Defining and Testing EMR Usability:

Principles and Proposed Methods of EMR Usability Evaluation and Rating

HIMSS EHR Usability Task Force

June 2009

HiMSS

CONTENTS

© 2009 Healthcare Information and Management Systems Society (HIMSS)

EXECUTIVE SUMMARY

Electronic medical record (EMR) adoption rates have been slower than expected in the United States, especially in comparison to other industry sectors and other developed countries. A key reason, aside from initial costs and lost productivity during EMR implementation, is lack of efficiency and usability of EMRs currently available. Achieving the healthcare reform goals of broad EMR adoption and "meaningful use" will require that efficiency and usability be effectively addressed at a fundamental level.

We conducted a literature review of usability principles, especially those applicable to EMRs. The key principles identified were simplicity, naturalness, consistency, minimizing cognitive load, efficient interactions, forgiveness and feedback, effective use of language, effective information presentation, and preservation of context.

Usability is often mistakenly equated with user satisfaction, which is an oversimplification. We describe methods of usability evaluation, offering several alternative methods for measuring efficiency and effectiveness, including patient safety. We provide samples of objective, repeatable and cost-efficient test scenarios applicable to evaluating EMR usability as an adjunct to certification, and we discuss rating schema for scoring the results.

INTRODUCTION

The rate at which EMRs have been adopted in clinic and hospital settings within the United States has lagged behind the adoption of information technology that has been common in other industries for more than 20 years.[19]

Multiple causes have been suggested including cost, resistance to change, fear or avoidance of technology, and ingrained patterns of behavior. Increasingly, however, usability has been acknowledged as a deterrent to adoption[40,51,35,23], and one that must be addressed.

In Brief: Adoption rates by physicians and hospitals have been slower than expected in the US, in part, due to poor efficiency and usability. We explore well-established usability principles and testing methods and propose methods to test and rate EMRs for efficiency, effectiveness and safety.

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. Clinicians lose productivity during the training days and for months afterward as they adapt to the new tools and workflow. Some productivity losses are sustained, mostly due to longer time needed for encounter documentation in complex patients[31].

Effective training and implementation methods affect user adoption rates as well, but training is both harder and more costly, and implementation is more complex and difficult when usability is lacking.

It has proved difficult for clinicians to evaluate EMR usability as part of the purchase process for several reasons. Proper evaluation by purchasers requires in-depth study using unfamiliar skills. Most users of one EMR often have not experienced other EMRs, so single product ratings are less helpful than those which compare systems.[9] There has been work done by third-party consulting groups to survey current users, but these results are rarely provided directly by clinical end-users. Industry survey instruments are generally not constructed to provide reliable usability data; they also only provide user satisfaction ratings–a single component of usability.

This paper will:

1. Describe and define usability as it pertains to the EMR.
2. Identify a set of well-established principles of usability and design.
3. Offer potential methods of assessing and rating EMR usability.

We submit that these principles and methods could be used by certification organizations to test and rate products for usability. Requiring this adjunct to certification may spur development of more usable EMR products, and allow decision-makers more confidence in choosing a product that will benefit clinicians.

What is Usability?

Usability is the effectiveness, efficiency and satisfaction with which specific users can achieve a specific set of tasks in a particular environment.[39] In essence, a system with good usability is easy to use and effective. It is intuitive, forgiving of mistakes and allows one to perform necessary tasks quickly, efficiently and with a minimum of mental effort. Tasks which can be performed by the software (such as data retrieval, organization, summary, cross-checking, calculating, etc.) are done in the background, improving accuracy and freeing up the user's cognitive resources for other tasks.

Usability evaluation is far broader than the simple process of measuring user satisfaction. Just as importantly, usability metrics include measures of efficiency, effectiveness, cognitive load and ease of learning. Usability emerges from understanding the needs of the users, using established methods of iterative design, and performing appropriate user testing when needed. There are a wide range of design and evaluation methodologies, both subjective and objective, which are continually growing in sophistication. Built-in webcams on modern laptop PCs, robust wireless networking, remote testing software, and compact, inexpensive video recorders make it increasingly easier to "test" in live clinical settings.

The Challenges of EMR Design

It is particularly challenging to develop excellent usability in EMR systems. There is a wide range of complex information needs, which vary from setting to setting, among different administrative, financial and clinician groups, and from task to task within a group. There are over 50 physician specialties (AMA specialty codes[2]) each with its own software needs, as well as the software needs of other clinical groups such as nurses, pharmacists, physical therapists, respiratory therapists, medical dieticians and others. Each discipline may have several different task scenarios in a working day, with each scenario demanding a different software interface design.

Clinicians are often mobile, going from room to room, hospital to clinic. They seldom give their full attention to the software. Their primary focus should be on the patient, and clinicians are often talking, listening or thinking while using the software. They often have a frequently changing agenda during a single patient workflow, and interruptions are common.

Administrative and financial issues complicate even routine tasks (providing billing codes, discovering drug formulary coverage, pursuing prior authorization) and vary widely with different insurers. There is a burgeoning impetus to measure quality of care, complicated by multiple standards.

It may be challenging for EMR developers to get access to clinician users for feedback or testing. Busy physicians allow only limited access for user-centered design work. Clinicians have other significant constraints that complicate usability evaluations, such as confidentiality concerns in all their encounters, the need to test in the actual work environment, and frequent interruptions in their workflow.

Scope of this Report

We discuss the usability of the EMR from the perspective of clinician users (physicians, nurses, pharmacists, physical therapists, respiratory therapists and others) in the ambulatory, inpatient and acute-care environments. We confined ourselves to issues of user-centered design and usability evaluation. These concepts apply to vendor product development processes, public product usability rating methods and vendor selection criteria for healthcare organizations. In addition, these methods should be applied in the configuration of highly adaptable systems during implementation. We do not otherwise address concerns of implementation, user training or change-management, though these issues do affect user adoption success rates.

Terminology note: An <u>EMR is a computer system</u> composed of multiple, integrated applications enabling clinicians to order, document and store patient information. The term electronic health record (EHR) is sometimes, and incorrectly, used interchangeably. In contrast, an <u>EHR is patient health information</u> from multiple care delivery organizations' EMRs, comprising a patient-centric, longitudinal view of a patient's encounters with healthcare providers.[13] For the purpose of this paper, the term EMR will be used, as we are addressing "systems"– vs. data.

USABILITY PRINCIPLES

In recent years, usability has become an increasingly prevalent topic in the health information technology (HIT) literature and media. Many HIT professionals, healthcare informaticians and researchers have clearly articulated design problems in the current generation of clinical applications.[40,20,8,50] The National Research Council (NRC) has asserted that today's clinical systems provide poor support for the cognitive tasks and workflow of clinicians.[28] These problems can dramatically impact user acceptance and productivity.

In Brief: Well known principles such as simplicity, naturalness, consistency, protection against cognitive overload and others, define good usability. Incorporate deep knowledge of the users' tasks, context and workflows.

Patient safety is a prominent concern in the literature. The Joint Commission (formally known as Joint Commission on Accreditation of Healthcare Organizations) recently issued Sentinel Event Alert 42 regarding technology-related adverse events (The Joint Commission, 2008). This safety alert included EMRs, computer physician order entry (CPOE) and clinical decision support (CDS) systems. They reported that approximately 25 percent of medication errors included in the 2006 Pharmacopeia MEDMARX involved computer technology as a contributing cause. The overwhelming majority of these (82 percent) stemmed from CPOE and other data entry functions. Many studies have documented the issues of alert fatigue, screen fragmentation, terminology confusion and lack of appropriate defaults in CPOE and CDS systems.[8,6,21]

Not all authors discuss these issues directly in terms of usability. Far fewer present these problems in terms of the underlying design principles being violated. It is this level that must be addressed in order to design applications that will achieve desired efficiency, broad usage (a prerequisite for "meaningful use") and avert system-facilitated safety errors.

Experts in usability and Human Factors have published many compilations of principles and guidelines to aid in designing the most effective user interfaces.[30,32,46,49] Most of these lists share certain core ideas. Principles must be evaluated for their importance to the particular context of use. Those that are of key importance to the design of an EMR system are discussed below. These principles were selected for discussion based on their contribution to two essential factors for clinician acceptance and system success:

> 1) Efficiency of use.
> 2) Minimizing likelihood of user error.

User errors have a direct relationship to potential patient safety. User errors may be either errors of commission or errors of omission:[18]

- Example, *errors of commission*:
    - Selecting the wrong patient, wrong medication, wrong dosage or wrong encounter.

- Example, *errors of omission:*
    - Overlooking or misinterpreting key data due to poor information display (e.g., overlooking critically abnormal lab result, or routinely dismissing a critically harmful drug-drug interaction warning).
    - Failing to complete a task (perhaps due to interruption) such as transmitting orders or signing documentation.

