Journal of Biomedical Informatics 78 (2018) 123–133

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Commentary

Ten factors to consider when developing usability scenarios and tasks for health information technology

Alissa L. Russ[a,b,c,*], Jason J. Saleem[d,e]

[a] Richard L. Roudebush Veterans Affairs (VA) Medical Center, Department of Veterans Affairs, VA Health Services Research and Development Service (VA HSR&D), Center for Health Information and Communication, United States
[b] Department of Pharmacy Practice, College of Pharmacy, Purdue University, West Lafayette, IN, United States
[c] Regenstrief Institute, Inc., Indianapolis, IN, United States
[d] Department of Industrial Engineering, J.B. Speed School of Engineering, University of Louisville, Louisville, KY, United States
[e] Center for Ergonomics, University of Louisville, Louisville, KY, United States

ARTICLE INFO

Keywords:
Usability
Human factors
Health information technology
Human-computer interaction
Scenarios
Tasks

ABSTRACT

The quality of usability testing is highly dependent upon the associated usability scenarios. To promote usability testing as part of electronic health record (EHR) certification, the Office of the National Coordinator (ONC) for Health Information Technology requires that vendors test specific capabilities of EHRs with clinical end-users and report their usability testing process – including the test scenarios used – along with the results. The ONC outlines basic expectations for usability testing, but there is little guidance in usability texts or scientific literature on how to develop usability scenarios for healthcare applications. The objective of this article is to outline key factors to consider when developing usability scenarios and tasks to evaluate computer-interface based health information technologies. To achieve this goal, we draw upon a decade of our experience conducting usability tests with a variety of healthcare applications and a wide range of end-users, to include healthcare professionals as well as patients. We discuss 10 key factors that influence scenario development: objectives of usability testing; roles of end-user(s); target performance goals; evaluation time constraints; clinical focus; fidelity; scenario-related bias and confounders; embedded probes; minimize risks to end-users; and healthcare related outcome measures. For each factor, we present an illustrative example. This article is intended to aid usability researchers and practitioners in their efforts to advance health information technologies. The article provides broad guidance on usability scenario development and can be applied to a wide range of clinical information systems and applications.

1. Introduction

Health information technology (HIT) interfaces have a long history of insufficient usability, which can dramatically impede a clinician's ability to provide efficient and safe patient care [1–5]. As a result, user-centered design, including usability testing, is becoming institutionalized for HIT development [6,7]. For example, clients who purchase electronic health records (EHRs) expect vendors to have their products certified by the ONC for HIT, which has only been the case over the last several years. Although this is a voluntary program in the U.S., this development suggests a shift in understanding the importance of incorporating user-centered design and usability testing in the development of HIT. Creating appropriate scenarios and tasks for usability testing can be critical for achieving a successful HIT product. In our experience with usability assessment, the most challenging part of designing a rigorous usability test is scenario development. The quality of the usability test and corresponding results are strongly influenced by the scenarios provided to test participants. Usability scenarios, and associated tasks, influence the types of usability issues uncovered [8], and subsequently, the quality and accuracy of usability testing. For example, scenarios and associated tasks often center on issues that might impact patient safety; that is, they describe circumstances around a hazard and usability testing can demonstrate the extent to which that hazard is mitigated by the organization's risk management process. For healthcare applications, we define a usability scenario as a description that provides clinical context for the end-user during usability testing. When end-users are healthcare professionals, scenarios most often consist of a patient case, which may be based on real or fictitious patient data. Scenarios can include personas [9] and can represent a number of clinical circumstances, from common to rare events, such as

* Corresponding author at: Department of Pharmacy Practice, College of Pharmacy, Purdue University, West Lafayette, IN, United States.
E-mail address: aruss@purdue.edu (A.L. Russ).

https://doi.org/10.1016/j.jbi.2018.01.001
Received 31 July 2017; Received in revised form 19 December 2017; Accepted 8 January 2018
Available online 09 January 2018
1532-0464/ Published by Elsevier Inc.

routine appointments and emergencies, respectively [10]. Scenarios provide some context for the end-user during usability testing. A simple scenario might be: "You are seeing a new patient who has a recent diagnosis of diabetes." While scenarios represent higher-level descriptions [11], tasks specify *what* the end-users should try to accomplish via the healthcare application. For instance, the above scenario might include a task to "order a diabetic eye exam". Each scenario includes at least one task, but more often incorporates several tasks [12]. Tasks focus the usability tests on specific software systems and design features so that relevant usability data can be collected and the goals of the evaluation can be achieved.

A handful of resources are available to facilitate the use of scenarios for usability tests of healthcare applications. For instance, the Office of the National Coordinator (ONC) for Health Information Technology outlined specific EHR capabilities that vendors should evaluate via scenarios during electronic health record (EHR) certification [13,14]. These range from testing the display of the patient's demographics and problem list in the EHR to the broader computerized provider order entry system [14]. These criteria are intended to promote improved user interfaces and foster consistency across EHR evaluations, thereby facilitating the EHR certification process [14–16]. The ONC outlines basic expectations for usability testing [14,17,18], but these documents do not describe factors to consider when developing scenarios nor do they provide guidance on how to develop usability scenarios to evaluate software features. In addition, the focus is on EHRs, rather than a broader set of health information technologies. The National Institutes of Standards and Technology (NIST) assembled several valuable reports to guide usability evaluations, including processes to improve usability methods for EHRs [7], and a template for effectively reporting usability results [19]. At least one NIST report provides some basic guidance on task development, but the topic is not discussed in-depth. Similar to NIST, the Federal Drug Administration (FDA), International Organization for Standardization (ISO), and International Electrotechnical Commission (IEC), provide basic guidance on usability methods, including some instruction on scenario and task development, for use with medical devices [20–22]. In other work, Vincent and Blandford describe the utility of scenarios across the product design lifecycle - as applied to medical devices - and explain how they developed scenarios for infusion pumps [12]. Their article describes clinical data sources used to inform scenarios, how they aggregated these data, and their process for confirming the clinical validity of the scenarios they developed. Their focus was on how to generate empirically grounded and validated content for scenarios, for use across the product lifecycle, rather than provide broader guidance on the multiplicity of factors to consider when developing usability scenarios.

