a proprietary or confidential nature that is disclosed by [*Test Company*], or otherwise acquired by the Participant, in the course of today’s study.

By way of illustration, but not limitation, Confidential Information includes trade secrets, processes, formulae, data, know-how, products, designs, drawings, computer aided design files and other computer files, computer software, ideas, improvements, inventions, training methods and materials, marketing techniques, plans, strategies, budgets, financial information, or forecasts.

Any information the Participant acquires relating to this product during this study is confidential and proprietary to [*Test Company*] and is being disclosed solely for the purposes of the Participant’s participation in today’s usability study. By signing this form the Participant acknowledges that s/he will receive monetary compensation for feedback and will not disclose this confidential information obtained today to anyone else or any other organizations.

Participant’s printed name: _____

Signature: _____ **Date:** _____

Appendix I: Usability Ratings

One commonly used usability rating scale is Brooke's. In 1996, he published a "low-cost usability scale that can be used for global assessments of systems usability" known as the System Usability Scale or SUS.⁵⁴ Lewis and Sauro (2009) and others have elaborated on the SUS over the years. Computation of the SUS score can be found in [Brooke's paper](#), or in Tullis and Albert (2008).⁵⁵

	Strongly disagree					Strongly agree
1. I think that I would like to use this system frequently.	<input type="checkbox"/>					
	1	2	3	4	5	
2. I found the system unnecessarily complex.	<input type="checkbox"/>					
	1	2	3	4	5	
3. I thought the system was easy to use.	<input type="checkbox"/>					
	1	2	3	4	5	
4. I think that I would need the support of a technical person to be able to use this system.	<input type="checkbox"/>					
	1	2	3	4	5	
5. I found the various functions in this system were well integrated.	<input type="checkbox"/>					
	1	2	3	4	5	
6. I thought there was too much inconsistency in this system.	<input type="checkbox"/>					
	1	2	3	4	5	
7. I would imagine that most people would learn to use this system very quickly.	<input type="checkbox"/>					
	1	2	3	4	5	
8. I found the system very cumbersome to use.	<input type="checkbox"/>					
	1	2	3	4	5	
9. I felt very confident using the system.	<input type="checkbox"/>					
	1	2	3	4	5	
10. I needed to learn a lot of things before I could get going with this system.	<input type="checkbox"/>					
	1	2	3	4	5	

⁵⁴ Brooke, J.: SUS: A "quick and dirty" usability scale. In: Jordan, P. W., Thomas, B., Weerdmeester, B. A., McClelland (eds.) *Usability Evaluation in Industry* pp. 189--194. Taylor & Francis, London, UK (1996). SUS is copyrighted to Digital Equipment Corporation, 1986.

Lewis, J R & Sauro, J. (2009) "The Factor Structure Of The System Usability Scale." in *Proceedings of the Human Computer Interaction International Conference (HCI 2009), San Diego CA, USA*.

⁵⁵ Tullis, T. & Albert, W. (2008). *Measuring the user experience*. Morgan Kaufman.

Appendix J: Incentive Receipt and Acknowledgment Form

Acknowledgement of Receipt

I hereby acknowledge receipt of \$_____ for my participation in a research study run by *Test Company*.

Printed Name: _____

Address: _____

Signature: _____ Date: _____

Tester/Researcher - Printed Name: _____

Signature of Tester / Researcher: _____

Date: _____

Witness - Printed Name: _____

Witness Signature: _____

Date: _____

Glossary of Acronyms

Acronym	Definition
ABA	Applied Behavior Analysis
ADA	Americans with Disabilities Act
AR	Army Regulation
CDRH	Center for Devices and Radiological Health
CIF	Common Industry Format
DOD	Department of Defense
EHR	Electronic Health Record
EHRUT	EHR Application Under Test
EMR	Electronic Medical Record
EUP	EHR Usability Protocol
FAA	Federal Aviation Administration
FDA	Food and Drug Administration
GAO	General Accounting Office
GUI	Graphical User Interface
HFE	Human Factors Engineering
HFPMET	Human Factors Premarket Evaluation Team
HIMSS	Healthcare Information and Management Systems Society
HIT	Health Information Technologies
HSI	Human-System Integration
ISO	International Organization for Standardization
MANPRINT	US Army Manpower and Personnel Integration
MPT	Manpower, Personnel and Training
MU	Meaningful Use
NCC MERP	National Coordinating Council for Medication Error and Reporting and Prevention
NOIS	National Online Information Sharing
NRC	Nuclear Regulatory Commission
ODE	Office of Device Evaluation
OSHA	Occupational Safety and Health Act
PRM	Program Review Model
SUS	System Usability Scale
TAG	Technical Advisory Group