Testing methods which measure efficiency, effectiveness, ease of learning and user satisfaction have been developed to take these usability principles into account. To use the methods properly, the principles behind them should be well understood.

Simplicity

Simplicity in design refers to everything from lack of visual clutter and concise information display to inclusion of only functionality that is needed to effectively accomplish tasks. A "less is more" philosophy is appropriate, with emphasis being given to information needed for decision making.[29,26] The more complex an application, the more important this principle becomes. Clinical systems are complex as well as information dense–it is essential for efficiency as well as patient safety that displays are easy to read, that important information stands out, and that function options are straightforward. Simplicity as a principle should not be interpreted as "simple." Clear, clean screen design requires substantially more effort than cluttered displays; it also may mean that some complexity has been removed from the

surface and moved "under the hood." Simplicity applies to any design regardless of the experience level of the target user.

Naturalness

Naturalness refers to how automatically "familiar" and easy to use the application feels to the user. Factors that contribute to this feeling include terminology used in the interface and how well the design and screen flows map to the users tasks and expectations.[30] This is critical to clinical applications since it is extremely difficult to provide extensive training, especially to clinicians whose time is limited and fragmented. Good workflow design can contribute significantly to efficiency and reduce cognitive load. "Natural" workflow can vary dramatically from one specialty to another–or in an acute setting, from one department to another. An Emergency Department's workflow is very different from that of an inpatient medical-surgical unit. Like simplicity, naturalness also contributes to error reduction.

Consistency

External and internal consistency are important to the design of any application. External consistency primarily has to do with how much an application's structure, interactions and behaviors match a user's experience with other software applications. The more a user can apply prior experience to a new system, the lower the learning curve, the more effective their usage, and the fewer their errors. An internally consistent application uses concepts, behavior, appearance and layout consistently throughout.[30,46,49] Predictability is another important factor in enabling efficient use and reducing errors.

Minimizing Cognitive Load

While this principle may sound a bit esoteric, it is essential for a complex, information dense clinical application. Clinicians, in particular, are almost always performing under significant time pressure and in environments bursting with multiple demands for their attention. Combined with the staggering information load faced by today's providers, this can be a recipe for cognitive overload, which could negatively impact patient safety.

Presenting all the information needed for the task at hand reduces cognitive load. For example, when reviewing results of a lipid profile, the provider will want to see the patient's latest and prior results, the medication list, the problem list and allergy list all in the same visual field so that decisions and subsequent actions may be performed without changing screens. Displaying information organized by meaningful relationships is one method of providing cognitive support to the user.[50,28]

An EMR must not only assist with task performance and decision making, but strive for "transparency." In design terms, transparency means that use of the software application does not create too many intrusive thoughts for the user like "How do I …?", "What does this do…?" or "Where is...?" These mental interruptions can cause the user to lose their thought process about the task or decision making process in which they are engaged. In other words, the user should not have to think too much about the application itself.[22]

Cognitive load is increased by any aspects of a design that do not follow the principles of simplicity, naturalness and consistency. It is also increased if a user is required to rely on memory (recall) rather than visual recognition, if a user must try to remember information from one screen to another, what a button really does, or what name something is called as in an "orderables" list. High information density, poor feedback to the user and inadequate cues for data entry fields also affect cognitive load.

Efficient Interactions

One of the most direct ways to facilitate efficient user interactions is to minimize the number of steps it takes to complete tasks and to provide shortcuts for use by frequent and/or experienced users.[46] While this is somewhat stating the obvious, it is included here because of its importance to the user acceptance of a clinical application. Other examples of designing for efficient interactions include auto-tabbing; good default values; large enough list and text boxes to limit scrolling; and preventing the need for frequent switching between keyboard and mouse.[49] Somewhat less obvious factors include attention to minimizing the amount of visual searching required to locate information and the distance the cursor must travel to make selections. Excessive cursor movement and visual scanning both contribute to user fatigue and frustration.

Forgiveness and Feedback

Forgiveness means that a design allows the user to discover it through exploration without fear of disastrous results. This approach accelerates learning while building in protections against unintended consequences.[30,46] This is especially helpful if training is limited. Good feedback to the user supports this goal by informing them about the effects of the actions they are about to take. Campbell et al.[8] provide an analysis of the types of unintended consequences related to CPOE. Forgiveness and feedback work together to reduce user errors and provide graceful recovery when mistakes are made. Good feedback also reassures the user that their actions have had the desired effect. Like consistency, these principles are standard in the design of any application, but of special importance in a clinical information system due to the impact they can have on user errors as well as cognitive load.

Effective Use of Language

All language used in an EMR should be concise and unambiguous. Terminology used also must be that which is familiar and meaningful to the end users in the context of their work; no terms related to computers, technology, HL7, databases, etc. should appear in the user interface. This applies to everything: labels, descriptions, pick lists and error messages.

Text should never be displayed in all upper case; this is considered "shouting." It is more difficult and takes longer to read, and increases perceived density. Even if lists of orderables or terms are received by the EMR in upper case, they should be translated to title case before display in the interface. Rare exceptions include one or two word messages that are intended to draw the attention of the user.[26,49,38]

Abbreviations and acronyms should only be displayed when they are commonly understood and unambiguous.[49,50] Information that must be spelled out but takes more space than available should have ellipses inserted to indicate there is more–with the full text available on mouse-over. This is in part a

patient safety issue. For efficiency, however, a larger number of common abbreviations and synonyms should be available to the user for the purposes of data entry and searching, expanding if necessary for display.

A language issue specific to EMR design is the need to capture structured (discrete) clinical terms from provider documentation such as visit notes, allergies and problem and medication lists. This data is used to identify clinical relationships in patient records, drive decision support functions, eliminate redundant data entry and supply coded data elements to administrative and reporting functions. "Meaningful use" criteria for health reform will likely include requirements for the routine capture of coded clinical data.

The challenge with discrete clinical data entry is the presentation of structured terminology in the user interface. Vocabulary must be efficient to navigate, presented in terms familiar to clinical practice (e.g., instead of billing) and at the appropriate level of granularity. Interface terminology is a complex issue and an active area of Medical Informatics research.[36]

Effective Information Presentation

Appropriate Density

While density of information on a screen is not commonly measured (though it can be), it is a very important concept to be cognizant of when designing EMR screens. In clinical applications, there can be so much relevant information to display it can be tempting to pack as much as possible onto a screen. However, visual search times and user errors increase in proportion to density. It is challenging to balance providing all the necessary information and limit the number of screen changes while maintaining an appropriate screen density. Testing actual users will reveal when the balance has been reached.

Character count, resolution, font, font size and grouping techniques impact visual density. Screen elements such as lines, buttons, controls, scroll bars and icons also contribute to density, which is yet another reason that simplicity is so important.

According to ergonomic recommendations for information presentation on computer screens,[15] an upper limit of 40 percent density is appropriate for character based displays (the percentage of potential character positions actually filled by characters). Graphical user interfaces must be even less due to the other elements contributing to perception of density.

An important means of reducing density is viewing data at a summary level before drilling down to detail. Roughly, the "80/20" rule applies to summary screens–80 percent of the time the information at the summary level is sufficient for decision making and is the most frequently needed information; 20 percent of the time the user will need to delve deeper.

Meaningful Use of Color

Color is one of several attributes of visual communication. It is singled out here for discussion due to how poorly it has been utilized in many system designs to date. Skillful use of color certainly contributes to a user interface that is pleasing in appearance. However, aesthetics should be the last consideration

for using color in any task-oriented application. First and foremost, color should be used to convey meaning to the user. This includes all aspects of information presentation, navigation, differentiation of screen areas and state representation of controls. Everything in the user's task area of the screen, including navigation bars, needs to obey a meaningful color scheme. Purely aesthetic use of color should be limited to design of icons, logos and banner areas.

Simplicity and consistency are both key principles in the use of color. For color to convey meaning, there cannot be a larger number of colors used than the user can remember, and they must be used consistently throughout the application. For instance, if bright yellow is used as a "highlighter" color to emphasize the name of the patient whose orders are currently being entered, then bright yellow should *only* be used as a highlight color for key information. Inconsistent or gratuitous use of color increases the likelihood of user error due to misinterpretation or oversight of important details; the meaning will be lost.

To accommodate users with color-blindness, all meaning conveyed with color must also be differentiated with a second visual mechanism ("redundant encoding") such as font characteristics or fill pattern. For example, if red is used to display critical lab values then the characters should also be bolded, increased in size or some other characteristic. It is highly recommended that displays be designed in grayscale prior to adding color to ensure that all meaning is represented. If not, the inability to differentiate colors also may lead to user errors that have patient safety consequences.