A systemic review of usability evaluations, published in 2016, states that "there is currently little guidance on how to perform systematic evaluations of EHRs" [23]. We did not identify any articles that provided overarching guidance for usability scenario development in the healthcare setting. Therefore, the objective of this article is to outline some of the key factors to consider when developing usability scenarios and tasks for computer interface-based health information technologies. This article is informed by usability literature and the authors' experience conducting usability tests with a variety of healthcare applications. This article provides guidance on the design of scenarios and tasks, such as the content, wording, and level of sophistication, with special considerations for healthcare applications. This article is *not* intended to provide comprehensive guidance on usability evaluation nor associated usability methods and measures. Instead, we focus on the design of usability scenarios to be used in conjunction with usability testing involving end-users. We describe scenarios intended for one or more of these end-users: health administrators, healthcare professionals, patients, and patient caregivers. We also discuss scenario design in relation to both formative and summative usability testing. The motivation for this article is to provide a more consolidated resource on usability scenario development for usability experts new to healthcare, clinical and IT staff on collaborative usability teams, and students specializing in HIT usability. This paper is expected to aid novice through expert usability evaluators, with broad applicability across healthcare technologies.

2. Background

Collectively, the authors have experience on numerous usability projects, either as a team leader or collaborator. Example projects described in Table 1 represent a variety of applications and a diversity of end-users. As shown in Table 1, the number of scenarios, and associated tasks, can vary widely. A task may focus on a very specific human-computer interaction (e.g., "order a hemoglobin A1c") or require longer, more complex interactions with the technology. An example of the latter is "please complete medication reconciliation", a task that has been reported to take 13–83 min in clinical practice, depending on factors such as patient complexity and the number of medications to be reconciled [24–26]. In reality, the number of scenarios and tasks needed for usability analysis is not the important factor. What *is* important is that they address the intended objective of usability testing and are carefully constructed to provide valid usability data. The next section addresses factors that should inform the development of usability scenarios and tasks.

3. Ten factors that inform usability scenario development

We describe 10 factors (Table 2) that influence scenario development and organize these into three main categories: (1) project goals; (2) validity; and (3) other considerations. In the text, each factor is discussed in greater detail, with illustrative examples to highlight key issues and particularly challenging aspects of scenario design, especially as it relates to healthcare.

3.1. Project goals

3.1.1. Usability objectives and type of evaluation

Usability objectives guide the type of usability test and, therefore, the associated scenario and task design [7]. If the primary objective is discovery of usability issues, then a qualitative-based usability test using 'think-aloud' technique [45] is typically warranted. Qualitative usability testing often occurs during the formative stages of design and scenarios are typically more open and exploratory. However, qualitative methods are useful for both formative and summative testing.

If the primary objective, however, is to measure end-user performance, then quantitative usability testing is needed with usability measures such as time on task, usability errors, and user satisfaction, etc. [35]. Quantitative usability testing can be useful in both the formative and summative stages of design. For quantitative testing, scenarios and tasks associated with collection of performance-based measures should be more scripted and close-ended, with clear end points and associated measures (Table 3). This is because tasks that are open and exploratory often introduce a lot of variation in the data collected, making performance-based comparisons problematic; therefore, focused tasks are warranted.

Integrating qualitative and quantitative usability testing is often useful and warranted. This mixed-methods approach introduces some special considerations for scenarios and tasks, discussed later in Section 3.2.3, *scenario-related bias and confounders*.

3.1.2. Role of end-users

The type of end-users is important to consider when developing usability scenarios. The type of end-users who will use a given HIT application may range from novice to expert, and vary in age, health literacy, level of clinical proficiency, and many other factors. Clear processes for usability specification are outlined elsewhere [35], and here we discuss how scenarios and tasks are shaped by the end-users'

Table 1
Description of a sample of usability projects. For more details on the scenarios, see the referenced publication in column one. The third column indicates the number and types of end users that were included in the project as well as the number of usability scenarios.

Technology Focus	Usability testing goal(s)	End User(s); # *Scenarios*	Description/Example
1. Computerized decision support for anticoagulation [27]	Pilot test the usability of a Web-based atrial fibrillation decision support tool.	8 physicians; *9 outpatient scenarios*	Scenarios varied in complexity, with some patients taking warfarin and other patients not yet prescribed therapy. Physicians used the tool to help decide whether to start or discontinue anticoagulation (i.e., warfarin) therapy.
2. Clinical reminders [28]	Reduce usability barriers for clinical reminder use and evaluate the usability of a prototype versus current system.	16 nurses; *5 patient scenarios*	Each patient scenario included multiple clinical reminders for nurses to address (e.g., advance directives screen, depression screen, influenza vaccine, pain screening)
3. Clinical reminders [29]	Comparative evaluation of an original vs. redesigned reminders to assess if the display influences providers' prioritization of reminders.	16 physicians; *1 patient scenario*	End-users were instructed to prioritize which clinical reminders they would complete under time pressure. (i.e., "There are 5 clinical reminders to complete, but you need to see your next patient in 5 min.")
4. Web-based computerized decision support, stroke prevention [30]	Assess usability and understand potential implementation barriers for a new decision support tool for stroke prevention.	10 stroke care providers; *1 scenario*	To simulate a clinic visit, providers interacted with the patient actor and the decision support tool for stroke prevention. For the scenario, the patient had been recently discharged from the hospital after experiencing a stroke, and had multiple risk factors for experiencing another one.
5. Personal health record (PHR) [31]	Conduct a usability assessment of VA's MyHealtheVet to inform future design of this and other PHRs.	24 patients[b]; *4 scenarios*	Example: "You would like to record your level of physical activity last week….create a record of what you did last week."
6. Medication alerts [32]	Comparative usability study to assess whether a novel alert prototype improves usability and reduces prescribing errors compared to the current alert design.	20 prescribers; *3 fictitious patient scenarios*	Example: "The patient has been taking spironolactone for heart failure. Spironolactone was prescribed about a month ago and he only has a few tablets left. Begin renewing spironolactone."
7. Tool to support asynchronous medication reconciliation [33]	Iteratively evaluate and improve the usability of a medication reconciliation tool, in preparation for a clinical research trial	20 healthcare professionals; 10 patients;*1 scenario*	Physicians, nurses, pharmacists, and patients were asked to review and update the medication list via the tool. Healthcare professionals used a real patient chart. Patients used their actual medications.
8. Rapid usability evaluation method [34][a]	Implement usability methods to aid the work of clinical informaticists so they can more rigorously evaluate health IT before implementation	Physicians, medical residents; *Various scenarios*	Example: "Pretend that the information below includes your notes on the test patient (Pt X) from rounding. You anticipate that Pt X will be ready for discharge sometime tomorrow. Please use this information to complete the Anticipated Discharge Note template as best as you can."
9. Dialysis services software program[a]	A formative usability test to document usability issues experienced by administrators	5 healthcare administrators; *2 scenarios*	Example: "View all of the available output summaries and graphs from [this hospital] site profile to make a decision on the best dialysis make/buy strategy for your facility."
10. Palliative Care Template[a]	A formative usability test of a national-level template to document common usability issues	5 palliative care providers; *1 patient scenario*	Providers worked through the computerized palliative care template and completed the template using the test patient scenario.

