

NISTIR 7804

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

Robert M. Schumacher

User Centric Inc.

Emily S. Patterson

Ohio State University

Robert North

Human Centered Strategies, LLC

Jiajie Zhang

Univeristy of Texas - Houston

Svetlana Z. Lowry

National Institute of Standards and Technology

Matthew T. Quinn

National Institute of Standards and Technology

Mala Ramaiah

National Institute of Standards and Technology

U.S. Department of Commerce

Rebecca M. Blank, Acting Secretary

National Institute of Standards and Technology

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



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

Robert M. Schumacher

User Centric Inc.

Emily S. Patterson

Ohio State University

Robert North

Human Centered Strategies, LLC

Jiajie Zhang

University of Texas - Houston

Svetlana Z. Lowry

National Institute of Standards and Technology

Matthew T. Quinn

National Institute of Standards and Technology

Mala Ramaiah

National Institute of Standards and Technology

September 2011

U.S. Department of Commerce

Rebecca M. Blank, Acting Secretary

National Institute of Standards and Technology

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



Acknowledgments:

We gratefully acknowledge the contributions of Ron Kaye (FDA), Molly F. Story (FDA), Quynh Nhu Nguyen (FDA), Michael C. Gibbons (Johns Hopkins University), Patricia A. Abbott (Johns Hopkins University), 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), Mala Ramaiah (NIST) and Emile Morse (NIST) to the development and expert review of this report.

DRAFT

Table of Contents

<i>Table of Contents</i>	4
1 Executive Summary	7
2 Background: Impact of Usability on Electronic Health Record Adoption	9
3 Proposed EHR Usability Protocol (EUP)	11
3.1 Positioning	11
3.2 Overview of the Protocol Steps	11
3.3 What the EHR Usability Protocol Is Not	13
4 EUP Foundation: Research Findings Defining EHR Usability Issues and Their Impact on Medical Error	14
5 Expert Review of the User Interface and Evaluation Methods	16
5.1 A Model for Understanding EHR Patient Safety Risks	16
5.2 Definition of Usability and Associated Measures	21
5.3 Objective, Measurable Dimensions of Usability.....	24
5.4 Categorization of Medical Errors.....	25
Summary.....	26
6 Expert Review/Analysis of EHRs	27
6.1 Criteria and Best Principles for Selecting Expert Reviewers	27
6.2 Protocol for Conducting a Review	28
7 Summative: Protocol for Validation Testing EHR Applications	30
7.1 Purpose of Protocol and General Approach.....	30
7.2 Overview of Protocol Steps	32
8 Conclusion	49
Appendix A: Government Best Practices of Human-System Integration (HSI)	50
Appendix B: Form for Expert Review	56
Expert Review of EHR	57
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.....	58
1A. Patient identification error	58
Actions are performed for one patient or documented in one patient’s record that were intended for another patient.....	58
1B. Mode error	59
Actions are performed in one mode that were intended for another mode.....	59
1C. Data accuracy error	60
Displayed data are not accurate.....	60
1D. Data availability error	61

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. 61

1E. Interpretation error	62
Differences in measurement systems, conventions and terms contribute to erroneous assumptions about the meaning of information.....62	
1F. Recall error	63
Decisions are based on incorrect assumptions because appropriate actions require users to remember information rather than recognize it.63	
1G. Feedback error	64
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.64	
1H. Data integrity error	65
Decisions are based on stored data that are corrupted or deleted.65	
2. Visibility of System Status	66
The system should always keep user informed about what is going on, through appropriate feedback within reasonable time.66	
3. Match Between System and the Real World	67
The system should speak 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.67	
4. User Control and Freedom	69
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.....69	
5. Consistency and Standards	70
Users should not have to wonder whether different words, situations or actions mean the same thing. Follow platform conventions.....70	
6. Help Users Recognize, Diagnose and Recover From Errors	71
Error messages should be expressed in plain language (NO CODES).71	
7. Error Prevention	72
Even better than good error messages is a careful design that prevents a problem from occurring in the first place. 72	
8. Recognition Rather Than Recall	73
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.73	
9. Flexibility and Minimalist Design	74
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.)74	
10. Aesthetic and Minimalist Design	75

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.75

11. Help and Documentation.....	76
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.	76
12. Pleasurable and Respectful Interaction with the User.....	77
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.	77
13. Privacy.....	78
The system should help the user to protect personal or private information belonging to the user or his/her patients.	78
Appendix C: Scenario 1: Ambulatory Care – Chronic Complex Patient; Mid-Level Provider	79
Appendix D: Scenario 2: Inpatient Care – Cardiac Patient; Physician.....	82
Appendix E: Scenario 3 Critical Care – Cardiac Patient; Nurse.....	86
Appendix F: Recruitment Screener Adapted from NIST IR 7742.....	87
Appendix G: Example Tester’s Guide.....	90
Appendix H: Informed Consent and Non-Disclosure Forms Adapted from NIST IR 7742.....	97
Appendix I: Usability Ratings.....	100
Appendix J: Incentive Receipt and Acknowledgment Form.....	101
Appendix K: Form for Reporting Potential Use Errors.....	102
Appendix L: Form for Tracking Resolution of Potential Use Errors.....	103
Appendix M: Example of Data Summary Form.....	105
Glossary of Acronyms.....	106
Further Reading.....	107

1 Executive Summary

This document summarizes the rationale for an Electronic Health Record (EHR) Usability Protocol (EUP) that encompasses procedures for (1) expert evaluation of an EHR user interface from a clinical perspective and a human factors best practices perspective, and (2) validation studies of EHR user interfaces with representative user groups on realistic EHR tasks.

This document begins with a brief overview of the problem at hand: Why is EHR usability critical? There is a sentiment among clinical users that EHRs are harder to use than they need to be and can introduce “use error” that could have potential negative implications on patient care. This document centers on improving user performance of EHRs through system developer demonstration of application of human factors best practices and user-centered design principles.

Within this document there is a detailed description of research findings relating to the 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. As a backdrop, these findings motivate and guide the clinical aspects of the expert evaluation procedure.

Another intent of the EUP is to provide detailed systematic steps for conducting validation studies (that is, summative usability tests). The validation study procedure guides the reader to help insure that the application’s user interface is free from critical usability issues and supports error-free user interaction with the EHR. The sample forms for data collection and test scenarios provided in the appendices are guidance only; the development teams can choose other scenarios or modify these examples as necessary for their medical context.

This document presents a three-step process -- that incorporates both the evaluation and validation procedures -- for design evaluation and human user performance testing for an EHR, see Figure 1. This process is focused on increasing safe use of the EHR and increasing ease of use of the EHR by users. The steps are as follows:

1. Usability/Human Factors Analysis of the application during EHR user interface development,
2. Expert Review/Analysis of the EHR user interface after it is designed/developed, and
3. Testing of the EHR user interface with users.

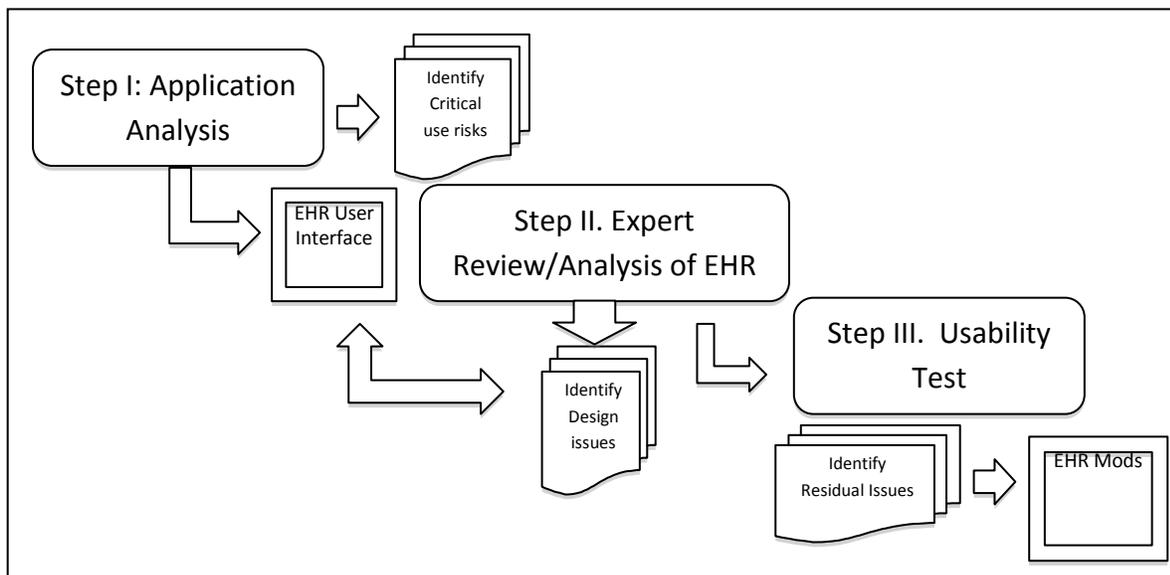


Figure 1. Three-step process for design evaluation and human user performance testing for the EHR

Step One: During the design of an EHR, the development team incorporates the users, work settings and common workflow into the design. Two major goals for this step that should be documented to facilitate Steps Two and Three are: (a) a list of possible medical errors associated with the system usability, and (b) a working model of the design with the usability that pertains to potential safety risks.

Step Two: The Expert Review/Analysis of the EHR step compares the EHR’s user interface design to scientific design principles and standards, identifies possible risks for error and identifies the impact of the design of the EHR on efficiency. This review/analysis can be conducted by a combination of the vendor’s development team and/or by a dedicated team of clinical safety and usability experts. The goals of this step are: (a) to identify possible safety risks and (b) identify areas for improved efficiency.

Step Three: The Testing with Users Step examines the critical tasks identified in the previous steps with actual users. Performance is examined by recording objective data (measured times such as successful task completion, errors, corrected errors, failures to complete, etc.) and subjective data (what users identify). The goals of this step are to: (a) make sure the critical usability issues that may potentially impact safety are no longer present and (b) make sure there are no critical barriers to decrease efficiency. This is accomplished through vendor-evaluator team review meetings where vendor’s system development and evaluation teams examine and agree that the design has (a) decreased the potential for medical errors to desired levels and (b) increased overall use efficiency due to critical usability issues.

The balance of this document describes the overall usability protocol recommended (with examples provided in appendices) and summarizes research findings on the relationship of usability and patient safety applicable to EHRs. *It is our expectation that the potential for all of these 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.*

2 Background: Impact of Usability on Electronic Health Record Adoption

Experts have identified shortcomings in the usability of current EHR systems as a key barrier to adoption and meaningful use of these systems.