Naturalness is accomplished by adhering to cultural conventions of color meaning. In the United States, the following color interpretations are commonly understood. Comprehensive guidelines on use of color have recently been developed by HFES[15]; see also HHS[49] and Accessibility Forum.[1]

> **Red:** Stop, Hot, Danger, Error, Extreme Warning, Severe Alert, Emergency, Alarm
> **Yellow**: Caution, Potential or Mild Warning, Requires Attention, Slow, Moderate Alert
> **Green**: Go, Safe, Normal, Good, Proceed
> **Blue**: Cold, Advisory

Readability

Screen readability also is a key factor in objectives of efficiency and safety. Clinical users must be able to scan information quickly with high comprehension. The pace and frequent interruptions in clinical workflow guarantee that decisions will sometimes be made based upon cursory screen review. Simplicity, naturalness, language use, density and color all contribute to readability. In addition, guidelines recommend using a font size of no smaller than 12-point for important content and never smaller than 9-point as defaults. Differences in visual acuity make it necessary to allow users to modify text size as needed. System settings for color, fonts and font size should always be respected.[1,26,49,15] San serif fonts can be read more easily in computer displays than serif fonts. High contrast between text and background is also important; black on white is the most readable.

Preservation of Context

Preservation of context is a very important aspect of designing a "transparent" application. In practical terms, this means keeping screen changes and visual interruptions to a minimum during completion of a particular task. Visual interruptions include anything that forces the user to shift visual focus away from the area on the screen where they are currently reading and/or working to address something else, and then re-establish focus afterward. The most frequent violator is the dialog box, which also tends to obscure a significant part of the screen. Dialog boxes should be kept to a minimum. For instance, when a dialog or message box is triggered, it should appear in-context (adjacent to or just below the control that triggered it). This limits visual searching and makes it feel like it is a natural part of the current task. All of these boxes should also be as small as possible without compromising their usability.

Another important guideline associated with preservation of context is that of directness. In part, this is a component of the "what-you-see-is-what-you-get" philosophy–if you change something on the screen, you should see the change immediately and in the format expected. An aspect of directness that sometimes falls through the cracks is to avoid "modes." In data entry, this sometimes occurs in the form of "viewing" vs. "entry" modes; these should not be separate.[30] If a user is viewing information on a form that they have permission to edit, they should be able to do so, in context. This does *not* mean that information collected via a particular form (e.g., allergies) shouldn't be displayed elsewhere in the system as view-only. However, any data presented that is potentially user-editable, should have a mechanism for taking the user directly to the appropriate entry form if updating is desired.

See *ANSI/HFES 200: Human Factors Engineering of Software UserInterfaces*[15], *Windows User Experience Interaction Guidelines*[26] and *Research-Based Web Design & Usability Guidelines*[49] for current, comprehensive guidelines on designing for optimal usability. This last work is unique in that it's inclusion criteria for guidelines was research-based evidence rather than expert opinion (parallel to the "evidence-based medicine" concept). While targeted at Web site design, most of these guidelines apply equally well to Web-based and desktop applications. Each guideline is accompanied by research references and the equivalent of 5-star scores for a) relative importance, and b) strength of the evidence.

Having reviewed the essential principles of EMR usability, we now move to a discussion of evaluation methods.

USABILITY EVALUATION AND RATING METHODS

The people who select and use EMR software are making a commitment that cannot easily be reversed. Costs of implementation are typically high, and the costs of abandoning an implementation or switching to another product are vastly higher. Reliable usability rating schemes offer product purchasers a tool for comparing products before purchase or implementation. These methods can foster competition and innovation by making excellent usability visible to the entire community of purchasers and users.[47]

In Brief: Objective methods are available to evaluate EMR usability. We focus particularly on efficiency, effectiveness AND SAFETY, but also discuss user satisfaction and cognitive workload. A 5-star rating system can help inform EMR purchasers about an EMR's usability.

Organizations evaluate and communicate usability information for a number of reasons. Commercial organizations evaluate usability as part of product improvement, and as a means of differentiating their products from competitors. The government evaluates usability to assure safety standards are not compromised. Independent groups such as *Consumer Reports* provide evaluation data to help consumers make informed purchases. This section will describe two such programs that provide potential EMR rating organizations with perspectives on evaluating usability in products that must maintain high safety standards.

1. National Highway Traffic Safety Administration Child Safety Seat Usability Rating Program

The National Highway Traffic Safety Administration (NHTSA) Ease of Use rating program was designed to create market forces that encourage child seat manufacturers to include user-friendly features, labeling, and instruction manuals and to provide consumers additional data as they make child safety seat purchase decisions. The initial mandate for such a rating program was part of the Transportation Recall Enhancement, Accountability, and Documentation (TREAD) Act in 2000. NHTSA had the requirements that the program must be repeatable and must be objective.[41]

NHTSA's Ease of Use rating program evolved over several years. The chosen approach uses trained evaluators to assign ratings to features considered to impact the usability of installing child safety seats. Weighted averages for each category and overall usability are communicated to the public through the use of a 5-star rating system.[41]

Several key aspects of the NHTSA rating program may provide insight to EMR certification organizations. First, it took two years from the mandate to the first implementation of NHTSA's rating program. Second, NHTSA collected data to evaluate the effectiveness of their program from the initiation of the program. Third, NHTSA recognizes that the Ease of Use rating program needs to be flexible and change as the market changes. In 2008 NHTSA updated their Ease of Use rating program to:

- Make improvements where the initial program was weak.
- Make updates where the initial program had become obsolete.
- Accommodate new safety seat features that did not exist when the rating program was initiated.
- Add a pre-evaluation program to encourage continued innovation.

2. FDA and Human Factors Regulation and Guidelines for Device Manufacturers

The Food and Drug Administration (FDA) requires device manufacturers to follow Human Factors regulations and provides guidance to ensure safe use of medical devices. In 1997 the FDA presented the

final rule for Good Manufacturing Practice. As part of this quality control mandate, medical device manufactures are required to demonstrate adherence to good design and manufacturing processes. The objective of the Human Factors aspect of the regulation is aimed at minimizing user errors that could cause patient injury or death. The regulation calls for design input, design verification and design validation.[17]

The essence of the regulation is that Human Factors activities are to be conducted throughout the design and development of a medical device. Design input calls for manufacturers to establish and maintain procedures that ensure design requirements are appropriate and address the intended use of the device, user needs and patient needs. Design verification requires manufactures to establish and maintain procedures for verifying the design input. Human Factors activities may include task analysis, functional analyses, user studies, prototype tests and mock-up reviews. Design validation requires that the device conforms to the defined user and patient needs, and assures safe use in both intended and unintended uses of the device. A risk analysis aimed at minimizing user error that can lead to patient injury or death must be included as part of design validation.[16]

Key aspects of the FDA's regulation and guidance useful to an EMR certification process include the following: First, the FDA provides guidance, placing the responsibility on the manufacturer to be educated in Human Factors and to select appropriate methods to meet the FDA's regulations. Second, the FDA endorses and requires manufactures to adhere to standards developed by other standards organizations in addition to their own. Third, the FDA holds manufacturers accountable to Human Factors regulations through field inspections, premarket reviews, and post-market surveillance. In each situation the FDA instructs manufacturers to provide evidence for appropriate Human Factors analyses and tests for the product under review.[17]

Evaluation Methods and Metrics

Depending on the reason for testing (e.g., early design vs. differentiation between interfaces) measurement methods differ. Usability experts approach product evaluation as a process. As such, there are specific goals for each phase of the process and there are specific activities appropriate to address phase specific goals.[42] Usability evaluation methods are often described as being primarily "formative" or "summative" in nature. Formative evaluation is used to inform and improve the product design during the development process. Summative usability testing is a validation exercise to evaluate a product at the end of the development process.

Usability is the result of careful design and evaluation throughout product development. During the design and development process, formative usability activities are carried out in support of defining the application, understanding the user and user workflow, and making iterative improvements to the product. The data gathered during these activities tend to be more qualitative and descriptive. The findings from formative usability activities are meant to describe and define users and user needs and product features, as well as have an impact on the design of the product's user interface.

This phase is clearly the responsibility of the software vendor. In addition, these methods should be applied in the configuration of highly adaptable systems during implementation; configuration can involve a high degree of screen "design" and workflow engineering. Formative usability activities include but are not limited to:

- Contextual Inquiry
- Focus Groups
- Stakeholder Meetings
- Affinity Diagramming
- Task Analysis
- Risk Assessment
- Expert Review
- One-on-one usability testing

Later in the development process, summative usability activities are carried out to refine the product. They also may be done after product completion to validate the usability of the product, or compare it with competitor products. Recommended usability rating activities clearly fall after product completion and should be summative in nature. Summative usability activities include but are not limited to:

- Expert Review
- Performance Testing
- Risk Assessment
- One-on-one usability testing

Summative usability activities each have specific goals which they appropriately address. The data gathered during these activities tend to be more quantitative and objective. However, some summative research activities are subjective. Expert reviews as a means to validate usability introduce subjective expert input. The findings from summative activities are meant to validate and confirm usability. If a vendor has employed an iterative user-centered design process through the product development process, there should be few surprises that arise in summative usability testing.