[a] Quality improvement project.
[b] Patient caregivers were encouraged to participate in usability testing if they normally would assist the patient with the technology.

role in healthcare. If end users' roles vary with using HIT, then different scenarios and tasks are typically warranted for each role. This is because the context, needs, interaction, and workflow across roles (e.g., physician versus nurse) and healthcare settings (e.g., outpatient versus inpatient care) – which scenarios should generally account for – can be quite different. On the other hand, for some applications, different types of end-users may use the technology for similar clinical roles and activities and only one test group may be necessary. An example of this is computerized provider order entry (CPOE) for medications, where testing could include one group of various prescribers - physicians, clinical pharmacists, nurse practitioners – who are asked to complete the same set of scenarios and tasks [32].

A needs analysis can be conducted for each application to determine if there are distinct roles among end-users, and whether distinct corresponding scenarios are needed for usability testing. Inviting input from healthcare professionals and administrators during the needs analysis can aid in defining the potential range of end-user roles across care settings, facilities, and healthcare organizations, as well as the types of scenarios that are clinically appropriate [12]. For example, if the application is computerized clinical reminders, a nurse may be expected to complete clinical reminders associated with routine screening (e.g., tobacco use, alcohol use, and pain assessment). In contrast, a physician may be expected to complete clinical reminders associated with diabetes management or cardiovascular risk reduction [46]. Therefore, distinct scenarios and tasks for usability testing of clinical reminders should be developed, accordingly. As another example, suppose the project goal is to evaluate computerized consults. During clinical practice, primary care providers enter consult requests, and specialists respond to consult requests [47]. Each role involves very different interactions and workflows with the application, necessitating different scenarios and tasks for usability testing.

There may be occasions when time and resources do not permit including more than one user role in a usability test and practical tradeoffs must be considered. In this case, performing a usability test with one user role is better than not performing the test at all. If necessary, it may be possible to combine user roles based on similar skills, experience, and background. However, we strongly encourage testing multiple user roles if at all possible when there are distinct roles for end-user groups that will be using the application.

3.1.3. Target performance goals

For quantitative, performance-based usability testing (see Section 3.1.1, *usability objectives and type of evaluation*) target goals can be set for each usability attribute of interest and the associated task and

Table 2
Factors to consider for scenario and task development. Ten factors are outlined below along with a description. The right hand column lists additional resources and/or example articles that relate to the factor.

Factors	Description	Related references
Project goals		
1. Objectives of usability testing	Usability testing objectives should guide scenario and tasks so appropriate usability data can be collected and the objectives are achieved • Scenarios and tasks for qualitative usability testing are generally more open and exploratory to aid discovery of usability issues • Scenarios and tasks associated with collection of performance-based measures (i.e., quantitative testing) should be more scripted and close-ended rather than open and exploratory • Task development often centers on software functions that are most frequently used, that are essential for clinical care, and/or have patient safety implications. Alternatively, there may be emphasis on evaluating only a specific part of the healthcare application.	[28,35,36]
2. Role of end-user(s)	When the healthcare application is anticipated to be used by multiple, disparate end-user groups to complete dissimilar clinical activities, different scenarios and tasks may be necessary for each user group	[35,37]
3. Target performance goals	For performance-based usability testing, the tasks that comprise scenario(s) become the "measuring instruments" for the quantitative metrics. For example, the design team may decide *a priori* that their healthcare application should support the completion of specific tasks (a, b, c…) in a specified amount of time, with no errors, etc. See also Table 3	[35]
4. Evaluation time constraints	Design scenarios to fit within the constraints of end-users' availability and the projects' allotted time for the usability session	[32,38]
Validity		
5. Clinical Focus	Scenarios should generally be clinically pertinent	[12]
6. Fidelity	When developing scenarios and tasks, consider the extent to which the scenarios need to represent the complexity of actual clinical situations and the real-world clinical environment	[12,39]
7. Scenario-related bias and confounders	Scenarios and tasks should be developed to: (1) avoid influencing end-users' expectations and perceptions of the software; (2) avoid guiding end-users on how to use the software in the 'right way'; (3) avoid confounding usability results	[40,41]
Other considerations		
8. Embedded probes	A probe is an error or tricky situation that is intentionally inserted as part of a fictitious scenario or task in order to assess the clinical or safety implications of a specific aspect of health information technology	[42,43]
9. Minimize risks to end-users	Consider potential psychological and clinical risks to end-users, including providers and patients, when planning a usability test. Develop introductory scripts, scenarios, and tasks that minimize these risks	[44]
10. Healthcare related outcome measures	Scenarios should be chosen or developed to accurately assess any clinical outcome measure(s) of interest	[32]

Table 3
Illustrative examples of usability specifications and how usability tasks can be used as measuring instruments. There are many different types of specifications and metrics that can be incorporated into usability testing and examples below are *not* intended to be comprehensive.