The Healthcare Information and Management Systems Society (HIMSS) Usability Task Force, in their white paper “Defining and Testing EMR Usability: Principles and Proposed Methods of EMR Usability Evaluation and Rating”,¹ identified the importance of improving EHR usability as follows:

We submit that usability is one of the major factors – possibly the most important factor – hindering widespread adoption of EMRs. Usability has a strong, often direct relationship with clinical productivity, error rate, user fatigue and user satisfaction—critical factors for EMR adoption. (p.2)

The President’s Council of Advisors on Science and Technology in December of 2010² framed the issue this way:

...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. (Emphasis added, p.10)

Poor usability of EHR applications has a substantial negative effect on clinical efficiency and data quality. One of the steps to improving the overall state of the usability of EHRs was the recent publications of NIST IRs 7741 and 7742 which provided guidance to the vendor community on user-centered design and reporting of usability testing results.

Examples of usability issues that have been reported by health care workers are below:

- Some EHR workflows do not match clinical processes create inefficiencies,
- Poorly designed EHR screens slow down the user and sometimes endanger patients,
- Large numbers of files containing historical patient information are difficult to search, navigate, read efficiently, and identify trends over time,

¹ Belden J., Grayson R., Barnes J. Defining and Testing EMR Usability: Principles and Proposed Methods of EMR Usability Evaluation and Rating. HIMSS. June 2009. Available at: http://www.himss.org/content/files/HIMSS_DefiningandTestingEMRUsability.pdf

² Report issued by the President’s Council of Advisors on Science and Technology “Realizing the Full Potential of Health Information Technology to Improve Healthcare for Americans: The Path Forward” December 2010.

- Warning and error messages are confusing and often conflicting
- Alert fatigue (both visual and audio) from too many messages leading to users ignoring potentially critical messages, and
- Frustration with what is perceived as excessive EHR user interaction (mouse clicks, cursor movements, keystrokes, etc.) during frequent tasks.

All these contribute to frustration and fatigue, and ultimately impact patient care, but what does it mean to have a 'usable' application? How can we make EHRs easier to use? The EUP emphasis should be on ensuring that necessary and sufficient usability validation and remediation has been conducted so that use error³ is minimized.

DRAFT

³ "Use error" is a term used very specifically to refer to user interface designs that will engender users to make errors of commission or omission. It is true that users do make errors, but many errors are due not to user error *per se* but due to designs that are flawed, e.g., poorly written messaging, misuse of color-coding conventions, omission of information, etc.

3 Proposed EHR Usability Protocol (EUP)

3.1 Positioning

The proposed EUP builds on best practices of current procedures and guidance from various government agencies for systematic application of human factors in the system development process. The Federal government has a history of successful integration of human factors into the design and development of systems, as summarized in *Appendix A*.

The EUP emphasis should be on ensuring that necessary and sufficient usability validation and remediation has been conducted so that (a) use error is minimized and (b) use efficiency is maximized. Thus, the EUP focuses on identifying and minimizing critical usability issues associated with less than optimal EHR interface design. The intent of the EUP is to validate, by way of systematic steps, that the application's user interface is free from critical usability issues and supports error-free user interaction with the EHR.

3.2 Overview of the Protocol Steps

The EUP consists of three major steps:

- (I) **Usability Analysis** led by the development team, which identifies the characteristics of the system's anticipated users, use environments, scenarios of use, and use related usability risks that may induce medical errors,
- (II) **Expert Review/Analysis**, an independent evaluation of the critical components of the user interface in the context of execution of various use case scenarios and usability principles, and
- (III) **User Testing**, involving creation of a test plan and then conducting a test that will assess usability for the given EHR application including use efficiency and presence of features that may induce potential medical error.

Step I: Usability Analysis. The first step of this three step process is to 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 identifies the critical use related risks per each use scenario. Included in this analysis are:

- **Who are the users?** Descriptions of the EHR users including their job classifications (physician, registered nurse, etc.), knowledge, training related to system use, and user characteristics
- **What is the EHR's work area like?** Ambient conditions of use including lighting, noise, vibration, distraction, workstation layout, etc.
- **What does the user do?** Description of the major actions that the user performs 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 represents best practice in capturing the hierarchy of steps during user interaction and is recommended for documenting use scenarios.

- **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 of use, 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 should include 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.

Step II: Expert Review/Analysis. The second step in the overall EUP process is an Expert Review/Analysis of the application's user interface. Usability/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 have identified potential critical usability issues in user interfaces of EHRs. These findings have resulted in user interface design review protocols for EHR user interfaces. These reviews are focused on identifying potential user interface design-induced use errors and barriers to efficient use 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 due to drug confusion. 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 produced by current research efforts funded by the Office of the National Coordinator for Health Information Technology. This program (known as SHARPC) is developing a Rapid Usability Assessment Protocol based on known user interface design best practices, e.g., consistency.⁴ Specifically, this method involves the use of three expert evaluators applying fourteen 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. The violations range

⁴ Zhang, J. (2011) Personal communication. Zhang J, Walji MF. TURF: A Unified Framework of EHR Usability. Journal of Biomedical Informatics, 2011

from ‘advisory’ (moderate concern) to ‘catastrophic’ (critical to safety or efficiency 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 and overall use efficiency to identify the most critical user interface problems that should be addressed through dialogue 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: User Testing. The final step in the EUP 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. A fuller description of the Summative Protocol is presented in *Section 7* of this document.

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 NIST IR 7742.

The results of this summative test of the application should reflect an error-free or near error-free user interface. The critical analysis of this study is in the examination, through post-testing interview techniques, of any observed errors or task execution difficulties, confusions, hesitations, corrected mistakes, etc.

3.3 What the EHR Usability Protocol Is Not

The type of usability testing protocols described in the EUP here are focused on the optimization of workflow and efficiency and safety of interactions in the EHR application or system. This evaluation, focused on user interface design evaluation, is meant to be independent from factors that engender creativity, innovation or competitive features of the system. The usability testing process documented here does not question an “innovative feature” being introduced by a designer, but could identify troublesome or unsafe implementation of the user interface for that innovative feature.

4 EUP Foundation: Research Findings Defining EHR Usability Issues and Their Impact on Medical Error

This section provides an in-depth look at research findings on critical EHR usability issues and their potential impact on reducing medical error. The scope of this review is limited to discussion of the relationship of usability and potential medical error. This section provides the technical foundation for early design evaluations. 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.

Others have identified poor usability as one of the major obstacles to adoption of health information technology at various levels^{5,6 7,8} Therefore, improving usability can be expected to enhance EHR adoption, and efficiency of use, while reducing user frustration, costs and disruptions in workflow.⁹ In addition, secondary users (defined as users not providing clinical care directly to a patient, most notably researchers seeking to improve quality of patient care¹⁰ and reduce health disparities^{11,12,13,14,15,16}) are anticipated to benefit from the electronic data generated by EHR use.

EHRs offer great promise for improving healthcare processes and outcomes, including increased patient safety. As with any health information technology, usability problems that can adversely impact patient safety with EHRs can be assessed, understood and controlled. In this section, research findings were synthesized on EHR usability issues and their potential impact on patient safety. These findings provide the technical foundation for the EUP for detecting and mitigating potential use errors.

Emerging evidence suggests that the use of health information technology (HIT) may help address significant challenges related to healthcare delivery and patient outcomes.¹⁷ For example, three recent

⁵ Miller, Robert H. and Ida Sim. (2004). Physicians' use of electronic medical records: Barriers and solutions. *Health Affairs*, 23(2), 116-126.

⁶ Yusof MM, Stergioulas L, Zugic J. Health information systems adoption: Findings from a systematic review. *Stud Health Technol Inform*. 2007;129(pt 1):262-266.

⁷ Smelcer JB, Miller-Jacobs H, Kantrovich L. Usability of electronic medical records. *J Usability Studies*. 2009;4:70-84.

⁸ EMR 2011: The Market for Electronic Medical Record Systems. Kalorama Information. March 1, 2011.

<http://www.kaloramainformation.com/EMR-Electronic-Medical-6164654/> Accessed May 23, 2011.

⁹ Kaufman D, Roberts WD, Merrill J, Lai T, Bakken S. Applying an Evaluation Framework for Health Information System Design, Development, and Implementation. *Nursing Research*: March/April 2006 - Volume 55 - Issue 2 - pp S37-S42.

¹⁰ For example, see the Quality Data Model project by the National Quality Forum, which has Version 3.0 posted at http://www.qualityforum.org/Projects/h/QDS_Model/Quality_Data_Model.aspx

¹¹ Gibbons MC, Ed. (2011). *eHealth Solutions for Healthcare Disparities*. Springer.

¹² Gibbons MC, Casale CR. Reducing disparities in health care quality: the role of health IT in underresourced settings. *Med Care Res Rev*. 2010 Oct;67(5 Suppl):155S-162S. Epub 2010 Sep 9.

¹³ Bakken S, Currie L, Hyun S, Lee NJ, John R, Schnall R, Velez O. Reducing health disparities and improving patient safety and quality by integrating HIT into the Columbia APN curriculum. *Stud Health Technol Inform*. 2009;146:859.

¹⁴ Muñoz RF Using evidence-based internet interventions to reduce health disparities worldwide. *J Med Internet Res*. 2010 Dec 17;12(5):e60.

¹⁵ Viswanath K, Kreuter MW. Health disparities, communication inequalities, and eHealth. *Am J Prev Med*. 2007 May;32(5 Suppl):S131-3.

¹⁶ Gibbons MC. A historical overview of health disparities and the potential of eHealth solutions. *J Med Internet Res*. 2005 Oct 4;7(5):e50

¹⁷ Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The Benefits Of Health Information Technology: A Review Of The Recent Literature Shows Predominantly Positive Results. *Health Aff* March 2011 vol. 30 no. 3 464-471.

reports suggest that the use of HIT may improve health care outcomes¹⁸ and reduce patient mortality.¹⁹ In addition, the use of HIT is a key component of a national strategy to improve healthcare quality and patient safety.²⁰ On the other hand, a prior study found that patient mortality unexpectedly increased following the introduction of an EHR in a pediatric hospital.²¹ The potential linkage between EHR usability and patient outcomes should be further studied.

Currently, the experience base in the United States with meaningful use of EHRs to provide care in hospital and outpatient settings is limited. Over the next few years, this experience base is anticipated to grow rapidly. Given that one in three patients has been estimated to be potentially harmed during hospitalization,²² the potential for the EHR to improve patient safety may be significant.

Usability issues may 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.²³ As a result, the FDA has placed increased emphasis on testing of device user interfaces in pre-market approval as evidenced by recent publication of the agency's human factors guidance.²⁴

Usability is not a new issue in safety related industries where human error can have severe consequences. Lessons learned from decades of experience using human factors and usability methods from these 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, surgicenters, 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.²⁵

¹⁸ See Gibbons MC, Wilson RF, Samal L, Lehman CU, Dickersin K, Lehmann HP, Aboumatar H, Finkelstein J, Shelton E, Sharma R, Bass EB. Impact of consumer health informatics applications. *Evid Rep Technol Assess (Full Rep)*. 2009 Oct;(188):1-546; and Cebul RD, Love TE, Jain AK, Hebert CJ. Electronic Health Records and Quality of Diabetes Care. *N Engl J Med* 2011; 365:825-833.