An important tradeoff to consider in any usability evaluation is the testing environment. As soon as the usability evaluation is moved out of the actual environment and into a test environment, much of the complexity caused by the environment is removed. As such, a system that is rated high in usability in a test environment may not be easy to use in the context of the actual environment rife with interruptions and changing work priorities. Usability testing best practices submits that it is always best to conduct formative usability testing in the environment that is closest to the user's actual environment. Software vendors should ensure that their designers and developers have the opportunity to experience their end product in use in a clinical setting. This process could help make tremendous strides toward minimizing "disconnect" between what the user needs an application to do–and what it actually does–or how it does it.

Usability specialists have started developing automated methods for usability testing. Automated usability test tools typically evaluate user interfaces against design requirements. The Web site http://usability.gov provides an automated usability tool to federal agencies (*Automated Usability*). In addition, commercial products are available for automated usability testing of software applications and Web-based solutions.

Evaluating Efficiency

Efficiency, as a test metric, is the speed with which a user can successfully accomplish the task at hand. Research activities aimed at evaluating efficiency includes expert review and efficiency studies. Expert Review is a Human Factors expert review of the product. As part of the review, the usability specialist identifies areas in the product where the product conforms or fails to conform to Human Factors best practices. There is some amount of subjective judgment involved in an expert review.[25]

There are a number of variants on one-one-one usability tests aimed at evaluating efficiency. A typical efficiency study calls for an expert, intermediate or novice user to complete specific key tasks with the application. Performance data is collected. Sessions are frequently recorded with special software that captures interactions with the graphical user interfaces and matches the interactions with time stamps. The results are used to evaluate the efficiency of the product.[25]

The most common measures of efficiency:

- Time to perform a particular task.
- Number of key presses or interactions to achieve task.
- Number of screens visited to complete a specific workflow scenario.
- Number of Back button uses.
- Time to execute a particular set of instructions.

Evaluating Effectiveness

Effectiveness is the accuracy and completeness with which a user can achieve task goals. Risk analysis is a collection of techniques for identifying the most likely human error points in a system. A comprehensive risk analysis will identify, quantify and mitigate risks with iterative assessment and implementation throughout product development. These techniques have been used for many years in numerous industries such as the space program, shipping and nuclear energy. Early on it was learned that human failures were much more difficult to predict than mechanical or electronic components.[24]

The Joint Commission, Veteran's Administration, the FDA, and the Department of Defense have spent a great deal of time and effort developing variants of Failure Modes and Effects Analysis (FMEA) to identify and analyze risks in hospital and other healthcare processes, medical device development and other complex systems. FMEA is one of the most widely used forms of "bottom up" risk analysis and is the prevalent form of risk analysis in the automotive and aviation industries.[43] An FMEA for analyzing human error is a systematic process examining the user's workflow for points where error could occur.

The probability and severity of likely errors are evaluated, and appropriate mitigation for each potential error is identified.

In addition to FMEA, which has a well-developed formalism, a more generic form of risk analysis is a "topological risk analysis." The topological risk analysis is a viable first analysis step to examine the process in enough depth to define its layout (or topology). A topological analysis can help identify a number of risk elements such as "single-point" failures and "common-mode" failures.

A single-point failure would be any action by the clinician that results in harm, injury or death to the patient without a redundant safety check in place. An example would be prescribing a drug the patient is allergic to because drug allergies were not displayed on the ordering screen, and decision support for drug-allergy checking was not yet implemented. A common-mode failure is when multiple actions by the clinician stem from a single cause that results in harm, injury, or death to the patient. For example, a prescribing error caused by forcing the provider to access multiple screens and hold details in memory to complete the prescribing process. As risks are identified, detailed analysis and mitigation efforts should concentrate at these points.[24]

A usability rating process can be developed by adapting risk assessment methodologies to objectively evaluate the potential for user error. Certain design factors can lead to user error which would have patient safety implications. Assessing an EMR user interface for the presence/absence of these design factors provides an important means of evaluating EMR effectiveness. Examples of user errors with patient safety implications are provided in Appendix C. In-depth work on examining how user interface design choices can compromise patient safety is being conducted by the ePrescribing & Common User Interface programs of the National Health Service (NHS) in the U.K. This group has designed a safety-focused usability evaluation method based on "error-traps" similar to the patient safety checklist concept proposed in this document.[11] The NHS also is in the process of developing specific guidelines for safe on-screen display of medication information.

Effectiveness studies are a class of one-on-one usability tests that involve collecting measures of effectiveness when users complete specific key tasks with the application.

Common measures of effectiveness include but are not limited to:

- Number or rate of errors
- Path taken to complete task
- Severity of errors
- Requests for help

Evaluating Ease of Learning

Improving usability has been shown to improve ease of learning or learnability.[37] The more a user can apply prior experience to a new system and the greater the internal consistency (use of consistence concepts, behaviors, layout, etc.) the lower the learning curve. When a system is forgiving of mistakes

and allows discovery through exploration, it fosters faster learning by reducing the user's fear of unintended consequences. Errors, path taken to complete tasks and requests for help each correlate with how well a user knows the system.

Ease of learning can be evaluated in terms of the time it takes the user to reach a specified level of proficiency and in terms of the time it takes a user who has never seen the system interface to successfully accomplish basic tasks. It is important to consider learning throughout the lifetime of use of a product.[33]

Possible measures of ease of learning include but are not limited to:

- Time to achieve expert performance.
- Number of icons remembered after task completion.
- Time spent using manual.
- Time to perform a particular task after a specified period of time away from the product.
- Time to perform task compared to an expert.
- Number of times the Help function is accessed.

For a more complete review of metrics, methodologies, and guidelines regarding usability and ease of learning see Grossman, et al.[14]

Evaluating Cognitive Load

Many methods for measuring cognitive load involve complex testing that require the skills of cognitive psychologists or experienced Human Factors engineers. However, there are a few well developed and validated instruments that are administered as simple questionnaires. Two examples are the NASA Task Load Index (NASA-TLX) and the Subjective Workload Assessment Technique (SWAT). It is possible to simplify these methods such that they may be administered fairly easily.[10] Cognitive load is such an important issue for clinicians that this should be considered for inclusion in an EMR usability rating program.

Evaluating User Satisfaction

The definition of usability typically includes reference to user satisfaction. User satisfaction is a person's subjective response to their interaction with a system. When evaluating usability, satisfaction can be addressed in several ways. A common approach uses Likert-scale questionnaires asking users to rate their satisfaction with various aspects of the product (e.g., on a scale of one to 10). Typically this is done immediately after hands-on usability task performance and at the end of a usability test session. What is weak about this approach is that the method has not been developed under scientific scrutiny. Other researchers use more scientific rating tools such as the System Usability Scale.[7,48,5,44] These scales are stronger because the tools are accompanied by measures of reliability and validity. Research suggests that user satisfaction does not correlate well with other more rigorously obtained measures of usability such as effectiveness and efficiency.[4,12]

We agree that user satisfaction is one component of usability. However, because of the subjective nature of evaluating user satisfaction, we will not provide recommendations concerning measuring user satisfaction as part of a usability rating program. We do highly recommend that EMR purchasers perform structured user satisfaction testing as part of their EMR selection process. Aggregation of user satisfaction data from current EMR end-users would also be of value if collected using appropriate instruments and methodology.

Selecting Tasks for Evaluation

Test tasks should be selected based on how the data are going to be used. When evaluating *efficiency*, tasks focused on user and system efficiency should be considered: that is, tasks that will be frequently performed by users and tasks that are known to be inefficient. When evaluating *effectiveness*, tasks focused on user effectiveness should be considered: that is, tasks that are deemed difficult to perform and tasks that are known to be at risk for user error based on prior evaluations. User observations, workflow analysis and task analysis are methods used to identify frequent and error-prone tasks. Surveys and interviews are methods used to solicit information from users; however, user behaviors should always be directly observed when possible because users are not always accurate in describing what they actually do.