End-user group	Usability attribute	Example task	Usability metric	Baseline level	Target level
Primary care physicians	Initial user performance (i.e., learnability)	Task: Complete the clinical reminder for colorectal cancer screening	Time on task	5 min	3 min
Primary care nurses	First impression	Task: Please access the clinical reminder dashboard and then complete the questionnaire to rate your initial impression.	Satisfaction questionnaire, average rating across users	Ave ≥ 4[a]	Ave ≥ 6[a]
Primary care physicians	Post-initial performance	Task: Complete the clinical reminder for colorectal cancer screening	Usability video, average number of errors	2 errors	0 errors

[a] for this illustration, 1 = strongly dissatisfied and 7 = strongly satisfied with the healthcare application

usability metric. Hix and Hartson define a usability attribute as a "general usability characteristic to be measured for the interface" [37]. For each attribute, usability tasks serve as measuring instruments to assess whether the *a priori* target goals are realized (Table 3). For example, suppose a new ordering system is being evaluated for lab tests, with a target goal of taking no more than two minutes per laboratory order with no errors occurring. Consequently, a scenario is developed for a patient needing several lab tests ordered; a task is created for each order; and the data collected from each task can be used to measure efficiency and number of errors. For this type of usability testing, a usability specification or target table is useful. As one approach, Hartson suggests using a table similar to Table 3 [35]. The value to be measured, or usability metric, describes the kind of data to be obtained for the usability attribute. The 'baseline level' is a benchmark level of the usability metric. This is often the level that has been measured for the current version of the system that the design team is trying to improve. Or, if there is no current version, it is an initial level that is established so that there is some baseline for comparison in future design iterations. The target level is the goal for the final design. One way to establish a target level for time on task, for example, is by using a keystroke-level model to predict task execution time [48,49].

3.1.4. Evaluation time constraints

Healthcare is a time-pressured, high workload environment, with clinicians often on-call, working long hours, nights, and weekends to provide patient care. Thus, the length of the usability session influences the feasibility of end-user recruitment and success of completing the evaluation. To promote project success, the length of the usability session should be conducive for participants' schedules. We find that healthcare professionals are more willing to participate in shorter sessions (e.g., 30–60 min) [32,50], while patients are often able to participate in longer usability sessions (e.g., 60–90 min) [38]. One strategy to keep the session within the planned time constraints is to consider whether any technology-based tasks are *extraneous* to the usability

Table 4
Overview of three types of usability scenarios, along with advantages and disadvantages of each for usability testing of healthcare technologies.

Scenario type	Advantages	Disadvantages
a. **Real-Life scenario** *Based on real-life patient data that have not been altered in any way. May be embedded within a functional EHR or extracted for use in a laboratory setting.*	• High fidelity • Clinically realistic, captures real-life complexity • Since scenarios are already developed, often saves substantial time for the project team; likely more efficient to use. • Likely to be more natural, realistic for end-users during usability testing • May include interesting clinical situations that are unanticipated and unlikely to be captured by fictional scenarios.	• Access to real-life cases may be challenging, especially for those outside the healthcare system • Potential confidentiality concerns if video recording the computer screen, actual patient data during usability testing • IRB approval/patient consent may be required for scenario development • If usability testing occurs in a naturalistic setting during clinical practice: o Primarily useful for evaluating technology post-implementation o Cannot insert safety probes o Patient data are dynamic, and may change during the usability testing process (not as good for consistency across end-users, or A vs. B design comparisons) • Cannot tailor scenario to test specific features of the technology or outcomes of interest
b. **Fictitious scenario** *Developed by the project team. Does not include any actual patient data.*	• Can be used for usability testing at any point in the development lifecycle, including pre-implementation • Unlimited flexibility in scenario design • Ability to insert safety probes • Can tailor scenario to test specific features of the technology and outcomes of interest • Facilitates comparison of two technologies under the same conditions	• Ranges from low to high fidelity • Often extremely time consuming to develop • Requires close partnership with clinical team members who often have limited time • May be less complex or true to life • Risk of scope creep for scenarios, since a real patient will generally have much more data than what are feasible to present in a fictional scenario; tough choices on scenario scope
c. **Combination scenario** *A scenario that is based, in part, on actual patient data but also includes some elements that are fictitious or have been intentionally manipulated in some way.*	• Generally higher fidelity than a purely fictional scenario • Offers same advantages as fictitious scenarios (see above), but generally higher fidelity and clinical realism • Once the factual patient data are obtained, requires less time to construct than an entirely fictitious scenario • Can insert safety probes to test specific usability hypotheses	• May be hard to identify real-life patient scenarios of interest that match the usability testing objectives • May be time consuming or difficult to extract patient data from the database, EHR, etc., for use • Requires close partnership with clinical practitioners, who often have limited time • Similar confidentiality, IRB requirements as real-life scenario (see above)

objectives. Extraneous tasks can be shortened, eliminated, or perhaps intentionally facilitated by the scenarios or usability moderator [50]; this is one instance where intentional bias (i.e., instructions) via scenario wording can be leveraged to guide participants in an appropriate manner. In addition, set reasonable time limits for completion of task(s) to ensure end-users attempt the full set of tasks during the usability session. If the end-user exceeds the time limit, the moderator can ask the end-user to move onto the next task. Related to timing, it is also helpful to have a protocol in place to handle instances where a healthcare professional is paged or interrupted during the usability session, since healthcare professionals might be participating in usability testing when they are on break or on call. This may involve pausing the usability session briefly and turning off recording devices, asking the end-user to return later to complete the remaining tasks, or rescheduling the session.

3.2. Validity

3.2.1. Clinical focus

For health information technologies, scenarios and tasks should represent real-world activities that are relevant for clinical practice and patient care [19]. This is a form of external validity and specifically ecological validity. In other words, it is often important that the scenarios in the usability test are sufficiently realistic such that results can be generalized to the clinical settings and situations in which the phenomenon being studied would naturally occur. Other forms of validity are also important for usability testing (e.g. internal validity – including test, criterion, content, construct, and face validity). Here, we focus specifically on ecological validity, which can achieved through scenario and task design.

Usability tasks should be informed by an understanding of the context of technology use and associated clinical workflow. Thus, it is often informative to conduct field observations, interviews, or ethnographic-type studies [46,51] prior to research-level usability testing. It can also be very helpful to include at least one or two practicing clinicians or other representative end-users on the scenario/task development team. Developing scenarios and tasks generally requires both usability and clinical expertise, and in our experience, this often necessitates pairing a usability specialist with a clinical representative to develop the materials in close partnership. The choice of tasks depends on the usability objectives, but some common criteria are to choose tasks that occur frequently in clinical practice or that have important clinical or patient safety implications. Depending upon the usability testing objectives, tasks may consist of a wide range of activities: some with clear end-points (e.g., “send the secure message”), open-ended tasks (e.g., assess the patient’s risk for stroke), and even self-generated tasks (e.g., “proceed with how you would use the EHR”) [8].