¹⁹ Longhurst CA, Parast L, Sandborg CI, et al. Decrease in hospital-wide mortality rate after implementation of a commercially sold computerized physician order entry system. *Pediatrics*. 2010;126:14-21.

²⁰ Institute of Medicine, *Crossing the Quality Chasm: A New Health System for the Twenty-first Century* (Washington: National Academy Press, 2001).

²¹ Han YY, Carcillo JA, Venkataraman ST, et al. Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system [published correction appears in *Pediatrics*. 2006;117(2):594]. *Pediatrics*. 2005; 116(6):1506–1512

²² Classen, D.C., et al. Global Trigger Tool' Shows That Adverse Events In Hospitals May Be Ten Times Greater Than Previously Measured. *Health Affairs* 3(4), 581-589.

²³ Comments from FDA spokesperson at AAMI conference June 25. Washington DC.

²⁴ See Draft Guidance: Applying Human Factors and Usability Engineering to Optimize Medical Device Design issued on June 22, 2011. <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM259760.pdf>

²⁵ National Research Council (2011). *Health Care Comes Home: The Human Factors*. Committee on the Role of Human Factors in Home Healthcare, Board on Human-Systems Integration, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

5 Expert Review of the User Interface and Evaluation Methods

In this section, we will:

1. Discuss a model for understanding EHR patient safety risks,
2. Define usability and associated measures,
3. Propose a model approach to understand the usability of EHR systems, and
4. Define categories of medical errors.

5.1 A Model for Understanding EHR Patient Safety Risks

Emanuel and colleagues²⁶ have defined patient safety as:

a discipline in the health care sector that applies safety science methods toward the goal of achieving a trustworthy system of health care delivery. Patient safety is also an attribute of health care 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 error is 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 created for medical error events by Mokkarala, Zhang and colleagues after reviewing a number of similar taxonomies described in their article²⁷ 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.

²⁶ Emanuel L, Berwick D, Conway J, Combes J, Hatlie M, Leape L, Reason J, Schyve P, Vincent C, Walton M. In: Henriksen K, Battles JB, Keyes MA, Grady ML, editors. *SourceAdvances in Patient Safety: New Directions and Alternative Approaches* (Vol. 1: Assessment). Rockville (MD): Agency for Healthcare Research and Quality; 2008 Aug. *Advances in Patient Safety*.

²⁷ Mokkarala P, Brixey JJ, Johnson TR, Patel VL, Zhang J, Turley JP. Development of comprehensive medical error ontology. In: *AHRQ: advances in patient safety: new directions and alternative approaches*. Rockville, MD: Agency for Healthcare Research and Quality; July 2008. AHRQ Publication Nos. 080034 (1-4).

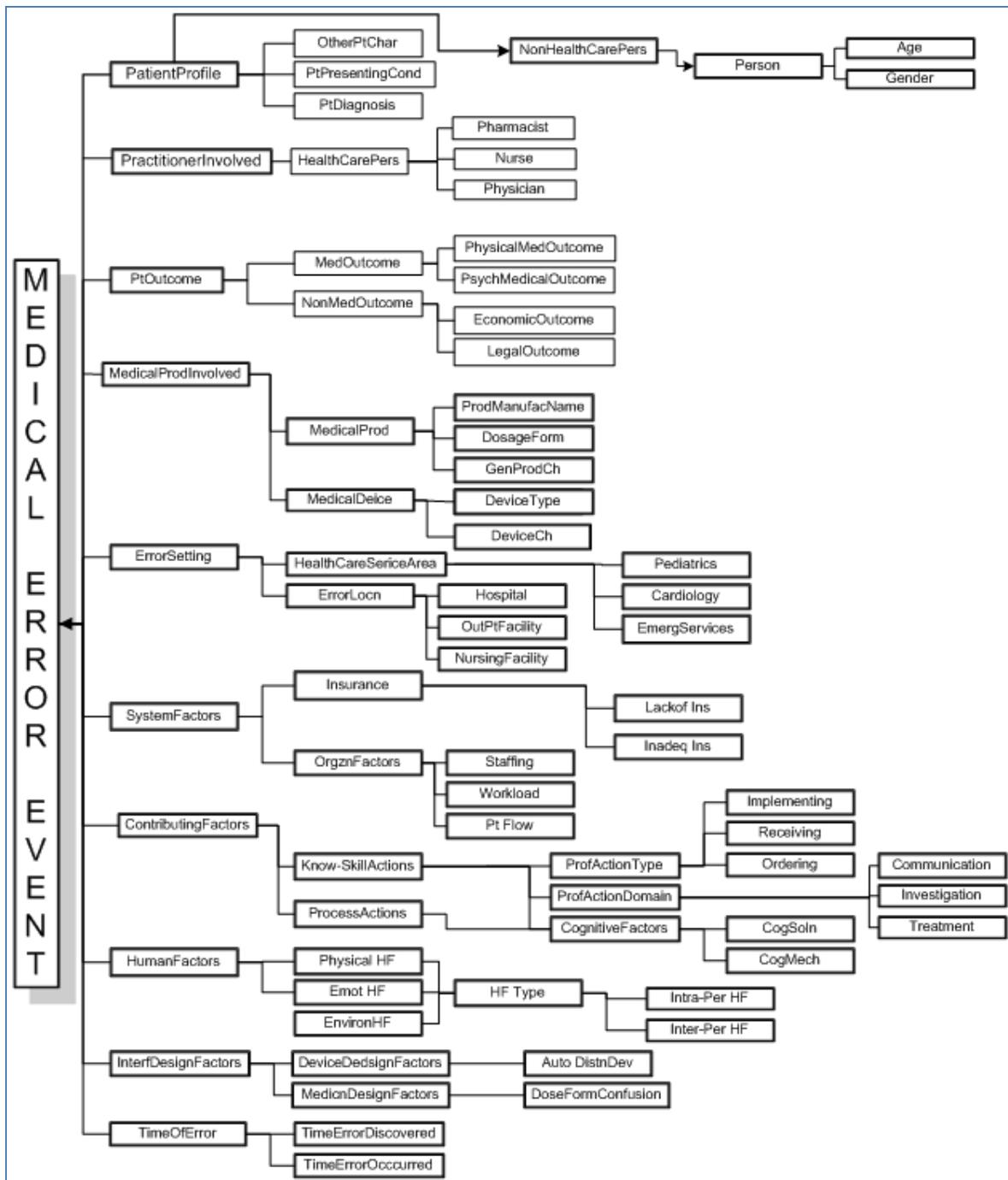


Figure 2. Taxonomy of contributors to medical error events (Mokkarala et al., [2008]; used with permission)

A summary of research findings on critical use errors and how the potential for patient safety is moderated by risk factors and suggested by evaluative indicators is displayed in Figure 3.

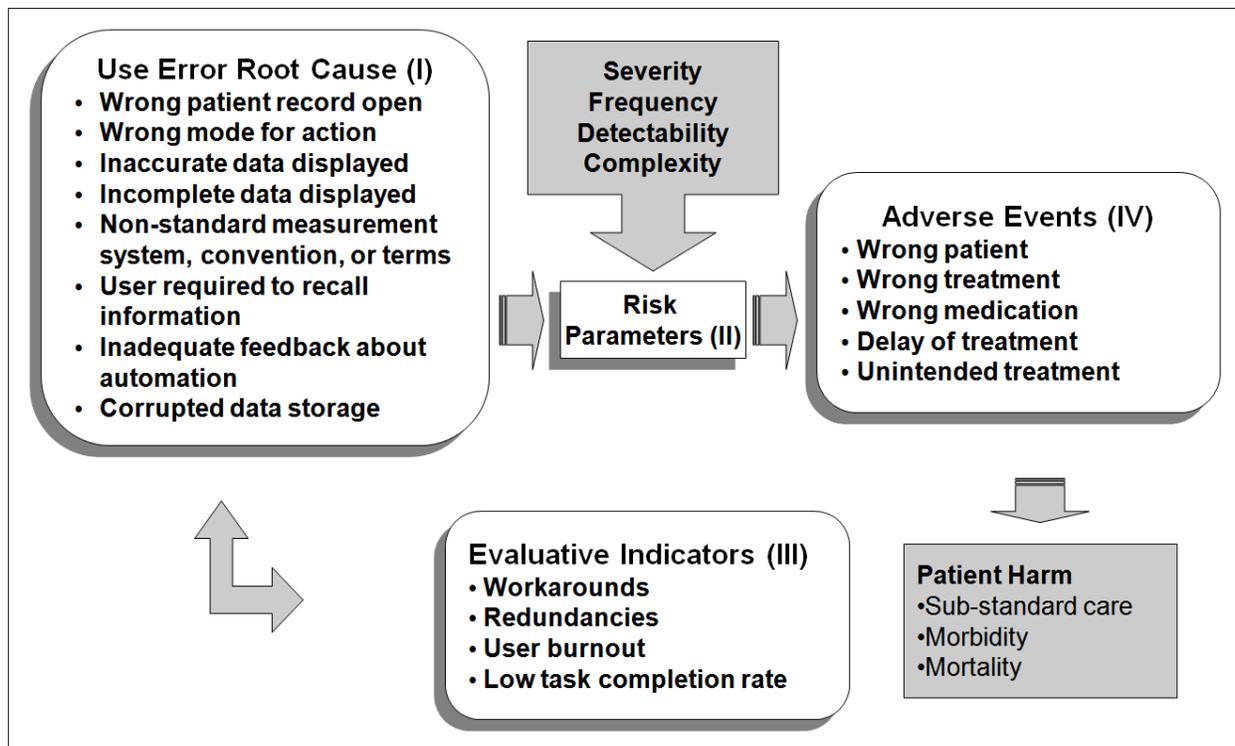


Figure 3. A model for analysis and understanding of use related risks of EHR systems.

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 that are identified in direct observations of system use by the evaluation team in the environment of use or 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 and **hypothetical illustrative** examples are:

- **Patient identification error:** Actions are performed for one patient or documented in one patient’s record that were intended for another patient (e.g., the wrong limb was removed because two patient identifiers were not displayed and it was not possible for the surgeon who was not there when the patient record was opened to verify the patient identity without double-billing the patient).