One of the challenges of evaluating usability in EMRs is the complexity of user tasks, workflows and the user environment. Consider the task simply stated as "refill a medication." The actual clinical workflow includes a combination of elements. Specifically, the provider needs to consider the following:

1. Past data points (e.g., medication history, last visit date, relevant lab values, last clinic note)
2. Future data points (e.g., next lab or visit date).
3. Medical evidence personalized for the patient (e.g., what is the goal cholesterol for this patient, how often do labs need to be checked on this medication).
4. Contextual relevance of #1-3: Where is the patient in the lifecycle of this medication (e.g., did they just start it or have they been on it for five years, reason for refill).
5. Task of formally creating/approving the refill.
6. Considerations of cost and formulary coverage, and possible alternative products with better formulary coverage.
7. Communicating with their assistant or the pharmacy.

For the purposes of usability rating, we recommend selecting test tasks that encompass entire workflows ("scenarios"). EMRs are complex systems and usability of complex systems must include the interactions between the information and user sub-tasks that make up the actual work.

Once tasks are selected, "successful completion" criteria (such as error states, deviations and others) must be defined for each task. Examples of test tasks and test scenarios are provided in Appendix A.

5-Star Usability Rating System

The five-star rating system is readily recognized, since it is a common scheme used with consumer products in many commercial Web sites. Development work is needed to define a usability rating system (e.g., 5-star = excellent, 4-star= good, etc.) that can be used to communicate the results of a usability rating program to EMR purchasers.

Consideration must be given to definition of the scale and assigning metric interval cut-offs. As an example, should the rating scale consist of equal intervals (e.g., 5-star is 0 – 59 seconds; 4-star is 60 – 69 seconds, etc.) or should the rating scale consist of intervals based on the normal distribution (e.g., 3-star is the middle 50th percentile).

Another consideration is the granularity of the reporting system. One option is to combine the scores from the tasks in each category (efficiency and effectiveness) and report one global star rating. A second option is to report separate star ratings for each category of measure. Initially this may only mean separate ratings for efficiency and effectiveness; ratings for factors such as cognitive load and ease of learning should also be considered.

The most important aspect of developing a star rating system is defining the benchmark metrics for each measure. For efficiency measures, goal task times should be based on the needs of the clinicians in actual clinical practice. Expectations should be both high and attainable. Appendix B includes benchmark examples.

Certification and EMR Usability Rating

During our survey of the literature, we learned that the Certification Commission for Healthcare Information Technology (CCHIT) was actively considering adding usability as a criterion to its EMR certification process. Subsequently, we contacted them to understand their overall program requirements, and to better understand the needs of certification organizations. CCHIT is a private nonprofit organization with the sole public mission of accelerating the adoption of robust, interoperable health information technology by creating a credible, efficient certification process. They described the following characteristics they found desirable in the development of a usability program:

- The program must be objective.
- The program must be repeatable.
- The program must be cost efficient to implement.
- The program should focus on evaluating efficiency and patient safety.
- The program should evaluate products that are ready for market.
- A good approach would be to rate usability on a scale similar to star programs seen in consumer products.
- Usability rating should be an adjunct to product certification without affecting certification outcome.
- The usability rating system adopted should not be a pass/fail model

© 2009 Healthcare Information and Management Systems Society (HIMSS)

RECOMMENDATIONS

In Brief: We offer specific recommendations for action to rating organizations:

- **Start small.**
- **Develop measurements.**
- **Define the process.**
- **Create a 5-star rating system.**
- **Improve with time.**
- **Encourage others to do their part.**

As the certifying organization implements its usability rating program, based on our research and prior knowledge of usability principles and practice we propose the following approach:

Start small.

Begin usability rating with a focus on simple efficiency and effectiveness measures, including some initial patient safety checkpoints. Don't let it become a multi-year process to set up an initial program. Learn from and adapt Human Factors processes established by the FDA, NHTSA and other organizations.

Develop measurements.

Devise objective measures of efficiency that takes into account time on task and number of user interactions. Develop objective measures of effectiveness that takes into account system characteristics that impact patient safety. We recommend initially testing effectiveness using a pre-defined checklist of system interactions that have patient safety implications if not handled well by the system. Appendix C illustrates patient safety checklist examples; also see Fone and Lewis.[11]

Create a 5-star rating system.

Define the usability rating system using 5-stars based on an absolute standard against benchmarks. An absolute standard allows each product to stand on its own and demonstrate progress over time rather than in comparison to other products. This requires that benchmarks be established in advance defining target scores for each measurement. Benchmarks should reflect user needs in actual clinical practice. See Appendix B for sample benchmarks.

Initially, report star ratings on a few measures. In subsequent years, additional measures can be added.

Define the process.

Select a set of test tasks for evaluating efficiency that a) are frequently performed by providers; b) are at risk of being inefficient; and c) allow evaluation of tasks and workflow. Begin with simple (but carefully planned) scenario-based user testing similar to the "discount usability engineering" methods described by Nielsen.[30] Recall, however, that the processes should be engineered to be summative in nature.

For efficiency evaluation we recommend a multi-step approach:

1) The vendor does a walk-through explanation to the rating organization's selected intermediate or expert clinical users.

2) The vendor performs the test tasks; task times are recorded as measures of expert performance.

3) The rating organization's selected users perform test tasks for "intermediate user" measurement. As much as possible, the group of user test participants should represent a mixture of role types (e.g., physician, nurse, medical assistant, physical therapist) performing tasks appropriate to their role. The rating organization's selected users should be experienced in two or more EMRs, but not the one being tested. The number of user participants necessary to produce meaningful results will need to be evaluated. In general, summative testing requires more participants than formative testing.

For effectiveness evaluation, an evaluator will need to determine the presence or absence of patient safety items from the checklist developed. See Appendix C for example effectiveness tests for patient safety.

Improve with time.

Usability rating programs should evolve in sophistication over a multi-year period. Eventually they should include measures of cognitive load and ease of learning. In future years, consider also testing naïve users. Evaluate the effectiveness of the program itself on an annual basis. Update scenarios, tasks, methods and measures to reflect any needed improvements as well as evolution in the EMR marketplace and usability best practices.

It may be helpful to the consumer to break the star system into categories as it becomes more complex, e.g., a product may score 4 stars for efficiency, 3 stars for effectiveness and 3 stars for ease of learning. Due to its subjective nature, we recommend that user satisfaction be left for potential customers and third parties to evaluate.

Encourage others to do their part.

Encourage vendors to utilize iterative design with formative user-based research throughout the design and development process with summative usability evaluation before launch. At the same time educate clinical decision-makers to assess EMR usability as part of their EMR purchase and system configuration processes.

REFERENCES

1. Accessibility Forum. *Quick Reference Guide to Section 508 Resource Documents.* September 15, 2003. Accessed June 8, 2009. Available at: www.accessibilityforum.org/paper_tool.html.

2. American Medical Association. AMA specialty codes. (n.d.). Accessed April 23, 2009. Available at: www.ehealthlist.com/ama_specabbrev.asp.

3. Usability.gov. Automated Usability Test Environment (UTE) Tool Fact Sheet. (n.d.). Accessed June 7, 2009. Available at: www.usability.gov/refine/UTEfactsheet.html.

4. Bailey B. *Getting the Complete Picture with Usability Testing*. March 2006. Accessed June 7, 2009. Available at: www.usability.gov/pubs/030106news.html.

5. Bangor A, Kortum PT, Miller JA. An empirical evaluation of the System Usability Scale (SUS). *Int J Hu-Comp Interact.* 2008;24(6):574-594.

6. Bates DW, Kuperman GJ, Wang S, Gandhi T, Kittler A, Volk L et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. *JAMIA. 2003;10*(6):523-530.

7. Brooke J. SUS: a 'quick and dirty' usability scale. In Jordan PW, Thomas P, Weerdmeester BA, McClelland AL (Eds.). *Usability Evaluation in Industry.* London: Taylor and Francis. 1996.

8. Campbell EM, Sittig DF, Ash JS, Guappone KP, Dykstra RH. Types of unintended consequences related to computerized provider order entry. *JAMIA.* 2006;*13*(5):547–556.

9. Edsall RL, Adler KG. User satisfaction with EHRs: report of a survey of 422 family physicians. *Fam Prace Manage.* 2008;*15:*25-32.

10. Embrey D, Blackett C, Marsden P, Peachey J. Development of a Human Cognitive Workload Assessment Tool. July 2006. Accessed June 8, 2009. Available at: www.mcga.gov.uk/c4mca/research_report_546.pdf.

11. Fone J, Lewis K. Safety & usability—methods for evaluating the user interface. April 28, 2009. Accessed June 19, 2009. Available at: www.mit.edu/~juhan/nhs_public/Safety-focused-usability-evaluation_28-Apr-2009_Kit-Lewis_for_HC_v2.pdf.

12. Frøkjær E, Hertzum M, Hornbæk K. Measuring usability: are effectiveness, efficiency, and satisfaction really correlated? *Proceedings of the SIGCHI conference on Human Factors in computing systems*, (pp. 345 – 352). New York, NY, USA. 2000.