In addition to including clinical representatives on the scenario/task development team, at least three other strategies can be used to confirm and strengthen clinical validity. First, clinical scenarios can sometimes be informed by published clinical guidelines, case reports, or references [32,52]. Second, similar to “member checking” approaches in qualitative research [53], clinical representatives can be invited to review written scenarios drafts for clinical relevance [9]. Discrepancies among clinicians may be resolved via consensus discussion meetings, such as those used in qualitative research [54], or a Delphi process [55,56]. Alternatively, Vincent et al. describe a detailed review process for integrating clinical feedback into scenarios for medical devices [12], which may inform scenarios for other types of HIT. Third, scenarios and tasks can be piloted with a few end-users prior to usability testing. For

these validation processes, clinical representatives or end-users should ideally be external to the project team [9].

3.2.2. Fidelity

Another consideration for scenario/task development is the extent to which these components need to represent the complexity of the real-world clinical environment. Scenarios and tasks may incorporate several different features that together comprise overall fidelity. For example, certain patient data in the scenario may be based on real patient data and other data may be fictitious. The accuracy and precision of data are other features to consider. Also, do certain patient data need to change over time - over the course of the scenario as they would in real-life? All of these features are important to consider for overall scenario fidelity. These considerations are especially important if scenarios will be entirely fictional. Table 4 outlines some advantages and disadvantages of real-life versus fictional scenarios for usability testing of healthcare applications. The appropriate level of fidelity for scenarios depends upon usability objectives and constraints.

In many instances, it is more practical to create fictitious scenarios for usability testing rather than use actual patient data, even though fictitious patient data may reduce the fidelity of the scenario. This is especially the case for usability testing in a laboratory environment, which may have limited access to clinical systems and real patient data. Due to the dynamic, ever-changing nature of patient data in healthcare applications (e.g., patients' medications are continuously being ordered, adjusted, and discontinued; patients receive new lab results and diagnoses), fictional scenarios are typically necessary for comparative testing of two different designs, so that HIT designs can be evaluated under the same conditions. Fictional scenarios can also facilitate usability testing of prototypes and innovative technology designs that have not yet been incorporated into EHRs nor implemented into clinical care. In addition, conducting usability testing of the EHR using real patient data is typically prohibited for non-clinical care purposes without special approvals. For research especially, it is often necessary to use fictional scenarios in conjunction with a EHR demonstration version or mock-up, since most functions of an implemented EHR cannot be used without real-life downstream effects. For example, usability testing of the laboratory ordering system can trigger actual ordering of laboratory tests for the patient.

It is sometimes possible to mix real patient data with fictitious patient data (Table 4, combination scenario), as an intermediary option between the real-life scenario and fictitious scenario. Researchers in other domains have developed methods to mix previously collected "real" data with fictitious data and have also developed methods for mixing experimental interventions with live operations in order to assess usability. For example, Bass et al. (2011) created an experimental design for studying weather forecaster and emergency manager interaction (and related software) such that the weather forecasters participated outside of their regular job duties – by providing fictional, real-time weather assessments and warning decisions – while emergency managers participated as part of their normal operating hours [57]. This type of design allowed for experimental interventions on the forecaster side of the interaction while also preserving the real-life context of the emergency manager interaction during their regular job duties.

If usability testing is being conducted in a laboratory environment [39], then usability evaluators should also consider how much the scenarios should simulate clinical workflow, interruptions, and other events that occur in a clinical environment. Clinicians experience a great deal of interruptions during their clinical work [58], but interruptions rarely occur in a laboratory environment unless they are included as part of the broader usability simulation and associated scenarios. If the objective of the usability test is to evaluate an early design of an interface, including functionality and navigation, then simulating interruptions and the complexity of the "real life" clinical environment in general is not necessarily helpful. Furthermore, if a comparative usability test is being conducted to rigorously evaluate the strengths and weaknesses of two different technology designs under the same exact conditions, (e.g., as in ref. [32]) then interruptions may interfere with the assessment because, if they are not consistent across the two HIT designs, then interruptions can confound usability data (see Section 3.2.3). On the other hand, if the objective of the usability test is to evaluate clinical performance with a HIT tool under the conditions of high cognitive workload or interruptions, then it is important that scenarios simulate the complexity of the clinical environment.

The fidelity of the prototype itself that is being used in the scenario is another consideration if the system being tested is in development or is a proposed redesign. A low fidelity prototype may be a paper prototype or simple wireframe. These low fidelity prototypes are not necessarily faithful representations of HIT details such as look, feel, and behavior [35]. Rather, they give high-level, more abstract impressions of the intended design. With a low-fidelity prototype, it may be difficult to develop a corresponding, high-fidelity scenario that mimics the real-world complexity of the clinical environment: low-fidelity prototypes may constrain usability testing to the use of low-fidelity scenarios. High fidelity prototypes, conversely, include details of appearance and interaction behavior. Although high-fidelity prototypes are much more realistic, interactive, responsive, and representative of the real final product than a low-fidelity prototype, the added expense of developing one may not be warranted if the objectives of the usability test can be accomplished with a low-fidelity prototype. Therefore, for many usability tests, lower fidelity for scenarios are more pragmatic and efficient for project timelines, yet captures high quality usability data [20]. Combination scenarios appear to be underutilized in usability projects and may help to balance the fidelity and feasibility of usability testing.

3.2.3. Scenario-related bias and confounders

It is important to minimize bias and confounders in healthcare studies, such as clinical trials, and usability studies are no exception. In this article, we focus on bias and confounders as related to usability scenarios. Much of this section focuses on bias and confounders that are introduced by scenarios (i.e. scenario-induced). For example, the wording of usability scenarios and tasks can introduce bias by influencing usability outcomes in favor or against a HIT design. However, we also include examples of how scenarios and tasks can sometimes be strategically used to mitigate bias and confounders associated with other aspects of usability testing.

We discuss two types of bias: influencing end-users' perceptions and inappropriately guiding end-users. For the first type of bias, influencing end-users' perceptions, imagine if a task read: "Use this new, improved design to complete clinical documentation". Terms "new" and "improved" might shape end-users' expectations of the tool. This terminology may raise end-users' expectations for the tool, which could inflate their positive feedback and satisfaction ratings for the technology, or alternatively, may cause end-users to give more harsh ratings if their elevated expectations are unmet. The general approach to mitigate this type of bias is to refer to the software design(s) in a neutral manner, even if you believe (or have some evidence) that it is a stronger or weaker design. This is especially important when comparing two or more designs head-to-head. For comparative usability testing, it is preferable to state that the comparison is between "two different designs" rather than explaining it is an "old versus new" design comparison. The latter is acceptable for publication, but should be avoided with end-users. In a subtle example, a recent study evaluated a secure messaging, medication reconciliation technology, formally known by its acronym, "SMMRT". [33] An early draft of scenarios used this term, pronounced "smart", but researchers realized this may bias end-users' expectations. Thus, the term "medication tool" was used instead in the scenarios to describe the technology in a neutral manner.