- **Mode error:** Actions are performed in one mode that were intended for another mode (e.g., *direct dose vs. weight dose*: a 100 kilogram patient received a 100X overdose of a vasoactive drug because weight dosing (mcg/kg/min) was selected instead of direct dosing (mcg/min) due to lack of feedback about an unusual mode choice, no warning about an unusually high dose, and 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; e.g., *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, resulting in a pediatric patient nearly being administered a medication with a dangerously high dose).
- **Data accuracy error:** Displayed data are not accurate (e.g., 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 requires additional navigation, access to another provider's note, taking actions to update the status, or is not updated within a reasonable time (e.g., a patient received four times the intended dose of a medication because the comments field was not visible without being opened that explained that there were progressive dose reductions (taper dose) 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 (e.g., 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 (e.g., the wrong dose of a medication is ordered because, during the ordering process for an outpatient medication, if a one-time (once) schedule is initially selected, the user must enter the appropriate quantity manually, even though for regular orders the user can select the dose from a list).
- **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 (e.g., a patient received 8 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 change was instituted as a patient safety measure to avoid patients needing to understand how to administer partial tablets in the home setting, but no feedback was provided to the user that there was a change in the dose).
- **Data integrity error:** Decisions are based on stored data that are corrupted or deleted (e.g., a patient received more than one pneumococcal vaccine because documentation that the vaccine was given previously was automatically deleted by the EHR; specifically, when the physician selected the clinical reminder dialog box option "Order Pneumovax vaccine to be administered

to patient”, and then clicked on the *Next Button* to process the next clinical reminder for a diabetic foot exam, the text documenting the order disappeared from the progress note. If the physician had pressed the *Finish Button* instead of *Next*, then the text documenting the order of the vaccine would have been correctly saved in the chart).

Risk Parameters (II), defined as controllable or uncontrollable factors that affect variation in the magnitude of the potential risk due to a use error. Risk parameters²⁸ are:

- **Severity:** Magnitude of potential harm,
- **Frequency:** Probability of harm occurring,^{29,30}
- **Detectability:** Ease of recognizing use error that could lead to potential safety issues, and
- **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.³¹

Evaluative Indicators (III), 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, and observations of the system in use, gaps in optimizing user interaction design needs can be identified. In addition, use cases and scenarios for usability evaluations can be developed that are more likely to detect system flaws that create use error hazards or traps proactively. Preliminary categories are:

- **Workarounds:** Users identify differences between the system’s design and their locally adopted workflow,
- **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:** Users frequently initiating, but not completing, a task, and
- **Potential risk for patient safety:** Defined as accidental or preventable injuries attributable wholly or partially to the design or implementation of an EHR.

²⁸ Note that these factors collapse six theoretical dimensions of risk identified by Perrow in *Normal Accidents: Living with High-Risk Technologies*, by Charles Perrow, Basic Books, NY, 1984: Identity; Permanence; Timing; Probability; Victims; and Severity. 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.

²⁹ 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.

³⁰ 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.

³¹ 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>

Adverse Events (IV), 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 to be committed. These event outcomes are similar to the patient safety checklist items for EHRs developed by HIMSS³² 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, and
- **Unintended or improper treatment:** A patient receives unintended care due to misunderstanding how to provide care in the system or due to actions taken to test software, train users, or demonstrate software to potential customers.

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

- **Sub-standard care**, defined as lack of preventive care, wrong 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**, and
- **Mortality**

5.2 Definition of Usability and Associated Measures

The 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.”³³ These terms are further defined as³⁴:

³² HIMSS EHR Usability Task Force. Defining and Testing EMR Usability: Principles and Proposed Methods of EMR Usability Evaluation and Rating. June 2009. Accessed September 10, 2011 at http://www.himss.org/content/files/HIMSS_DefiningandTestingEMRUsability.pdf.

³³ ISO/IEC. 9241-14 Ergonomic requirements for office work with visual display terminals (VDT)s - Part 14 Menu dialogues, ISO/IEC 9241-14: 1998 (E), 1998.

- **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 to the use of the product, and
- **Context of use:** characteristics of the users, tasks and the 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 quality³⁵. 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, and
- **Accessibility:** usability and safety for users with specified disabilities.

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

Usability is correlated with how useful a system is perceived to be.^{36,37} Based on the rationale that usefulness cannot be meaningfully separated from usability, Zhang and Walji³⁸ describe how they have expanded the ISO definition for usability, resulting in three dimensions for usability:

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

³⁴ ISO/IEC. 9241-14 Ergonomic requirements for office work with visual display terminals (VDT)s - Part 14 Menu dialogues, ISO/IEC 9241-14: 1998 (E), 1998.

³⁵ ISO/IEC CD 25010.3: Systems and software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Software product quality and system quality in use models. ISO, First Edition March 1, 2011.

³⁶ 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.

³⁷ Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13 (3), 319-340.

³⁸ Zhang J, Walji M. (2011, in press). TURF: Toward a unified framework of EHR usability. *Journal of Biomedical Informatics*, 00, 000-000

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.”³⁹ Figure 4, from the 2007 American National Standard for Medical Devices on Application of usability engineering to medical devices⁴⁰ 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.

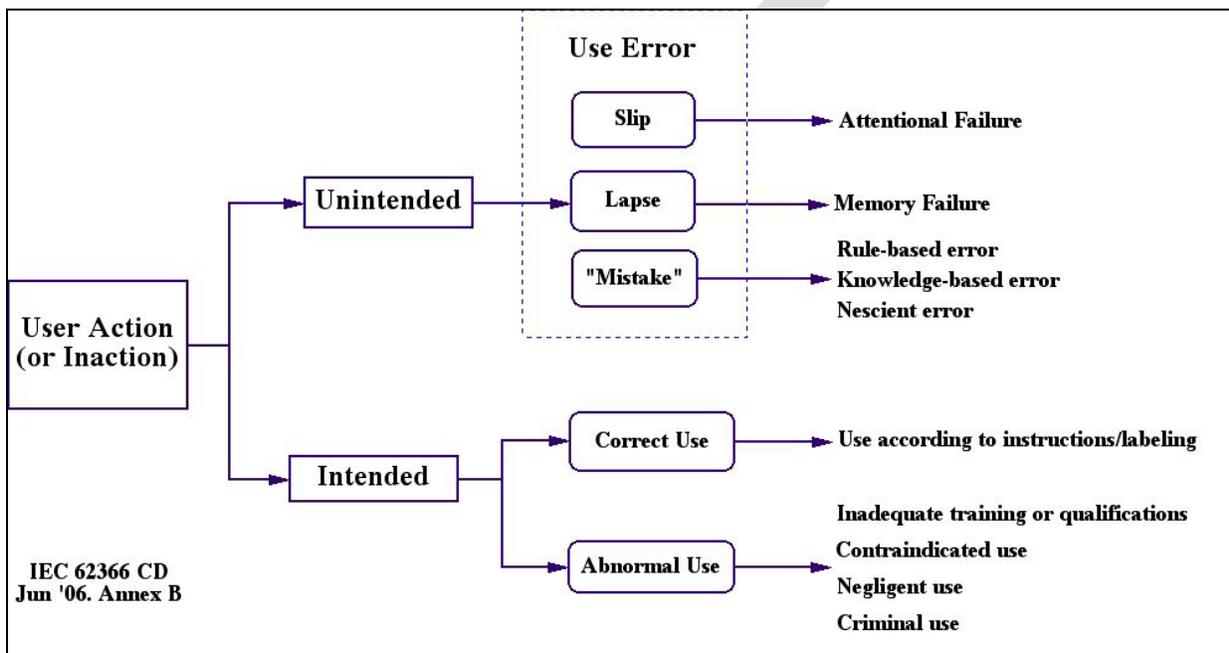


Figure 4. Relationship of user actions and use errors (from ANSI/AAMI/IEC 62366)

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”:

- Device use requires physical, perceptual or cognitive abilities that exceed the abilities of the user,
- The use environment can affect 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,

³⁹ ANSI/AAMI HE75:2009, Human factors engineering — Design of medical devices

⁴⁰ ANSI/AAMI/IEC 62366:2007 -- Medical devices - Application of usability engineering to medical devices

- 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, or
- Devices are used in ways that were anticipated but inappropriate and for which adequate controls were not applied.”⁴¹

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 of time, number of errors, etc. We elaborate on these methods and their application to the EUP in this document.

5.3 Objective, Measurable Dimensions of Usability

Systems that have better usability on these dimensions can be expected to be adopted more quickly and have fewer or no negative, undesirable consequences due to extensive user workarounds:

Efficiency, measured objectively as:

- Time to accomplish a task (average, standard deviation), and
- 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 correspondence of 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 text font size and type, and
- Appropriate and consistent use of measurement units (e.g., kilograms).

The remainder of this section regarding EHR usability focuses on critical use errors with EHR software that may lead to medical error and user interface design that promotes **use error prevention**, defined in ISO 25010 as a sub-characteristic of usability: “the degree to which the system protects users against making errors.” We anticipate that the potential for these use errors will be reliably and validly predicted based on a summative usability evaluation conducted by qualified usability/human factors professionals prior to EHR implementation/deployment.

⁴¹ "Applying Human Factors and Usability Engineering to Medical Device Design". Available online at: <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm259748.htm>. Draft version released June 22, 2011.

5.4 Categorization of Medical Errors

The effect of medical errors can be categorized by potential for patient harm. The National Coordinating Council for Medication Error 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).⁴² 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.

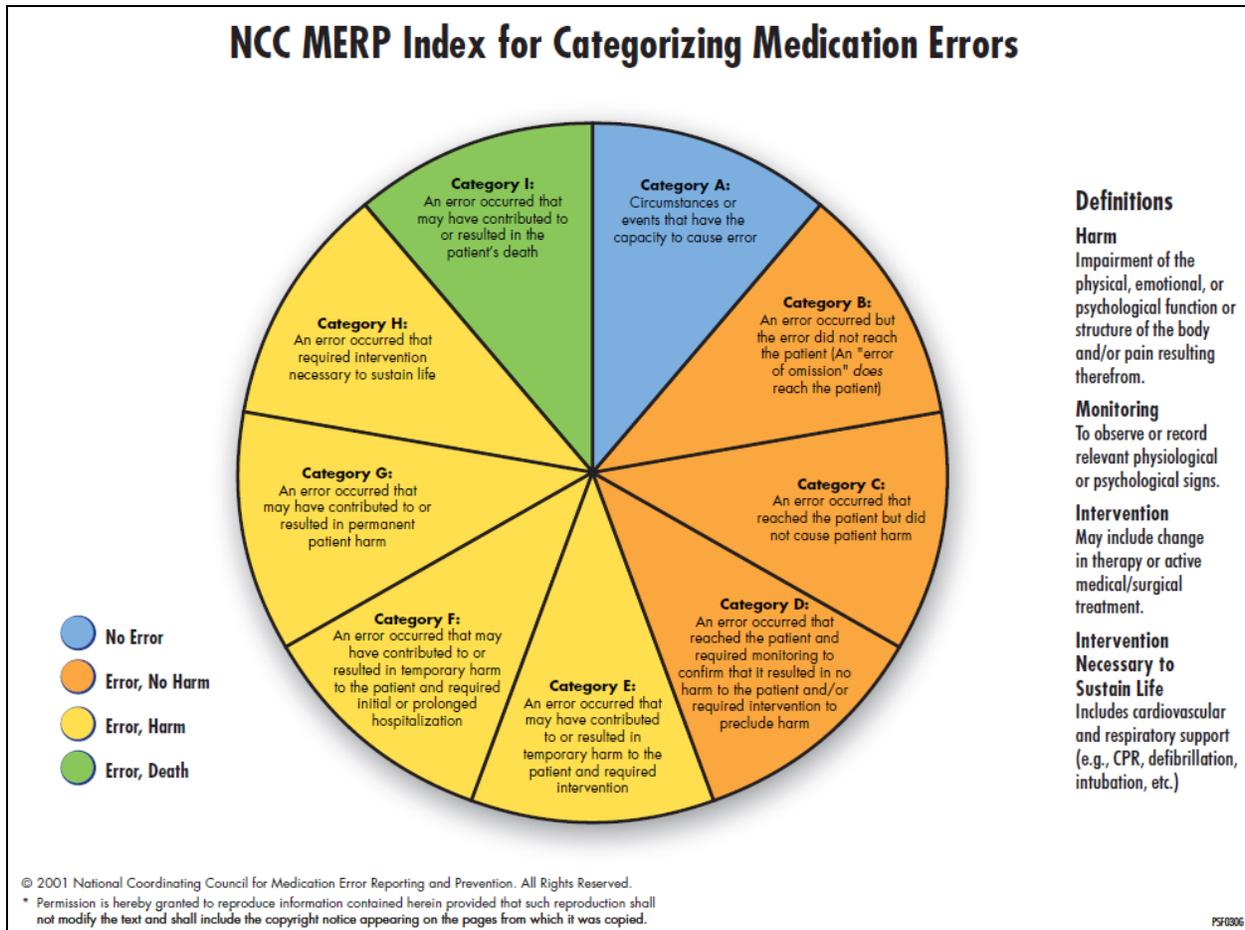


Figure 5. Levels of Patient Harm (National Coordinating Council, [2001]; used with permission)