13. Garets D, David M. *Electronic medical records vs. electronic health records: yes, there is a difference.* HIMSS Analytics. January 26, 2006. Accessed June 17, 2009. Available at: www.himssanalytics.org/docs/WP_EMR_EHR.pdf.

14. Grossman T, Fitzmaurice G, Attar R. A survey of software learnability: metrics, methodologies and guidelines. Proceedings of the 27th international Conference on Human Factors in Computing Systems (Boston, MA, USA, April 04 - 09, 2009) (pp. 649-658). New York, NY: CHI '09. ACM.

15. Human Factors and Ergonomics Society. 2008. *ANSI/HFES 200:Human Factors Engineering of Software UserInterfaces.* Santa Monica, CA: Human Factors and Ergonomics Society.

16. *Human Factors Implications of the New GMP Rule Overall Requirements of the New Quality System Regulation*. 2009. Accessed June 6, 2009. Available at: www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/PostmarketRequirements/Human Factors/ucm119215.htm.

17. Information for Manufacturers and Distributors. 2009. Accessed June 7, 2009. Available at: www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/PostmarketRequirements/Human Factors/ucm119190.htm#1.

18. Institute of Medicine. *Patient Safety: Achieving a new standard for care.* 2004. (Aspden P, Corrigan JM, Wolcott J, Erickson SM, Eds.) Washington, D.C.: The National Academies Press.

19. Jha AK, DesRoches CM, Campbell EG, Donelaln K, Rao SR, Ferris TG, et al. Use of electronic health records in U.S. hospitals. *N Engl J Med.* 2009;360: 1628-1638.

20. Khajouei R, Jaspers MW CPOE system design aspects and their qualitative effect on usability. *Studies in Health Technology Informatics.* 2008;*136:*309–14.

21. Krall MA, Sittig DF. (2002). Clinician's assessments of outpatient electronic medical record alert and reminder usability and usefulness requirements. *AMIA 2002 Symposium Proceedings*, (pp. 400-404).

22. Krug S. (2006). *Don't Make Me Think!* (2nd ed.). Berkeley, CA: New Riders.

23. Kushniruk A, Borycki E, Anderson JG, Anderson MM. (2008). Combining two forms of simulation to predict the potential impact of interface design on technology-induced error in healthcare. *Proceedings of the 2008 Spring simulation multiconference* (pp. 497-504). San Diego, CA: The Society for Computer Simulation, International.

24. Maddox ME, Barnes HJ. (2004). Risky Business: Analysis Tools for Medical Devices and Processes. *Workshop presented at the 48th Meeting of the Human Factors and Ergonomics Society.* New Orleans, LA.

25. Mayhew DJ. Keystroke Level Modeling as a Cost Justification Tool. In Bias RG, Mayhew DJ (Eds.), *Cost-Justifying Usability, An Update for the Internet Age.* San Francisco, CA: Morgan Kaufmann Publishers. 2005.

26. Microsoft Corporation. *Windows User Experience Interaction Guidelines.* November 6, 2008. Accessed June 8, 2009. Available at: msdn.microsoft.com/en-us/library/aa511258.aspx.

27. National Institute of Standards and Technology. *Common Industry Specification for Usability - Requirements (NISTIR 7432).* June2007. Accessed June 7, 2009. Available at: zing.ncsl.nist.gov/iusr/documents/CISU-R-IR7432.pdf.

28. National Research Council. (2009). *Computational Technology for Effective Health Care: Immediate steps and strategic directions.* (Stead WW, Lin HS, Eds.) Washington, D.C.: The National Academies Press.

29. Nielsen J. *Designing Web Usability: The Practice of Simplicity.* Indianapolis: New Riders Publishing. 1999.

30. Nielsen J.. *Usability Engineering.* Chestnut Hill, MA: Academic Press. 1993.

31. Poissant L, Pereira J, Tamblyn R, Kawasumi Y. The impact of electronic health records on time efficiency of physicians and nurses: a systematic review. *JAMIA.* 2005;12:505-516.

32. Principles for Usable Design. (n.d.). Accessed June 8, 2009. Available at: www.usabilitybok.org/design/p287.

33. Quesenbery W. (2001). What Does Usability Mean: Looking Beyond 'Ease of Use'. *Proceedings of the 48th Annual Conference, Society for Technical Communication* . Available at: www.wqusability.com/articles/more-than-ease-of-use.html.

35. Rose AF, Schnipper JL, Park ER, Poon EG, Li Q, Middleton B. Using qualitative studies to improve the usability of an EMR. *J Bio Inform.* 2005;*38:*51-60.

36. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. A model for evaluating interface terminologies. *JAMIA,* 2008;*15:*65-76.

37. Saleem JJ, Patterson ES, Militello L, Asch SM, Doebbeling BN, Render ML. (2007). Using Human Factors Methods to Design a New Interface for an Electronic Medical Record. *AMIA 2007 Symposium Proceedings*, (pp. 640-644).

38. Sanders MS, McCormick EJ. *Human Factors in Engineering and Design.* New York, NY: McGraw-Hill. 1993.

39. Schoeffel R. The concept of product usability. *ISO Bulletin.* 2003;34:6-7.

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

41. Smith CS, Rockwell TE, Collins LA, Kim SD, Park BT. (2009). *NHTSA'S child safety seat usabilityrating program.* Accessed June 8, 2009. Available at: www.lifesaversconference.org/handouts2009/Smith2.pdf.

42. Step-by-Step Usability Guide. (n.d.). Accessed June 7, 2009. Available at: www.usability.gov/.

43. Stephans RA. *System Safety for the 21st Century: The Updated and Revised Edition of System Safety 2000.* Hoboken, NJ: John Wiley & Sons, Inc. 2004.

44. System Usability Scale. (2009). Accessed June 8, 2009. Available at: http://en.wikipedia.org/wiki/System_Usability_Scale_(SUS).

45. The Joint Commission. *Safely implementing health information and converging technologies*. December 11. 2008. Available at: www.jointcommission.org/SentinelEvents/SentinelEventAlert/sea_42.htm.

46. Tognazzini B. (n.d.). *First Principles of Interaction Design*. Accessed June 6, 2009. Available at: www.asktog.com/basics/firstPrinciples.html.

47. Tullis, T. S., & Albert, W. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics.* Burlington, MA: Morgan Kaufmann. 2008.

48. Tullis TS, Stetson JN. (2004). A Comparison of Questionnaires for Assessing Website Usability. *Usability Professional Association Conference Proceedings.*

49. U.S. Department of Health and Human Services. (2006). *Research-Based Web Design & Usability Guidelines.* Accessed June 8, 2009. Available at: www.usability.gov/pdfs/guidelines.html.

50. Walker J. Usability. In Walker JM, Bieber EJ, Richards F. (Eds.), *Implementing an Electronic Health Record System* (pp. 47-59). Springer. 2005.

51. Zheng K, Padman R, Johnson MP, Diamond HS. An interface-driven analysis of user interactions with an electronic health records system. *JAMIA.* 2009;16:228-237.

CONTRIBUTORS

Authors

Jeffery L. Belden, MD – Chair, HIMSS EHR Usability Taskforce

Rebecca Grayson

Janey Barnes, PhD

Usability Principles Workgroup

Patricia Alafaireet, MHA

Janey Barnes, PhD

Jeffery L. Belden, MD – Chair, HIMSS EHR Usability Taskforce

Edna Boone, MASS, CPHIMS – HIMSS Senior Staff Liaison

Jon Duke, MD

Rebecca Grayson

Andrew Hutson

Jasmin Phua

Tiana Thomas – Leadership Council, and Co-chair EHR Usability Principles Workgroup

Juhan Sonin

Penn White, MD - Leadership Council and Co-chair EHR Usability Principles Workgroup

Editors

Melanie Brodnik, PhD, RHIA – Vice-Chair, HIMSS EHR Usability Taskforce

Helen Volger, MSHA, CPHIMS

In Acknowledgement:

We would like to thank the many HIMSS EHR Usability Task Force members and others who provided their time and expertise for review and comment on the final drafts. Appreciation is also extended to the TIGER Initiative for permitting us use of the compendium of material organized by their Usability and Clinical Application Design Workgroup. This document is the synthesis of numerous experienced professionals in the Healthcare IT and Human Factors communities.