A second pitfall is that scenario and task wording may introduce bias by inappropriately guiding or making it easier for end-users to use the technology. This can consequently yield favorable, but false,

usability outcomes. Suppose a usability test is being conducted to assess a laboratory display for incoming medical residents, with the objective of evaluating the learnability of the technology. A task that states, "Using the menu on the left, pull up the laboratory results", would guide the resident on how to navigate to the laboratory results and falsely inflate outcomes for ease of learning. Instructions on how to appropriately interact with the technology should be avoided. This includes avoiding any terms that might offer unintentional clues to end-users about the navigation or software design. Likewise, if the project is intended to evaluate how HIT influences clinical decision-making, then scenarios should be crafted to avoid accidentally nudging end-users towards any particular clinical decision. Additionally, the order of tasks can bias usability results if the order helps facilitate the "appropriate" workflow and sequence of steps needed for effective use of the software. This type of bias can often be mitigated by a more general patient scenario, or a task that describes the overarching goal, rather than the specific steps of completion. Scenarios and tasks should provide guidance to end-users on *what* activities to complete, but *not how* to complete them [40,41].

Like clinical trials, confounders can positively or negatively influence outcomes and alter usability results. For instance, scenarios that isolate tasks or break normal task flow might introduce confounders, but sometimes might be necessary to fit tasks into a given test session or to collect certain measures. These factors should be considered when interpreting results and reporting findings. Another common example is when the same sample of end-users are asked to complete the same clinical scenarios for two or more HIT designs. This can introduce potential learning effects, due to the cross-over design, and end-users may complete tasks more easily, more quickly, and with fewer usability errors the second time around. Of course, one way to eliminate this potential learning effect is to use different end-user participants for each design. However, the time and resources to recruit separate user samples for each design is often impractical. If both designs use the same sample of end users, then one approach to mitigate the potential confounder of learning effects is to counterbalance the order of testing across end-users, so the first end-user completes design 'A' first, while the second end-user completes design 'B' first and so on. In addition, this type of confounder can may be mitigated by altering non-critical aspects of the scenario so that the end-user is less likely to recognize that that it is the same scenario. For example, this might include changing the name of the patient in a fictional scenario as well as changing non-clinical details. In addition, distractor tasks can be inserted in between scenarios to help mitigate learning effects. An example of a distractor task is to ask participants to complete a demographics questionnaire [59] in between the two designs. Another, likely more rigorous, approach is to use a washout period between the two designs – with the usability sessions scheduled day(s) or more apart – to reduce the likelihood that the end-user will remember the scenarios and their associated human-computer interaction [32]. There does not appear to be definitive guidance in the literature on the appropriate length of washout period for usability testing. We previously used a minimum two- week washout period and this approach seemed to be effective [32,60]. A potential downside to a washout period is that end-users may not be willing or able to participate in both usability sessions, and some may be lost to follow-up, resulting in incomplete usability data. A washout period is most likely to be feasible for usability projects that are occurring over several months, allowing more flexibility for participants' schedules.

In addition, it is often desirable to use a combination of qualitative and quantitative usability testing methods (see Section 3.1.1). It is especially important to consider scenarios and tasks when integrating these methods, since combining these approaches often introduces confounding effects. Scenarios and tasks can be used judiciously, however, to help mitigate confounding effects. Of primary concern is that 'think aloud' technique, which is central to qualitative testing, can confound quantitative, performance-based measures of time on task (i.e., efficiency). Specifically, there is evidence that think aloud impacts quantitative performance measures for tasks that take longer than about a minute to complete [61,62]. Nonetheless, there are several strategies for integrating qualitative and quantitative-based usability methods. One strategy for integrating these types of usability tests is to designate a portion of tasks for qualitative data collection via think aloud technique and a different set of tasks for quantitative, performance-based measures. (See example in Russ et al., 2014 [32]). A second strategy is to collect quantitative performance measures during the usability tasks, and then conduct a retrospective review of the usability video with the end-user using think-aloud at that time [63], or conduct a debrief interview while reviewing a video replay of the usability test [64]. A third strategy could include the use of multiple usability tests, or iterative testing, where each usability test has a different usability objective [65]. For example, an initial test, with 10 or so end-users, could use think-aloud technique with 5 tasks to uncover a list of major usability issues. The second round of usability testing, with the same or different set of end-users, and the same 5 tasks, could then be used for collection of performance-based measures.

3.3. Other considerations

3.3.1. Embedded probes

Embedded probes [42], sometimes referred to as "safety probes" [66], are especially useful tools to evaluate how well the technology helps users catch or prevent errors. The concept of embedded probes, at least within the healthcare and human-computer interaction context, appears to be first described by Patterson et al. 2004 [42]. A probe is something that is intentionally inserted as part of a fictitious scenario or task that is intended to be misleading, erroneous, or somehow challenging for the usability end-user as s/he uses the technology. A probe typically represents an actual, clinical phenomenon. For example, a fictitious patient scenario may contain duplicate medication orders, a missing laboratory result, or conflicting information that the end-user is expected to address during their interaction with the technology. A similar approach is described by Prichett and Hansman, who used testable responses in aviation scenarios to receive unambiguous accounting of the types of tasks for which the pilot was successful and had sufficient situation awareness [43]. Probes are generally informed by field observations or other sources of knowledge about real-world healthcare complexity. Patterson et al. created probes within scenarios for nurses so that their research team could clearly interpret outcomes for a bar coding medication administration technology [42]. One of their probes consisted of a patient with a medication order for amoxicillin, while having a documented allergy to penicillin. Researchers defined correct and incorrect responses to this and other probes to assess outcomes. Probes can also potentially be used to assess how the technology design might be contributing to rare, but serious, safety events that occur in clinical practice. For example, if a patient safety manager suspects that the usability of allergy entry is inadequate only for a specific medication, a probe could be used to test this hypothesis. Probes can be used in fictitious or combination scenarios (see Table 4). Probes are only a viable option when scenarios, tasks, or related aspects of the technology design can be manipulated and used in a laboratory or other simulated setting, such that the fictional probe – which is usually an inserted error or cause for concern - can be incorporated into the usability test without impacting real-life clinical care [42].