Further Reading

- Armijo, D., McDonnell, C., & Werner, K. (2009). *Electronic Health Record Usability: Evaluation and Use Case Framework* (AHRQ Publication No. 09(10)-0091-1-EF). Rockville, MD: Agency for Healthcare Research and Quality.
- Armijo, D., McDonnell, C., & Werner, K. (2009). *Electronic Health Record Usability: Interface Design Considerations* (AHRQ Publication No. 09(10)-0091-2-EF). Rockville, MD: Agency for Healthcare Research and Quality.
- Britto, M., Jimison, H., Munafo, J., Wissman, J., Rogers, M., & Hersh, W. (2009). Usability testing finds problems for novice users of pediatric portals. *Journal of the American Medical Informatics Association*, Sep–Oct, 660–669.
- Johnson, C., et al. (2005). A user-centered framework for redesigning healthcare interfaces. *Journal of Biomedical Informatics*, 38, 75-87.
- Institute of Medicine. (2012). *Health IT and Patient Safety: Building Safer Systems for Better Care*. Washington, DC: The National Academies Press.
- Kushniruk, A. & Patel, V. (2004). Cognitive and usability engineering methods for the evaluation of clinical information systems. *Journal of Biomedical Informatics*, 37, 56-76.
- McDonnell, C., Werner K., & Wendel L. (2010). *Electronic Health Record Usability: Vendor Practices and Perspectives* (AHRQ Publication No. 09(10)-0091-3-EF). Rockville, MD: Agency for Healthcare Research and Quality.
- National Institute of Standards and Technology. (2010). *Customized Common Industry Format Template for Electronic Health Record Usability Testing* (NISTIR 7742). Gaithersburg, MD: [://www.nist.gov/customcf/get_pdf.cfm?pub_id=907312](http://www.nist.gov/customcf/get_pdf.cfm?pub_id=907312) .
- National Institute of Standards and Technology. (2011). *Human Factors Guidance to Prevent Healthcare Disparities with the Adoption of EHRs* (NISTIR 7769). Gaithersburg, MD: [://www.nist.gov/customcf/get_pdf.cfm?pub_id=907991](http://www.nist.gov/customcf/get_pdf.cfm?pub_id=907991) .
- National Institute of Standards and Technology. (2010). *NIST Guide to the Processes Approach for Improving the Usability of Electronic Health Records* (NISTIR 7741). Gaithersburg, MD: [://www.nist.gov/manuscript-publication-search.cfm?pub_id=907313](http://www.nist.gov/manuscript-publication-search.cfm?pub_id=907313) .
- National Institute of Standards and Technology. (2010). *Usability in Health IT: Technical Strategy, Research, and Implementation* (NISTIR 7743). Gaithersburg, MD: [://www.nist.gov/customcf/get_pdf.cfm?pub_id=907316](http://www.nist.gov/customcf/get_pdf.cfm?pub_id=907316) .

- Phansalkar S., Edworthy J., Hellier E., Seger D., Schedlbauer A., Avery A., & Bates D. (2010). A review of human factors principles for the design and implementation of medication safety alerts in clinical information systems. *Journal of the American Medical Informatics Association*, Sep-Oct, 493-501.
- Ralston, J., Carrell, D., Reid, R., Anderson, M., Moran, M., & Hereford, J. (2007). Patient web services integrated with a shared medical record: patient use and satisfaction. *Journal of the American Medical Informatics Association*, Nov-Dec, 798-806. Correction in: Mar-Apr 2008, 265.
- Rosenbloom, S., Miller, R., Johnson, K., Elkin, P., & Brown, S. (2008). A model for evaluating interface terminologies. *Journal of the American Medical Informatics Association*, Jan-Feb, 65-76.
- Saitwal, H., Feng, X., Walji, M., Patel, V., & Zhang, J. (2010). Assessing performance of an electronic health record (EHR) using cognitive task analysis. *International Journal of Medical Informatics*, 79, 501-506.
- Saleem, J., Patterson, E., Militello, L., Anders, S., Falciglia, M., Wissman, J., Roth, E., and Asch, S. (2007). Impact of clinical reminder redesign on learnability, efficiency, usability, and workload for ambulatory clinic nurses. *Journal of the American Medical Informatics Association*, Sep-Oct, 632-640.
- Zhang, J. (2005). Human-centered computing in health information systems: Part I--Analysis and Design (Editorial). *Journal of Biomedical Informatics*, 38, 1-3. ([pdf file](#)).
- Zhang, J. (2005). Human-centered computing in health information systems: Part II--Evaluation (Editorial). *Journal of Biomedical Informatics*, 38, 173-175. ([pdf file](#)).
- Zhang, J., Johnson, T. R., Patel, V. L., Paige, D., & Kubose, T. (2003). Using usability heuristics to evaluate patient safety of medical devices. *Journal of Biomedical Informatics*, 36 (1-2), 23-30.
- Zhang, Z., Walji, M., Patel, V., Gimbel, R., & Zhang, J. (2009). Functional analysis of interfaces in U.S. military electronic health record system using UFuRT framework. *American Medical Informatics Association (AMIA) Conference Proceedings*.
- Zheng, K., Padman, R., Johnson, M., & Diamond, H. (2009). An Interface-driven analysis of user interactions with an electronic health record system. *Journal of the American Medical Informatics Association*, Mar-Apr, 228-237.