The Veteran's Health Administration has been a recognized leader since 2000 in reducing the potential of patient harm due to health information technology design and implementation. The VA Office of Information 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

⁴² Forrey RA, Pedersen CA, Schneider PJ. Interrater agreement with a standard scheme for classifying medication errors. *Am J Health-System Pharm.* 2007;64(2):175-181.

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, as detailed in Table 1.

Rating	Severity	Frequency	Detectability
4	Catastrophic: Failure can cause death, injury or illness that requires medical or surgical intervention to prevent permanent loss of function in sensory, motor, physiologic or intellectual skills to patient (or supporting evidence).	Frequent: Likely to occur immediately or within a short period (may happen several times in 1 week).	Remote: Failure is not detectable within the use of a single system.
3	Major: Failure can cause permanent lessening of bodily function (including but not limited to sensory, motor, physiological or intellectual) and disfigurement to patients.	Occasional: Probably will occur (may happen several times in 1 month).	Low: Requires individual knowledge of process to determine abnormality in system.
2	Moderate: Failure can cause injury or illness that requires medical or surgical intervention, or requiring increased length of care to patients.	Uncommon: Possible to occur (may happen sometime in 1 year).	Moderate: Adequate data is available, however requires further cognitive processing.
1	Minor: Failure causes no injury or illness, and requires no medical or surgical intervention other than first aid treatment. Requires no increased length of care to patients.	Remote: Unlikely to occur (may not happen in lifetime of software/system).	High: The defect is obvious and failures can be avoided.

Table 1. Veteran’s Health Administration’s triage ratings used for internally reported Patient Safety Issues due to information technology design

Summary

In this section, we presented a model for categorizing the effects of medical error applicable to EHR systems, and synthesized research findings to define categories of critical usability issues that could have the potential to induce use errors by professional healthcare personnel. These findings provide the technical risk-based foundation for determining what should be tested in Step Three of the EUP, User Testing. However, an overall EHR design objective should include proactive detection and elimination of potential use errors well before summative testing and system deployment. It is our hope that these findings will also be useful for any patient care context in which health information technology is used.

6 Expert Review/Analysis of EHRs

6.1 Criteria and Best Principles for Selecting Expert Reviewers

Principles of good user interface design are well defined in the literature. Nielsen and Molich (1990),⁴³ Norman (1986),⁴⁴ 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)⁴⁵ refined this early work and developed a set of principles that are still widely circulated today, including:

- The 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, and
- 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 which 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.

Experts participating in these reviews should be selected based on criteria of education, experience and expertise in user interface design and human factors. Each of the reviewers should have (a) an advanced degree in a human factors discipline, and (b) a minimum of three years of experience with EHRs and other health information technologies. Specific requirements are outlined below. These Expert Reviewers should:

- Have earned a Master's Degree or higher in one of the following areas or in a closely related area that pertains to human interaction with computing technologies.⁴⁶

⁴³ Nielsen, J., and Molich, R. (1990). Heuristic evaluation of user interfaces, Proc. ACM CHI'90 Conf. (Seattle, WA, 1–5 April), 249-256

⁴⁴ Norman, Donald A. (1986): Cognitive engineering. In: Norman, Donald A. and Draper, Stephen W. (eds.). "User Centered System Design: New Perspectives on Human-Computer Interaction". Lawrence Erlbaum Associates pp. 31--61

⁴⁵ Nielsen, Jakob (1994). Usability Engineering. San Diego: Academic Press. pp. 115–148. ISBN 0-12-518406-9.

- 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 should 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 research and in user interface design of EHRs and/or other health information technologies.

6.2 Protocol for Conducting a Review

The process is similar to that used in the SHARPC program and is adapted from Nielsen⁴⁷ as follows:

1. 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 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. Importantly, in the case of the EUP, the inspection is done on general usability principles and also on the use errors identified in Figure 3. *Appendix B* provides a detailed form for evaluating the user interface. 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 data sets that are to be present during the review. We recommend using the sample cases outlined in *Appendices C, D, and E* as the data.

⁴⁷ http://www.useit.com/papers/heuristic/heuristic_evaluation.html

- ❑ Each expert reviewer proceeds to review each page and evaluate that page with the checklist from *Appendix B*.
- ❑ When issues are identified, in the comment field, the expert should (a) document the page, (b) the context, (c) the severity, and comment 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 following the guidelines in Table 1 above.

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 comments and justification on the user interface good practices that are considered violated. Ultimately, the expert provides a completed expert review form that must be consolidated with the reviews done by other experts. 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?).

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.
3. Once all experts have completed the evaluations, a lead expert consolidates all the findings into a single report. This combined 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 the usability best practices, there are often fixes that are obvious to those responsible for making the change.

To be most effective, and identify the greatest number of problems, there should be more than one expert evaluator (Nielsen and Landauer, 1993)⁴⁸.

⁴⁸ Nielsen, J., and Landauer, T. K. 1993. A mathematical model of the finding of usability problems. *Proceedings ACM/IIFIP INTERCHI'93 Conference* (Amsterdam, The Netherlands, April 24-29), 206-213.

7 Summative: Protocol for Validation Testing EHR Applications

7.1 Purpose of Protocol and General Approach

The purpose of this protocol is to set out general testing conditions and steps that will apply to validation of the usability and accessibility testing of EHR applications. *To distinguish conditions and steps that are exclusive to testing accessibility with participants with disabilities, those conditions and steps that pertain to participants with disabilities are set out in italics.*⁴⁹ The term *participant* refers to a person who is a representative user of an EHR application. This person may or may not have disabilities. A participant may be in one of several representative user groups including, but not limited to the following: physician, physician's assistant/nurse practitioner, or nurse and could work in multiple clinical settings.

Formative vs. Summative Usability Tests. It should be noted that *summative* validation testing, the subject of this protocol, has different purposes and intent from *formative* usability testing, which represents a series of more informal evaluations with the user interface during the design process. Summative validation testing should allow participants to perform steps in a pre-determined set of use scenarios without intervention or interaction with test moderators, while formative tests may engage the users in dialogues with the moderator or “walk-through, talk-through” methods to isolate points of confusion and difficulty.

Performance difficulties and errors that occur when using EHRs are variable. Counting “failures” is insufficient to understand the usability and safety of the system because of this variation and because of scenarios that result in undesirable outcomes. Therefore a somewhat more challenging test process that involves describing the failures as well as their outcomes and whether patterns of similar or identical failures occur is necessary. If late design stage summative testing finds residual problems with the use of the system, those problems 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. This necessitates a combination of objective performance observation and post-test inquiry with participants to identify the source of any observed problems.

The EUP protocol should:

1. Provide a rationale for what user tasks are being tested, and how these test tasks relate to either (a) identifying user interaction leading to potential medical error, and/or (b) general interaction efficiency with the application.
2. Describe the participant selection criteria for the appropriate user test population.
3. 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.

⁴⁹ 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 for accessibility.

4. Describe the facility and equipment required to represent final or near-final design of the user interface including use of the appropriate application platform(s) that will be used (e.g., desktop computer, smart phone, touch screen, tablet PC, etc.) Where appropriate, multiple platforms of application delivery should be tested. An EHR can have different appearances on different platforms. That is, one EHR is available on a desktop PC and is also on a tablet PC. These instantiations might be considered to be separate and distinct software products even though they are distributed from the same company.
5. Provide a checklist of steps for testers (i.e., qualified usability/human factors professionals) of EHR applications to follow.

We strongly urge the designer/developer to create a systematic test plan before beginning any usability test. Elements of the testing can be found in many published volumes and are also outlined in Section 9 of NIST IR 7741 - [*NIST Guide to the Processes Approach for Improving the Usability of Electronic Health Records*](#). A detailed description of each of the steps in conducting the EUP follows later in this section.

Qualified usability/human factors professionals ('testers') should oversee the overall validation testing. Such professionals should be selected based on criteria of education, experience and expertise in human factors. It is recommended that the professional have (a) an advanced degree in a human factors discipline, and (b) a minimum of three years experience with EHRs and other health information technologies. These recommendations do not extend to all members of the testing team.

Two testers conduct the protocols; both must be present at each session. These two testers are:

1. An expert/test administrator who facilitates the testing and is in charge of 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 accessibility testing, the expert/test administrator must also be 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 accessibility testing, the assistant/data logger must also have some expertise in accessibility.*

In addition to the expert/test administrator and the assistant/data logger, there may be other testers present at sessions. The expert/test administrator and assistant/data logger may fill additional roles that may, alternatively, be filled by other testers. For example, the expert/test administrator and assistant/data logger may serve as recruiter, test facility selector, greeter, tester or systems administrator.

1. Testers who act as recruiters must be skilled in interviewing. *In the case of accessibility testing, 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 the case of accessibility testing, 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 act as greeters must be skilled in interacting with test participants. *In the case of accessibility testing, they must be skilled in interacting with people with disabilities.*
4. Testers who set up the testing environment must be skilled in usability testing and must know how to set up and operate the EHRUT (EHR System Under Test). *In the case of accessibility testing, 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.*
5. 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 and computers. *In the case of accessibility testing, 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.

7.2 Overview of Protocol Steps

This section sets out the major steps of the protocol in a temporal sequence so that testers can follow them sequentially:

1. Recruit and schedule participants.
2. Set up test environment.
3. Set up EHR systems and materials.
4. Greet, orient and instruct participants
5. Conduct the testing.
6. Debrief participants and ready the test environment for the next participant.
7. Store, analyze data and report the results.