HIMSS EHR Usability Taskforce

Patricia Alafaireet, MHA

Cecilia Backman, MBA, RHIA, CPHQ - Leadership Council

Janey Barnes, PhD

Jeffery L. Belden, MD - Chair

Lyle Berkowitz, MD, FHIMSS

Judi Binderman, MD, MBA

Edna Boone, MASS, CPHIMS – HIMSS Senior Staff Liaison

Melanie Brodnik, PhD, RHIA - Vice-Chair, HIMSS EHR Usability Taskforce

Jon Duke, MD

Robert Duthe, MBA - Leadership Council

Rebecca Grayson

Doron Gutkind

Shannon Houser, PhD, MPH, RHIA

Andrew Hutson

Ron Ribitzky, MD - Leadership Council

Juhan Sonin

MaryAnne Sterling - Leadership Council and Co-chair EHR Usability User Pain Points Workgroup

Sandra Stork

Carolyn Swanson - Leadership Council and Co-chair EHR Usability User Pain Points Workgroup

Tiana Thomas - Leadership Council and Co-chair EHR Usability Principles Workgroup

Juanita Threat - HIMSS Coordinator

Royce Uehara

Michael Van Ornum, RPh, RN, BCPS – Vice-Chair, HIMSS EHR Usability Taskforce

Albert Villari, MD - Leadership Council and Chair EHR Usability Vendor Best Practices Workgroup

Helen Volger, MSHA, CPHIMS

Penn White, MD - Leadership Council and Co-chair EHR Usability Principles Workgroup

APPENDICES

A. Test task and scenario examples

Top level: Clinical Scenarios – These are entire workflows consisting of a number of associated component test tasks. These scenarios are sufficiently complex to represent a clinician workflow worthy of testing.

Next level: Test Tasks – These are component tasks that occur frequently in clinical settings, or are tasks that are at risk for user error. Individually, they would be too simple to constitute a test workflow.

This set of examples is not meant to be exhaustive, but to serve as a starting point for types of scenarios and tasks that might be a part of usability testing.

Table A1. Clinical Scenarios

These are entire workflows consisting of a number of associated component test tasks. These scenarios are sufficiently complex to represent a clinician workflow worthy of testing.

				Design Principles										
				Simplicity	Naturalness	Consistency	Minimizing Cognitive Load	Efficient Interactions	Forgiveness & Feedback	Effective use of Language	Appropriate Density	Meaningful use of Color	Readability	Preservation of Context
No.	Scenario Name	Scenario Description	Features / Rationale	A	B	C	D	E	F	G	H	I	J	K
1	Acute UTI with "hey-doc" rash	New patient presents with 3 days of dysuria, hematuria, urgency and frequency. No fever, chills, or back pain. Later on, she remembers that she has an itchy rash between the toes of her left foot. Doctor orders Bactrim DS 1 tab bid x 3 days, phenazopyridine 200 mg tid after meals and recommends OTC terbinafine cream to apply bid x 10 days.	"Oh by the way" complaint not included in the initial reason for visit. Demonstrates how program handles multi-complaint visits and ease of charting a last minute addition.	+				+	+	+			+	

#	Scenario	Description	Considerations											
2	Chronic complex diabetic with LDL elevation	Chronic patient with HTN, Obesity, Type 2 Diabetes and elevated LDL comes in for a recheck of his weight and diabetes. Doctor wants fasting BS (in office), Lipid panel & HbA1c (sent out), VS including weight, diabetic foot exam, and intervening history before seeing patient. At end of visit, doctor increases glipizide from 5 mg bid to 10 mg bid.	Care for chronic disease. How does program handle instructions to staff before seeing the patient? How easily are outside lab orders handled? How efficiently are medication orders changed? Does system offer decision support for target LDL, aspirin therapy indications, reminders for periodic testing and immunizations?	+			+	+				+		+
3	Medication refill request	Respond to a medication refill request. Check medication history, patient problem list, drug prescribing information and lab tracking studies recommended for this medication.	Does design of display provide necessary information in a terse, aggregated fashion?	+	+		+	+	+	+	+			+
4	Depression initial visit	Established 53 yr old male with 3 months of depression symptoms. Not suicidal or psychotic. Order lab tests to look for medical causes of depression. Initiate treatment with SSRI. Print out a patient education handout for the patient.	Order TSH, CMP or BMP. Efficient access to patient education materials, and way-finding to the proper handout. Linking the EHR diagnosis to the Patient Education resource would save time.	+	+	+		+			+		+	+

© 2009 Healthcare Information and Management Systems Society (HIMSS)

5	Lab result letter & orders	Send a letter (or email) to the patient reporting on her abnormal thyroid test result, order thyroid medicine, and schedule repeat testing for six weeks from now.	Effective use of language appropriate for the patient receiving the communication. Page layout in the letter that makes communication effective. Clinician efficiency. Simplified data display for clinician. Non-intrusive decision support for selecting the proper thyroid test. How is receipt of the notification to the patient verified? What happens if the patient misses her retest?		+		+	+		+	+		+	

Table A2. Test Tasks.

These are component tasks that occur frequently in clinical settings, or are tasks that are at risk for user error. Individually, they would be too simple to constitute a test workflow.

No.	Task Name	Task Description	Features / Rationale	Simplicity	Naturalness	Consistency	Minimizing Cognitive Load	Efficient Interactions	Forgiveness & Feedback	Effective of language	Appropriate Density	Meaningful use of color	Readability	Preservation of Context
			Design Principles	A	B	C	D	E	F	G	H	I	J	K
1	Find LDL	Find the patient's latest LDL result.	Don't make the clinician calculate the LDL result.				+	+				+		
2	Count CAD risk factors	How many coronary artery disease risk factors does the patient have?	Does the system aggregate risk-factor data and present it concisely and appropriately for the task at hand? Reduce cognitive load. Simple data presentation.	+			+	+					+	+

#	Task	Question / Scenario	Heuristic / Rationale											
3	CAD risk score	What is the patient's risk of having a coronary-disease related event in the next 10 years?	Does the system aggregate risk-factor data and present it concisely and appropriately for the task at hand? Presenting Framingham risk score can guide clinician in making decisions about lipid-reduction therapy.	+	+		+	+		+		+	+	+
4	Change default pharmacy	How do you change a patient's pharmacy of choice? What happens if the new pharmacy is not approved by the patient's insurance plan?	Forgiveness and Feedback in event of error in data entry. Appropriate system defaults.	+	+	+			+	+				+
5	Drug-interaction alert & response	Prescribing new drug brings up drug interaction warning. Physician reviews warning, completes prescription order, and makes change in default setting for DI severity level threshold.	Avoids alert fatigue. Patient safety is at stake. Is information terse and actionable? Is severity threshold easily adjustable?	+	+	+	+	+	+	+	+	+	+	

Table A3. Potential Tasks that Need Additional Work to be considered a Test Task.

These are too vague to be component tasks, or have components that have no clinical consensus as to appropriate clinical response.

No.	Task Name	Task Description	Rationale for non-inclusion	Simplicity A	Naturalness B	Consistency C	Minimizing Cognitive Load D	Efficient Interactions E	Forgiveness & Feedback F	Effective of language G	Appropriate Density H	Meaningful use of color I	Readability J	Preservation of Context K
1	New orders.	Place new orders on a complex patient	This task needs more detail in order to define the task as a test task.	+	+		+	+	+					
2	Manage hypertension	Review your plan for managing the patient's hypertension.	This task is too vague. In order to be a test task, the task would need to have more detail.					+	+	+				
3	What kind of penicillin allergy?	What is the nature of the patient's penicillin allergy?	This task is too simple as stated and needs context regarding the need for the information and how the information is going to be used in a clinical decision.	+	+	+		+		+			+	+

| 4 | Orthopedic consult order | Order an orthopedics consult, with appropriate pre-visit testing. | This task needs more detail as there is not a standard approach for ordering consults. Should the clinician order the MRI of the knee, or let the orthopedic surgeon decide if it is needed? | + | + | | | + | | | | | | |

B. Benchmark examples

For usability ratings to offer more than comparative information, usability benchmark metrics need to be established. Benchmark metrics can be developed by measuring clinician users in actual clinical environments performing each task or scenario. Clinician user panels would then compare the best <u>actual</u> performances against user's perceived <u>ideal</u> performance, in order to develop a target score that better reflects actual user needs, as opposed to the current state of the art EMR performance. Some target criteria would be more straightforward, as either present or absent features.

See National Institute of Standards and Technology (2007) for a detailed description for benchmarking usability criteria.

Table 1 presents examples of target criteria for measuring effectiveness tied to patient safety.

Table 1. Target Criteria for Evaluating Patient Safety.

Patient Safety Checklist	*Effectiveness: Pass/Fail of Patient Safety Item*
Medication list displayed in Tallman lettering	Pass / Fail
Patient's drug allergies displayed on medication ordering screen	Pass / Fail

Table 2 presents examples of target criteria for measuring the efficiency of an EMR. In this example efficiency is defined as the average time for test participants to complete each specific task or scenario. User time and system response time (e.g., download times) should be included in the task time.

Table 2. Target Criteria for Measuring Efficiency.