3.3.2. Minimize risks to end-users

It is important to consider risks in human subjects research, including usability studies. An in-depth discussion of research ethics in healthcare is available in other literature (e.g., [44]). Here, we focus on some special considerations for introductory scripts, scenarios, and tasks in terms of minimizing risks to end-users, whether they are healthcare professionals, patients, or family caregivers. In usability testing, risks are often minimal and usually consist of psychological

risks and/or the potential loss of confidentiality, although the types of risks can vary depending upon the individual project. A carefully crafted introductory script or research consent form can help reduce some psychological risks. For example, healthcare professionals in particular may feel they are being "tested" in a usability evaluation, and these documents can remind end-users that the technology design is the focus of the evaluation, rather than their individual clinical or software skills. These documents can also let end-users know whether they will be completing scenarios and tasks with a prototype or software system that is separate from patient care or if they will be using a live, implemented technology with real clinical functionality and implications. Likewise, end-users can be informed about whether scenarios utilize fictitious or real clinical data. If usability testing is partially or fully embedded in real-life clinical practice or leverages actual patient data for scenarios, consider how to handle end-users' concerns if they identify a medical error as part of the testing session. (This *excludes* safety probes or errors that are intentionally embedded into the usability materials.) For example, during usability testing of an EHR, a physician might identify a potentially serious dosing error for a medication that was prescribed for the patient by someone else. As one approach, healthcare professionals within the same institution could be instructed to notify or contact the patient's provider(s) with any concerns. Likewise, patient end-users could be instructed to contact their provider with medical questions that arise. In some situations, the error or incident that is identified may be outside the scope of the usability evaluation. Prior to usability testing, the project leader should provide a protocol to the team on how to respond to these situations, and the project lead, or primary investigator, should be notified about any incidents. The usability team should have a plan in place for these types of situations, with input from Institutional Review Board institutions, especially if the evaluation involves a live, implemented technology.

3.3.3. Healthcare related outcome measures

For usability tests of HIT, outcomes of interest may include not only traditional target performance measures for usability (e.g., time on task, task success, usability errors, to name a few; see also Section 3.1.3) but also healthcare related outcomes. Examples might include measuring clinical decision-making in response to a computerized decision support tool [32]; how a display influences providers' interpretation of laboratory results; or how different interface designs influence a patient's ability to refill prescriptions through a personal health record [38]. Scenarios can be intentionally chosen or crafted so that healthcare related outcomes can be measured. Fictitious and combination scenarios (Table 4) are especially useful for this purpose since they are readily manipulated and provide a great deal of flexibility to focus the evaluation on specific clinical outcomes.

4. Discussion

In summary, we outlined 10 factors to consider when developing scenarios and tasks for usability testing. These factors may be used as a checklist (Fig. 1) to facilitate the development process for scenarios and tasks, and consequently may increase the rigor and validity of usability data. These factors are generally applicable to both applied usability projects designed for rapid feedback to a design team, as well as usability experiments as part of human factors or human-computer interaction research.

Some factors have been widely discussed in usability literature and texts (e.g., usability objectives, end-users) [35,67,68], although generally not in the context of scenario/task development for HIT. Other factors, such as embedded probes, bias and confounders, and health-related outcomes measures have received little to no attention.

Furthermore, some factors are likely easier to address during scenario/task development and there are sometimes tradeoffs between factors that warrant consideration during scenario development. As one example, a complex, high-fidelity scenario may require more time for end-users to complete, and therefore may not be conducive with the time constraints of the usability testing session, so a lower fidelity scenario is necessary. Additionally, when aligning scenarios with usability objectives, it might not be possible to avoid all types of potential bias and confounding. Avoiding bias in scenario wording may be more challenging for individuals who are new to scenario development or have experience writing instruction manuals. Somewhat contrary to an instruction manual, a basic approach for usability scenarios and tasks is to provide the end-user with the *minimum* amount of information that is necessary for them to know what activities they are expected to complete during usability testing. Despite the best efforts of scenario/task developers, given the complexity of scenarios and tasks, sometimes bias or confounders cannot be avoided, might not be caught in advance, or might be unintentionally introduced during usability testing

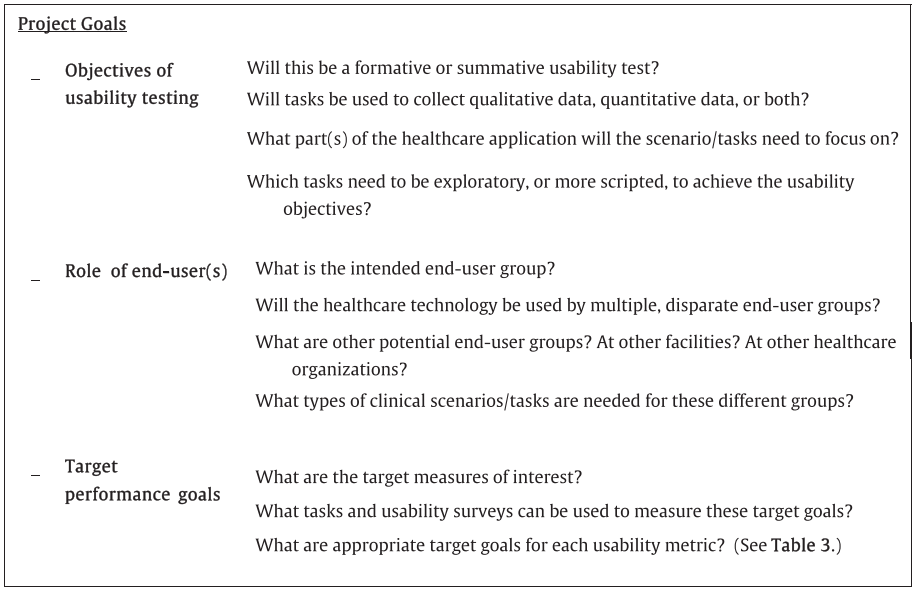

Fig. 1. Checklist of factors for consideration when developing usability scenarios and tasks. Definitions and examples of each factor are available in Table 2. This checklist may facilitate scenario/task development, but is not intended to be an exhaustive list.