Step 1. Recruit and Schedule Participants

The testing must meet all Federal and state legal requirements for the use of human subjects.

The test lab will typically have to over-recruit 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. The inclusion criteria are:

- Are currently practicing as a medical professional as one of the following:
 - Physicians (Attendings and Residents)
 - Mid-level providers (Advanced Registered Nurse Practitioners and Physician Assistants)
 - Nurses (Registered Nurses and Licensed Practical Nurses)
- Are literate in English,
- Have no significant commercial connection to any EHR vendor, e.g., no close relative as an employee or owner, and
- Do not have a background in computer science.

For physician participants, recruitment should be balanced as best as possible to the population demographics (See for example: *AMA Physician Statistics, AMA Physician Characteristics and Distribution in the US 2008, 2009*).

For mid-level provider participants, their recruitment should be balanced as best as possible to the 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 nurse participants, their recruitment should be balanced as best as possible to their population demographics (See: [2008 Health Resources and Services Administration \(HRSA\), Bureau of Health Professions, Division of Nursing](#)).

Inclusion criteria should be specified on the screener, such as: age, gender, length in position, current user, etc.

Step 1.1 Sample Size

The protocol for testing described in this document is aimed at demonstrating that the system's user interface design promotes both efficient and safe use via a final, summative usability test with representative users. The summative validation study, which is the focus of this EUP, 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 in her research that with 10 users 80% of the problems are found whereas 95% of the problems are found with 20 users. FDA's recommendation for summative validation tests of medical devices is to test a minimum of 15 users per distinct user group (see Appendix B of FDA's draft guidance of June 22, 2011 *Applying Human Factors and Usability Engineering to Optimize Medical Device Design*). Sauro

(2009) presents a very readable discussion on the topic; he, too, settles on a sample of about 20 representative participants per user group to capture most of the variance at a practical level of effort.

In the end, decisions about the number of user groups and the sample size per group have serious implications on experimental design and the validity of the research. Practically, these decisions are in the hands of the application team. Rationale for these decisions should be clearly documented in the final report.

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 NIST IR 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 blindness, low vision, manual dexterity, or mobility disabilities.*
- A recruiter recruits participants.
 - 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.
 - The participant may also want to supply contact information about a care-giver who will accompany the participant to the test location. In this case, a recruiter accepts this information as well.*

- 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 Test Environment

It is not easy to recreate a test environment that simulates the clinical environment. It is true that while the clinical space provides many more contextual clues than a lab or office environment, the tradeoff in cost is likely not worth the incremental value of performing the testing in a full out simulation environment. As such, we recommend the kinds of test setups in common practice in usability testing labs today. 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 room in which to carry out the testing. The test area should have the following characteristics:

- Test room:
 - Minimum dimensions: 10' x 10',
 - Ambient lighting should be in the range of 400-600 lx. If possible, use indirect lighting rather than overhead fixtures or direct sunlight so as to reduce glare.
 - Ambient noise levels should be below 40dB,
 - Adequate ventilation such that the participants do not report a stuffy or drafty feeling,
 - Temperature should be between 68 and 76 Fahrenheit
 - Relative humidity should be between 20 percent and 60 percent,
 - Each room will have one test PC for use by the participant, and
 - (Optional) One-way mirror allowing observers to view the participant using the application.
- Observation room:
 - Minimum dimensions: 10' x 10',
 - Ambient lighting should be in the range of 400-600 lx. If possible, use indirect lighting rather than overhead fixtures or direct sunlight so as to reduce glare. If using a one-way mirror, this room will be dark enabling observers to view the lighted test room.
 - Ambient noise levels should be below 40dB,
 - Adequate ventilation such that the observers do not report a stuffy or drafty feeling,
 - Temperature should be between 68 and 76 Fahrenheit;
 - Relative humidity should be between 20 percent and 60 percent, and
 - (Optional) One-way mirror allowing observers to view the participant 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 That the Test Facility Is Appropriate

- A tester ascertains that a test facility is accessible for parking, sidewalks, pathways, hallways, doorways, corridors and bathrooms per the guidelines in <http://www.ada.gov/stds.pdf>. 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 there is an evacuation procedure in place for the test facility *and that the evacuation procedure provides for evacuation of participants with disabilities.*
- In multi-story 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 the testing environment conforms to OSHA guidelines for air quality (see <http://www.osha.gov/SLTC/indoorairquality/index.html>).
- A tester assures that there is access to drinks and snacks.
- A tester assures the availability of clear floor space on which the EHRUT will be located.
- 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.
- A tester assures that the testing environment has room for all testers and observers; if there will be observers, to work without disturbing the participant, *including testers and observers with disabilities.*

Checklist: Testers Set up the Testing Environment

- A tester is responsible for the setup and configuration of the testing environment such that it conforms to all ADA/Applied Behavior Analysis (ABA) guidelines.
- A tester assures that furniture in the testing environment accommodates participants, testers and, if there are observers, that it accommodates the 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 the camera if a camera is to be used or distract the participant or impede the testers.
 - A tester sets up a place for a data collector to record observation data.
- A tester assures that seats for observers are placed so that observers can view the testing without distracting the participants, interfering with data collection, or interfering with participants' interaction with the EHRUT.
- A tester assures that the test facility and the testing environment are kept free of obstacles throughout all test sessions.

- A tester assures that the route(s) to and from the testing environment is/are free of all obstacles, e.g.,
 - obstacles on the floor
 - obstacles that hang from a ceiling
 - protruding obstacles.
- A tester assures that the testing environment is free of all obstacles, e.g.,
 - obstacles on the floor
 - obstacles that hang from a ceiling
 - protruding obstacles.
- A systems administrator is responsible for setting up the EHRUT.
 - 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 functions correctly.
- 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 sets up the testing equipment.
 - If a camera is to be 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 to capture the participant's comments.
 - A systems administrator is responsible to ensure that the data capture and logging software is properly capturing data on the EHRUT.
- A systems administrator sanitizes equipment that the participant will come into direct contact with.
 - Immediately before each participant session, a systems administrator sanitizes the EHRUT.
 - Immediately before each participant session, a systems administrator sanitizes the assistive device(s) to be used in that session.*
 - If a questionnaire is used, immediately before each participant session, if the participant is going to complete a questionnaire using a keyboard, a systems administrator sanitizes the keyboard.

Step 3. Set up EHR Systems and Materials

Step 3.1 Setting up the System

Setting up the EHR is the first step in the testing procedure. 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 the patient records and histories in *Appendices C, D, and E* loaded as data sets.

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 3.2 Tester's Guide and Test Tasks

The tester's guide is the script that the tester uses during the actual testing. An example of the guide can be found in *Appendix G*.

EUP will focus its test cases (at least initially) on functions that potentially could have impact on patient safety related to Meaningful Use (MU) criteria⁵⁰. 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 a 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. A scenario of use developed for this e-Prescribe MU use case would expose this scenario as being potentially harmful to patients not because the software functions fail, but because the functions do not provide sufficient and usable information to the healthcare professional involved in medication administration.

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

Sample Test tasks are presented in *Appendices C, D, and E*. These test cases are developed so that they can be evaluated on both a clinical and usability level.

Step 4. Greeting, Orienting and Instructing Participants

Participant background

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. This requires some means of accounting for getting the participants to a level of proficiency. In testing an EHR, there is a need to understand interface features that increase or decrease the likelihood of error efficiency and learnability.

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

1. Train participants with no experience on EHR,
2. Use participants with experience on the vendor EHR, and
3. Train participants with prior experience on a different EHR, but with no experience on the vendor EHR.

Each of these options presents its own considerations, so the best option is to select the approach that is most practical. Background 1 is probably out of the question as users with no prior experience (a) will be increasingly hard to find, and (b) will likely take too long to get to an adequate level of proficiency in a reasonable period of time. Background 2 brings in users of the vendor EHR and has two things to weigh. 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. Frensch & Sternberg (1989)⁵¹ showed that expert bridge players did worse than novices initially when some of the rules were changed. 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 (that is, 'think aloud' is appropriate for formative, qualitative testing not for summative, quantitative focused testing.). Background 3, while they do not have specific experience in the vendor EHR, participants 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 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 artifacts that must be managed. However, it is recommended that participants have at least one year of consistent clinical use of an EHR, and not the EHR being tested. These criteria must be used during the selection process.

⁵¹ Frensch, P. A., & Sternberg, R. J. (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.

The development team should develop and provide typical self-guided training that is given to users. This training period can last anywhere from zero minutes to two hours depending on the discretion of the development team. Training should represent as realistic a program as possible. It is understood that testing immediately following training biases performance, but as a practical matter it may be a compromise that must be made given the time, cost and complexity of this type of testing.

If the participant requires significant training, then consideration has to be given to the length of the testing session and how the session is managed.

Furthermore, the development team should make available manuals, help systems, or tip sheets as would be available in normal use to end users.

Participant Preparation

The test facilitators are responsible for instructing the participants on the test procedure. Follow these steps for each participant:

- Greet incoming participants and verify that they are here for the appropriate purpose.
- Provide informed consent forms for signature if required; *Appendix H* replicates examples from NIST IR 7742.
- Provide each participant with the application training required.
- Do not 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’.
- Finally, escort the participant to the test room.

Checklist: Escorting the Participant to the Testing Environment

- If, during recruitment, it has been established that the participant requires an escort to the testing environment,
 - A greeter greets the participant at the entrance to the test facility.
 - A greeter escorts the participant from the entrance to the test facility to the testing environment.

Checklist: An Expert/Test Administrator Greets and Orients the Participant

- An expert/test administrator welcomes the participant to the testing environment.
 - An expert/test administrator greets the participant when he or she enters the testing environment.
 - An expert/test administrator verifies that the participant has come to participate in the EHRUT test.
 - An expert/test administrator verifies that this is the participant scheduled for this session.
- An expert/test administrator informs all participants of routes to exits.

- If a participant has a disability, an expert/test administrator informs that participant of the accessible route(s) to accessible exits.*
- If a care giver accompanies a participant with a disability, an expert/test administrator informs the care giver about the accessible route(s) to accessible exits.*
- An expert/test administrator assures that all participants are aware of routes to exits *and if a care giver accompanies a participant with a disability, that the care giver is aware of accessible routes to accessible exits.*
- An expert/test administrator informs participants about emergency evacuation procedures.
 - If a participant has a disability, an expert/test administrator informs that participant about emergency evacuation procedures for people with disabilities.*
 - If a care giver accompanies a participant with a disability, an expert/test administrator informs the care giver about emergency evacuation procedures for people with disabilities.*
 - An expert/test administrator assures that all participants understand emergency evacuation procedures for the test facility and *if a care giver accompanies a participant with a disability, that the emergency evacuation procedures are understood by the care giver.*
- An expert/test administrator shows the participant to the area where he or she will complete paperwork prior to the test.