Task or Scenario	*Efficiency: Maximum acceptable task time*
Scenario 1. Acute UTI with "hey-doc" rash	\<n> minutes
Scenario 2. Chronic complex diabetic with LDL elevation	\<n> minutes

C. Patient safety checklist examples

This list provides samples of the types of design factors that could lead to user errors which would have patient safety implications. They are intended to be straightforward Pass/Fail tests. The resulting effectiveness measure could simply be how many of the chosen tests the product passed (e.g. "8 out of 10"). Opportunities for checking some of these may need to be built into testing scenarios, but most can be scenario-independent.

These "tests" focus on prevention of the following user errors or practices:

1.   Selection of the wrong patient or patient encounter.
2.   Selection of the wrong medication or dosage.
3.   Stepping away from a terminal without logging out or suspending the session.
4.   Stepping up to a terminal and taking action within someone else's active session.
5.   Overlooking or being unaware of critical patient information.

Patient Selection and Identification

- Patient's full name, unique ID, age (or DOB) and gender are prominently displayed on all chart screens.
- Patient's full name, unique ID, DOB and gender is the minimum set of identifiers displayed when selecting a patient to access their chart.

CPOE/CDS/ePrescribing

- Patient's drug allergies are displayed on the medication ordering screen.
- Patient's current medications are displayed on a single screen.
- Viewing of the patient's current medication list is at most one click away from the medication ordering screen.
- Actions to renew, discontinue or cancel are done directly from the current medication list (i.e. the user is not required to reselect drug from a pick-list).
- When ordering, the selected drug provides information on standard dosing, dosing range and appropriate field defaults.
- Similar drug names are differentiated using Tall Man lettering according to FDA recommendations.
- Orders are displayed in a list format as they are created, and may be reviewed and edited prior to transmission to the appropriate ancillaries/departments for processing.
- All elements of a medication order are included on the screen prior to ordering.

Log In and Log Out

- A single action will log out or suspend the user's session and bring up the Log In screen for the next user.
- All screens display the name of logged in user.

Information Display

- Abnormal results are readily differentiated visually from normal results using at LEAST two methods (i.e. "redundant encoding"). Methods may include meaningful use of color, change in typography (e.g. bold or larger font size), use of iconography or other innovative means.
- Patients who have *new* (unacknowledged) abnormal results are visually differentiated at the "desktop" function level (Provider Inbox, Census display, etc.) without having to open their chart to check.
- Patients requiring isolation have a *unique* (i.e. not same method as abnormal results) visual differentiation with redundant encoding.
- Results are never displayed without a normal range (if there is one) visible or readily available (e.g. on mouse-over), including on Trends or Graphs.

Documentation

- If the patient has more than one "open encounter" (documentation started but not yet signed off), it is straightforward for the user to identify and open the correct one for additions or completion.
- During documentation, the patient's current problem and medication lists are at most one click away for viewing and/or inserting elements into the note.

General

- Error messages explain the error in user-understandable terms plus describe steps necessary to recover from the error.
- Error message choices are straightforward, describing the user action directly. Button labels describe the resulting action directly rather than unclear Yes/No choices.
- Error messages allow the user to recover without data loss or data entry loss.

D. Usability Principles Workgroup Bios

Patricia Alafaireet, MHA - Ms. Alafaireet is the Director of Applied Health Informatics, working with HMI Group consulting. She also holds a clinical faculty appointment. She is currently pursuing a PhD in applied health informatics. Her research area of expertise lies in graphical user interface aesthetics for physician use and in the visual representation of data specifically designed to support physician practice.

Janey Barnes, PhD - PhD in Cognitive Psychology from UMass, Amherst (1992). Currently principle of User-View, Inc. 15 years as Human Factors consultant serving diverse domains including healthcare, transportation, telecommunications, and financial. User centered design experience with multiple hospital and ambulatory electronic medical records and patient health portals. A current member of HIMSS, HFES, and Triangle UPA (Usability Professionals Association).

Jeffery L. Belden, MD – Chair HIMSS EHR Usability Task Force - Practicing family physician with 30 years experience in clinical practice, 15 years of using four different EHRs. Past experiences in photography, film-making, layout and design, typography, and consulting in healthcare software design inform his approach to user-centered design. Faculty at the University of Missouri, Department of Family and Community Medicine, and on the affiliated faculty at the Information Experience Lab at the School of Information Sciences and Learning Technologies. Responsibilities include user training, implementations, collaboration in product development with Cerner on their ambulatory EHR, and collaboration with human-computer interaction colleagues in the IE Lab.

Edna Boone, MASS, CPHIMS – HIMSS Senior Staff Liaison with 20+ years health information technology experience in a community healthcare network setting. Experience includes planning, product selection, implementation and training for inpatient, ambulatory, patient health portals and health information exchange with over ten different vendor product lines.

Melanie Brodnik, PhD, RHIA - Dr. Brodnik is the Director of the undergraduate division in Health Information Management and Systems and the coordinator of the master's degree program in Health Informatics at The Ohio State University in Columbus, Ohio. She has 35+ years experience as an educator and practitioner in the field. She has been a member of HIMSS and AMIA for over 15+ years and in 2004 served as President of AHIMA. She has held numerous volunteer positions and has delivered numerous presentations at the state and national level. She has published in the *JAHIMA*, the *Journal of Health Information Management Research, Topics in Health Information Management* and other related

journals. Dr. Brodnik currently serves as managing editor of the legal textbook *Fundamentals of Law in Health Informatics and Information Management.*

Jon Duke, MD -Internist, Medical Informatics Fellow at Regenstrief Institute, getting M.S. in Human Computer Interaction at Indiana University. Has worked with two inpatient CPOE's and one outpatient EMR over the past 5 years. Currently involved in the development of the next generation system at Regenstrief.

Rebecca Grayson - Independent HIT consultant with 20 years experience primarily in the Ambulatory EMR domain, specializing in User Experience. Includes nine years leading requirements, design and usability for Kaiser Permanente EHR projects. Other significant work includes design and development of EHR/EMR systems for UCSF and Marquette General Healthcare System. Academic training in Medical Informatics and Human Factors. AMIA member since 1988; also current member of HIMSS, ACM and UPA (Usability Professionals Association).

Andrew Hutson - Graduate Research Assistant. 2010 Candidate for a Masters in Health Administration and a Master of Science in Health Informatics at the University of Missouri. 2 yrs experience working with Cerner's PowerChart.

Juhan Sonin - Juhan is the Creative Director of Involution Studios Boston, and has been the creative leader of four different organizations, producing work recognized by the BBC, the New York Times, Ars Electronica, National Public Radio and Billboard Magazine. Juhan has previously spent time at Apple, the National Center for Supercomputing Applications (NCSA), a handful of startups, and MITRE. He is also a lecturer on design and rapid prototyping at the Massachusetts Institute of Technology (MIT).

Jasmin Phua – Independent user experience and business process consultant with 10 years of experience designing systems across a broad-range of user interfaces; 6 years focused on the design of electronic medical records systems for the Social Security Administration's disability adjudication process while at Lockheed Martin.

Tiana Thomas – Co-lead for this document. Software development experience with 9 years spent on Cardiovascular Information Systems; 6 years focused on Technical Product Management of our suite of integration products including interfacing, IHE and clinical data registries. This includes leading the team through the full software development life cycle of requirements gathering, design, development and deployment.

Helen Volger, MSHA, CPHIMS - Ms. Volger contributes over twenty years experience with Healthcare Information Systems, including work with physician group practices, software vendors, insurance carriers, consulting firms and hospitals. Her background includes project management experience for strategic initiatives, as well as managing business and clinical application implementations and on-going support. She is a current Board Member of the

Virginia Chapter of HIMSS and a past Board Member for an HIS vendor's national user group. Her education includes a Bachelor of Science in Computer Science from the University of Pittsburgh. She recently earned a Master of Science in Health Administration through Virginia Commonwealth University and was an A.D. Williams Award recipient.

Penn White, MD, MSIDC – Co-lead for this document. Practicing clinician for 25 years, Master of Science in Information Design & Communications. Twenty years experience in medical informatics. Currently independent Clinical Practice HIT Consultant for ambulatory care providers.

The Technology Informatics Guiding Educational Reform (**TIGER**) **Initiative** aims to enable practicing nurses and nursing students to fully engage in the unfolding digital electronic era in healthcare by identifying best practices and effective technology capabilities for nurses. TIGER's goal is to create and disseminate action plans that can be duplicated within nursing and other multidisciplinary healthcare training and workplace settings. The TIGER Initiative represents a relationship between the Alliance for Nursing Informatics (ANI), with its 20 nursing informatics professional societies, the American Nurses Association (ANA), the Association of Nurse Executives (AONE), the American Association of Colleges of Nursing (AACN), the HIMSS nursing community and other associations, collectively representing over 2,000,000 nurses.