Checklist: Participant Completes the Consent Form

- An expert/test administrator 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, an expert/test administrator reads the consent form aloud for the participant.
 - An expert/test administrator asks the participant if the participant has any questions about the consent form, and answers any participant questions about the consent form.
- The participant signs both copies of the consent form.
- An expert/test administrator witnesses the participant's signature on both copies of the consent form.
 - If the expert/test administrator has witnessed the participant signing the consent form, the expert/test administrator signs the consent form as a witness.
- If, for any reason, the participant cannot sign the consent form, the expert/test administrator asks the participant if the participant consents.
 - If the participant consents, the expert/test administrator notes this on the consent form.
 - If the participant does not consent, the test is terminated.
- An expert/test administrator offers one copy of the consent form to the participant.
- An expert/test administrator retains the other copy of the consent form and makes it part of the records of the test session.

Step 5. Conduct the Testing

Validation testing is not the same as formative, qualitative testing. With formative testing, the idea often is to collect rich qualitative data to inform design, typically with only a handful of users. Stylistically, the participants are asked to think aloud and frequently there is a lot of questioning and probing after a task to dissect problem areas. By contrast, validation testing is much more hands off by the tester. Tasks are given, 'think aloud' is not used, and performance is recorded and measured. There is no debrief until after all the tasks have been done. Essentially, during validation testing extra care must be taken to avoid the appearance or reality of facilitator influence. While facilitator bias is of concern during formative testing, it is less of an issue than in validation testing.

Step 5.1 Test Administration

Once the participant has been seated, the participant is read the instructions as presented in NIST IR 7742 guidance on participant Instructions. The facilitator provides the context of the testing: (1) the participant will be asked to perform a number of typical tasks on a fictitious patient record. The facilitator is not able to help. (2) Each task will be read to the participant by the facilitator, and then the participant will receive a written card with the task. (3) The participant should do the task as quickly, but as accurately, as possible. The tasks will be timed and the participant should be stopped if their time exceeds the time limit for each task. (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. At this point, once the performance and rating data have been recorded, the tester may do a qualitative debrief with the participant.

Total test session time should not exceed two hours, excluding training.

Step 5.2 Collect Data

During the testing the facilitator provides the context of the test, and presents each task. Both the facilitator and a note taker who is observing (in a separate observation room) do data collection. The description of the performance and rating data to be collected (as well as data sheets) can be found in NIST IR 7742.

Additionally, and critically, since the focus of this protocol is detection and correction of use errors that affect patient safety, full and complete documentation and classification of use errors observed must be provided. Using this framework as backdrop, the data collection must include a plan for identification of use-related 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 or during the review of the test results.

Checklist: Participant Completes the Test Tasks

- An expert/test administrator presents general instructions to the participant. See the Tester's Guide example in *Appendix G* for the general instructions.
 - An expert/test administrator hands the general instructions to participants *including participants with mobility disabilities.*
or
 - An expert/test administrator offers to read the general instructions to participants with blindness or a low vision disability.*
or
 - An expert/test administrator places the general instructions on a height-adjustable table or cart for participants with a mobility or manual dexterity disability.*
- An expert/test administrator verifies that the participant can comfortably read 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 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.
- An expert/test administrator presents task instructions to the participant.
 - An expert/test administrator hands the 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 turn ballot pages 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 announces the end of the test sessions.
- The participant performs all test tasks.
- The participant completes the post-test questionnaire (System Usability Scale) 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 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 maximum allowable time.
 - An assistant/data logger notes if all tasks have been completed.
 - An assistant/data logger notes critical incidents.

- An assistant/data logger takes note of the participant's comments that reflect the participant's experience with the EHRUT.
- If a camera has been used, an assistant/data logger captures camera data throughout the entire time that the participant interacts with the EHRUT.

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

A critical step in conducting a summative validation test focused on uncovering any residual usability problems is the post-test interview with participants. If failures on critical tasks, performance with effort, or successful task completion with corrected errors were observed, these instances should be investigated with the participant as to the root cause of the problem. The objective of this session 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.

Step 6. Ready the Test Environment for the Next Participant

Once the testing is complete, the tester provides the participant with his/her incentive compensation and thanks the participant for his/her time.

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 who is blind or who has a visual or a manual dexterity 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: The Participant Leaves the Testing Environment and 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: An Assistant/Data Logger Collects 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 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.

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.
- A systems administrator resets the EHRUT to the state in which a new user would find it when approaching the EHRUT for daily use.
 - A systems administrator resets all adjustable aspects of the EHRUT, 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. which is marked with the user identification and session identification.

Checklist: A tester closes down the EHRUT and test equipment at the end of the 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 (if a camera has been used) and computer(s) (if computers have been used.)

Step 7. Analyze Data and Report the Results

Documentation of the performance results must be provided according to the layout of the Common Industry Format (CIF) as described in NIST IR 7742.

Additionally, a thorough classification and reporting of errors committed by the participants and those identified by the facilitator is required. This classification should conform to a framework presented in *Appendices K and L*. Two essential elements of this reporting are the priority and the mitigation plan. The tester must realistically estimate the severity, frequency and detectability of the error, and also a risk mitigation plan that describes how and when the interface will be fixed to minimize the error.

Checklist: Testers Analyze Data

- Testers analyze data
 - Each participant's performance on each task is recorded including: task time, task success criteria, efficiency (i.e., steps taken/optimal steps), errors, relevant verbalizations and task ratings.
 - Usability ratings (such as the System Usability Scale – SUS; see *Appendix M*) are also collected for each participant and an aggregate score is computed.

Checklist: Testers Report Test Findings

- Testers identify the EHRUT by name and version/release number.
- Testers report test findings in the CIF.
 - Testers report results on based on task success, path deviations, task times, errors, task ratings and usability ratings.

8 Conclusion

This document summarizes the rationale for an EUP that encompasses protocols for (1) expert evaluation of an EHR from a clinical perspective, a human factors best practices perspective, and (2) validation studies of EHR user interfaces with representative user groups on realistic EHR tasks.

This document centers on improving user performance with EHRs through application of human factors best principles. In addition, we presented areas where the usability could be improved 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 the 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 is an expert evaluation paradigm that consists of a review protocol for human factors and a clinical expert analysis of the user interface to identify and mitigate potential patient safety issues.

The three-step process described within this document for design evaluation and human user performance testing for EHR is focused on (a) increasing safe use of the EHR and (b) increasing ease of use of the EHR by users. The three steps to achieving this are:

1. Usability/Human Factors Analysis of the application during EHR user interface development,
2. Expert Review/Analysis of the EHR user interface after it is designed/developed, and
3. Testing of EHR user interface with users.

The protocol scope in this document describes the overall usability protocol recommended (with examples provided in appendices) and summarizes research findings on the relationship of usability and patient safety applicable to EHRs. *It is our expectation that the potential for all of these 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.*

The sample forms provided in the appendices are samples only; the development teams will need to modify them as necessary for their scenarios. Refer to *Appendix A* next for examples of the successful application of government best practices of HSI (Human-System Integration) .

Appendix A: Government Best Practices of Human-System Integration (HSI)

Throughout the past four decades, the US Government has systematically increased incorporation of human factors analysis, evaluation and testing requirements for government procured 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 HSI implementation 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.

US 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 in 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 Practice

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 have shown 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) Flightdeck 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 US. 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

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>

of the use of the device. The test results should demonstrate that the device has been optimized with 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*—
 - Identify all prospective device user groups and demonstrate an understanding of their potential capabilities and limitations,
 - Develop scenarios of use—high-level descriptions of user interactions involved when performing specific tasks,
 - Analyze all prospective environments and conditions of use for factors that may affect user performance,
 - Analyze the required perceptual, cognitive and physical demands associated with device interactions, and
 - Analyze 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.

DRAFT

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

DRAFT

Expert Review of EHR

System Title: _____

Release #: _____

Evaluator: _____

Date: _____

The ratings used in Table 1 are used here to categorize each violation of a usability principle. These can be classified 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, is the first patient record automatically closed?		
1A.3	When a second user opens a patient chart, is the first user automatically logged out?		
1A.4	When another application (e.g., imaging) is opened from within the EHR, does the display have a title or header with an accurate unique patient identifier?		
1A.5	When an 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), is the user alerted?		
1A.8	If multiple records for a patient have been merged, are users provided feedback about this action?		
1A.9	If information is copied from the record of one patient and pasted into another, is feedback provided to anyone viewing the record that the information was pasted from the record of a different patient?		

1A.10	If information is copied from a previous note of the same patient, is feedback provided to anyone viewing the record what information was pasted without further editing?		
-------	---	--	--

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 for testing new functionality)?		
1B.8	Are special modes (e.g., view only, demonstration, training) clearly displayed and protected against any patient care activities being conducted in them on actual patients?		

1C. Data accuracy error

Displayed data are not accurate.

#	Review Checklist	Severity Rating	Comments
1C.1	Is information truncated on the display (e.g., medication names and doses in pick list menu displays)?		
1C.2	Is information required to be actively refreshed to be displayed accurately?		
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 it removed from all displays?		
1C.8	If numbers are truncated to a smaller number of digits, is feedback provided immediately to the user and to others who view the information at a later time?		
1C.9	If outdated orders are automatically removed from the display, is feedback provided to the user?		
1C.10	If a user enters the date of birth of a patient instead of the current date, is the user alerted?		
1C.11	If a user enters the current date as the patient's birth date, is the user alerted?		

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 feedback provided that there is information in comments field or other background fields that is related to displayed information on a primary display? For instance, Medication order says 80 mg. Comments say "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" does user have to look in comments to know what the actual dose is?		
1D.2	Are complex doses displayed in ways that users could easily misunderstand what is intended on a particular day? (e.g., taper doses that display the original dose before weaning)		
1D.3	If the content of an unsigned note is only available to the provider who wrote the note, are other providers aware that there is an unsigned note?		
1D.4	Is information accurately updated in one place not transmitted to other areas or to integrated software systems? For example, changing dose on medications tab does not update dose where medication information is displayed on discharge summary—anywhere there is a list of medications with doses.		

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	Are generic or brand names of medications used consistently?		
1E.5	Is terminology consistent and organized consistently (e.g., clinical reminders use What/When/Who structure)?		
1E.6	Are negative structures avoided (e.g., Do you not want to quit?)		
1E.7	Are areas of the interface that are not intended for use by certain categories of users clearly indicated?		

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 user need to remember, rather than recognize, information such as medication doses, even for one-time orders?		
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 matches?		
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	Are attempts to save a patient record under a different name warned of the risk of destruction of patient data?		

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	Do automated changes to medication types, doses and routes trigger feedback to the user and the ability to easily undo?		
1G.2	Are changes to displays easy to detect and track?		
1G.3	Are automated merges of patient record data done with sufficient feedback and active confirmation from the user?		
1G.4	Do automated changes to medication orders (to match the formulary or for economic reasons) trigger feedback to the user?		

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 change stored data?		
1H.2	Is it possible to know who is blocking access to a data element or record?		
1H.3	Is it possible for corrupted backup data to update accurate patient information permanently?		
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)?		

2. Visibility of System Status

The system should always keep 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 GUI menus make obvious w 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 speak 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 and interdependent fields appear on the same screen?		
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	Have uncommon letter sequences been avoided whenever possible?		
3.10	Does the system automatically enter leading or trailing spaces to align decimal points?		
3.11	Does the system automatically enter commas in numeric values greater than 9999?		3.
3.12	Do GUI menus offer activation: that is, make obvious how to say "now do it"?		
3.13	Has the system been designed so that keys with similar names do not perform opposite		

	(and potentially dangerous) actions?		
--	--------------------------------------	--	--

DRAFT

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	Can users reduce data entry time by copying and modifying existing data?		
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?		3.
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	Can users set their own system, session, file and screen 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 labeled?		
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	Are vertical and horizontal scrolling possible in each window?		
5.8	Does the menu structure match the task structure?		
5.9	If "exit" is a menu choice, does it always appear at the bottom of the list?		
5.10	Are menu titles either centered or left-justified?		
5.11	Are menu items left-justified, with the item number or mnemonic preceding the name?		
5.12	Do embedded field-level prompts appear to the right of the field label?		
5.13	Are field labels consistent from one data entry screen to another?		
5.14	Are fields and labels left-justified for alpha lists and right-justified for numeric lists?		
5.15	Do field labels appear to the left of single fields and above list fields?		
5.16	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	Is sound used to signal an error?		
6.2	Are prompts brief and unambiguous?		
6.3	Are error messages grammatically correct?		
6.4	Do error messages avoid the use of exclamation points?		
6.5	Do error messages avoid the use of violent or hostile words?		
6.6	Do error messages avoid an anthropomorphic tone?		
6.7	Do all error messages in the system use consistent grammatical style, form, terminology and abbreviations?		
6.8	Do messages place users in control of the system?		
6.9	If an error is detected in a data entry field, does the system place the cursor in that field or highlight the error?		
6.10	Do error messages inform the user of the error's severity?		
6.11	Do error messages suggest the cause of problem?		
6.12	Do error messages provide appropriate semantic information?		
6.13	Do error messages provide appropriate syntactic information?		
6.14	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	If the database includes groups of data, can users enter more than one group on a single screen?		
7.2	Have dots or underscores been used to indicate field length?		
7.3	Is the menu choice name on a higher-level menu used as the menu title of the lower-level menu?		
7.4	Has the use of qualifier keys been minimized?		
7.5	If the system uses qualifier keys, are they used consistently throughout the system?		
7.6	Does the system prevent users from making errors whenever possible?		
7.7	Does the system warn users if they are about to make a potentially serious error?		
7.8	Does the system intelligently inter variations in user commands?	3.	
7.9	Do data entry screens and dialog boxes indicate the number of character spaces available in a field?		
7.10	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 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 long columnar fields broken up into groups of five, separated by a blank line?		
8.7	Are optional data entry fields clearly marked?		
8.8	Are borders used to identify meaningful groups?		
8.9	Is color coding consistent throughout the system?		
8.10	Is color used in conjunction with some other redundant cue?		
8.11	Is the first word of each menu choice the most important?		
8.12	Are inactive menu items grayed or omitted?		
8.13	Are there menu selection defaults?		
8.14	Do data entry screens and dialog boxes indicate when fields are optional?		
8.15	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	Do expert users have the option of entering multiple commands in a single string?		
9.4	Does the system provide function keys for high-frequency commands?		
9.5	For data entry screens with many fields or in which source documents may be incomplete, can users save a partially filled screen?		
9.6	If menu lists are short (seven items or fewer), can users select an item by moving the cursor?		
9.7	If the system uses a pointing device, do users have the option of either clicking on fields or using a keyboard shortcut?		
9.8	Does the system offer "find next" and "find previous" shortcuts for database searches?		
9.9	In dialog boxes, do users have the option of either clicking directly on a dialog box option or using a keyboard shortcut?		
9.10	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 Graphical User Interface (GUI) where menu sequence has already been specified, do menus adhere to the specification whenever possible?		
10.6	Are meaningful groups of items separated 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; for example, 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 the amount of required window housekeeping been kept to a minimum?		
12.6	Has color been used specifically to draw attention, communicate organization, indicate status changes, and establish relationships?		
12.7	Are typing requirements minimal for question and answer interfaces?		
12.8	Do the selected input device(s) match environmental constraints?		
12.9	If the system uses multiple input devices, has hand and eye movement between input devices been minimized?		
12.10	If the system supports graphical tasks, has an alternative pointing device been provided?		
12.11	Is the numeric keypad located to the right of the alpha key area?		
12.12	Are the most frequently used function keys in the most accessible positions?		
12.13	Does the system complete unambiguous partial input on a data entry field?		

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 completely inaccessible?		
13.2	Can protected or confidential areas be accessed with a certain password?		

DRAFT

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.

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.

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

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

Methadone – the patient is not sure of the dose

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.

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 ones. 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.

DRAFT

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.

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 NIST IR 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 to reflect your EHR's primary audience.*

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 held this position? _____ *[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. What 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.

DRAFT

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 you will take a look at 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. You will be asked to complete these tasks on your own trying to do them as quickly as possible with the fewest possible errors or deviations. Do not do anything more than asked. If you get lost or have difficulty I cannot answer help you with anything to do with the system itself. While this is not a timed test, there may be moments when I step in and move us on to another task. 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.]* Some of the data may not make sense as it is placeholder data.

We are recording the audio and screenshots of our session today. 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?

Preliminary Questions (X minutes)

What is your job title / appointment?

How long have you been working in this role?

What are some of your main responsibilities?

Tell me about your experience with electronic health records.

DRAFT

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. Find 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 / Notetaker 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 / Notetaker 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 / Notetaker 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?

Thank participant once completed.

Appendix H: Informed Consent and Non-Disclosure Forms Adapted from NIST IR 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: _____

DRAFT

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	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	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	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	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 (HCII 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: _____

Appendix K: Form for Reporting Potential Use Errors

No.	Potential use error	Mitigation Plan	Priority
1.C.1	Data accuracy error: Medication doses truncated in pick list menu makes it easy to pick the wrong dose	Do not truncate names at 40 characters, but instead display 75 characters and the remainder viewed upon mouse roll-over	High
1.F.6	Recall error: Physicians might forget that patients have allergies to medications while ordering, even though it is displayed	Provide pop-up “Are you sure?” alerts when a physician orders and a pharmacist verifies a medication order to which a patient has an allergy	High
1.C.3	Data accuracy error: Unable to document reasons other than available in pick lists for ordering medication	Add “other” option in pick list with optional free text entry	Moderate

Appendix L: Form for Tracking Resolution of Potential Use Errors

No.	Date Found	Date Fixed	Date Fix Released	Reported?	Contact	Resolution	Related Issues	Priority
2011-1.C.1	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	2011-1.C.3	High

[Clear – Closed, Green – awaiting fix, Yellow – Analysis ongoing, Red – newly reported, awaiting analysis]

DRAFT

Appendix M: Example of Data Summary Form

Task \ Measure	N	Task Success	Path Deviation	Task Time		Errors	Task Ratings 5=Easy
	#	Mean (SD)	Deviations (Observed / Optimal)	Mean (SD)	Deviations (Observed / Optimal)	Mean (SD)	Mean (SD)
1. [Find item on patient summary screen]							
2. [Use patient chart to find lab results]							
3. [Check vital signs]							
Etc.							

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

AMIA and its members have written extensively on the topic of usability of EHRs and other clinical systems, for example:

A. W. Kushniruk and V. L. Patel, "Cognitive and usability engineering methods for the evaluation of clinical information systems," *Journal of Biomedical Informatics*, vol. 37, pp. 56-76, 2004.

Britto, Maria T., Holly B. Jimison, Jennifer Knopf Munafo, Jennifer Wissman, Michelle L. Rogers, and William Hersh (2009). Usability Testing Finds Problems for Novice Users of Pediatric Portals *J Am Med Inform Assoc*. Sep–Oct; 16(5): 660–669.

C. M. Johnson, *et al.*, "A user-centered framework for redesigning health care interfaces.," *Journal of Biomedical Informatics*, vol. 38, pp. 75-87, 2005.

Phansalkar S, Edworthy J, Hellier E, Seger DL, Schedlbauer A, Avery AJ, Bates DW. (2010). A review of human factors principles for the design and implementation of medication safety alerts in clinical information systems. *J Am Med Inform Assoc*. Sep–Oct;17(5):493-501.

Ralston, James D., David Carrell, Robert Reid, Melissa Anderson, Maureena Moran, and James Hereford (2007) Patient Web Services Integrated with a Shared Medical Record: Patient Use and Satisfaction *J Am Med Inform Assoc*. Nov–Dec; 14(6): 798–806.
doi: 10.1197/jamia.M2302 Correction in: *J Am Med Inform Assoc*. 2008 Mar–Apr; 15(2) : 265.

Rosenbloom, S. Trent, Randolph A. Miller, Kevin B. Johnson, Peter L. Elkin, and Steven H. Brown (2008) A Model for Evaluating Interface Terminologies *J Am Med Inform Assoc*. Jan–Feb; 15(1): 65–76

Saitwal, H., Feng, X., Walji, M., Patel, V. L., & Zhang, J. (2010). Performance of an Electronic Health Record (EHR) Using Cognitive Task Analysis. *International Journal of Medical Informatics*, 79, 501-506.

Saleem, Jason J., Emily S. Patterson, Laura Militello, Shilo Anders, Mercedes Falciglia, Jennifer A. Wissman, Emilie M. Roth, and Steven M. Asch (2007) Impact of Clinical Reminder Redesign on Learnability, Efficiency, Usability, and Workload for Ambulatory Clinic Nurses *J Am Med Inform Assoc*. Sep–Oct; 14(5): 632–640.

Saleem, Jason J., Emily S. Patterson, Laura Militello, Shilo Anders, Mercedes Falciglia, Jennifer A. Wissman, Emilie M. Roth, and Steven M. Asch (2007) Impact of Clinical Reminder

Redesign on Learnability, Efficiency, Usability, and Workload for Ambulatory Clinic Nurses J Am Med Inform Assoc. Sep–Oct; 14(5): 632–640.

Zhang, J. (2005). -centered computing in health information systems: Part I--Analysis and Design (Editorial). *Journal of Biomedical Informatics*, 38, 1-3. ([pdf file](#))

Zhang, J. (2005). -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. L., Gimbel, R., & Zhang, J. (2009). Functional Analysis of Interfaces in U.S. Military Electronic Health Record System using UFuRT Framework. *AMIA Proceedings*.

Zheng, Kai, Rema Padman, Michael P. Johnson, and Herbert S. Diamond (2009) An Interface-driven Analysis of User Interactions with an Electronic Health Records System J Am Med Inform Assoc. Mar–Apr; 16(2): 228–237. doi: 10.1197/jamia.M2852