_ Evaluation time constraints	What are the time allocations for the usability testing session? Will end-users have adequate time to complete the scenarios/tasks? Are there any scenarios/tasks that need pre-set time limits? Or will end-users be permitted to work on scenarios/tasks for as long as they like?
Validity	
_ Clinical Focus	What types of scenario(s) and task(s) are needed to align with the clinical activities associated with the healthcare application? What types of clinical experts or end-users should review drafts of the scenarios and tasks to assess clinical relevance?
_ Fidelity	To what extent do scenarios/tasks need to represent the complexity of actual clinical situations? The real-world clinical environment? Can usability objectives be achieved by low fidelity scenarios?
_ Bias and Confounders	Are scenarios/tasks free of any language that could influence end-users' expectations of the healthcare application? Are scenarios/tasks free of any language that might facilitate end-users 'correct use' of the technology? What are potential confounders? Do the scenarios/tasks minimize confounders?
Other Considerations	
_ Embedded probes	What is potentially high-risk with the healthcare application? Are there any a-priori hypotheses of potentially risky usability errors that might occur with the healthcare application? What scenarios/tasks are needed to assess these risks? What information or errors might be inserted into the scenarios/tasks to evaluate these potential risks? For each probe, what are the pre-defined correct and incorrect actions? (Scenarios are then usually designed to capture binary outcomes for probes.)
_ Minimize risks to end-users	What are potential psychological, clinical, or other risks for each end-user group? What can be modified in scenarios/tasks to minimize these risks?
_ **Healthcare related outcome measures**	In addition to traditional usability metrics, are any clinical outcomes of interest? What data are needed to measure these outcomes? Have scenarios/tasks been adequately designed to collect these data?

Fig. 1. (*continued*)

procedures. If this occurs, then any aspects of scenarios or tasks that could unduly influence the interpretation of usability findings should be noted in usability summary reports and/or as research limitations.

The primary motivation for this article was to present more in-depth, consolidated recommendations for developing usability scenarios for use with HIT, but we also hope this article furthers the discussion on how to effectively evaluate and compare HIT systems. Assuming an EHR vendor or healthcare organization's goal is to achieve ONC certification, the current ONC criteria are sufficient. However, we recommend that vendors and organizations perform iterative usability testing post-certification for the purpose of gathering richer user data on how to continuously improve the EHR and its associated functions, including those not covered by the certification criteria. More meaningful usability testing can be achieved and specifically tailored to each unique product and healthcare environment by carefully considering these 10 factors to produce strong scenarios and tasks for usability testing.

The 10 factors we outline in this article is not intended to be an exhaustive list. These factors are based in large part upon our experience with usability projects, and while diverse, still has bounded scope. This commentary is intended to supplement, rather than supersede, any other guidance on scenario or task development. Along these lines, information on usability scenarios within this document represent one important aspect of a broader clinical risk management plan that, in addition to usability testing, should include many other safety techniques for HIT, such as failure mode and effect analysis and root cause analysis methods [69,70]. Further, while we have usability experience in research and quality improvement/applied projects [71], our expertise lies largely within the realm of research. Consequently, there may be additional considerations for scenario and task development that are not captured by the factors we present. Furthermore, much of our usability testing experience is with healthcare professionals as participants, and we have comparatively less experience with patient-

facing healthcare technologies. Literature on various types of scenarios has been published over many decades (e.g., [72,73]) and is quite broad. In the future, systematic reviews of scenario methods and their application to HIT might be useful to continue to advance the field in this area. In addition, in parallel to scenario development guidelines outlined in this paper, others are pursuing model-based design and testing methods to potentially achieve a more state of the art approach to scenario development. For example, Bolton and Bass describe a modeling effort that utilizes a system architecture composed of models including the human mission, human task behavior, and operational environment [74]. Others researchers have taken a similar modeling approach [75,76]. This type of model-based usability development has great potential for a more routine or streamlined approach to scenario design.

Also of note is work in other domains to create frameworks for the development of human-centered scenarios. For example, researchers in air transportation developed a modeling and simulation framework for analyzing the impact of different function allocations on task work and teamwork [77,78]. As part of the framework, they evaluate function allocations within a range of operational scenarios. Others in aviation developed a framework to support the creation of scenarios that assess system level performance while considering the system, the humans that interact with it, and the environment [79]. In another example, researchers developed a method for automatically generating usability measures from task models that enables analysts to use formal verification to perform an evaluation [80]. These frameworks and model-based approaches may be quite useful for healthcare applications as well, including HIT scenarios.

While this article is intended to benefit individuals with varying degrees of usability experience, it is not a substitute for usability training or expertise. Usability scenarios should be developed in collaboration with usability expert(s) who have formal training and prior experience with needs analysis, visualization, user interface design, and/or usability testing. Furthermore, well-designed scenarios do not supplant the necessity of a comprehensive needs analysis as a precursor to usability testing [35]. Even well-conceived usability scenarios can be ineffective if other aspects of the usability evaluation, such as appropriate evaluation design, moderator skill, usability analyst's expertise, or measures are lacking. This paper focused on scenarios, rather than usability measures, and the field could benefit from systematic reviews on usability measures and how these measures might be expanded to better assess clinical processes and healthcare related outcomes.

In conclusion, this article offers guidance on usability scenario and task development and can be applied to a wide range of clinical information systems and applications. The checklist in Fig. 1 may be used by usability researchers and practitioners to carefully construct meaningful scenarios and tasks to produce rich usability data. This article is expected to aid usability researchers and practitioners in their efforts to advance HIT.

Author contributions

AR conceived the idea for this article. JS drafted results sections 3.1.1–3.1.3, 3.2.2, and Table 3. AR drafted the rest of the manuscript. Both authors reviewed and critically edited the scientific content of this article.

Conflicts of interest statement

The authors declare there are no conflicts of interest.

Acknowledgements

We would like to thank Ms. Rachel Gruber for assistance with manuscript formatting and submission. This work was supported in part by the Center for Health Information and Communication, Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service, CIN# 13-416. Dr. Russ was supported in part by a VA HSR&D Research Career Development Award (CDA 11-214). Additional sources of funding for usability projects mentioned herein include: VA HSR&Dgrant #PPO 09-298 (PI: Russ); VA HSR&D grant IIR #14-059 (PI: Dr. Steven Simon); VA Stroke Quality Enhancement Research Initiative (QUERI); The VA Engineering Resource Center (VERC), specifically, the Center for Applied Systems Engineering (VA-CASE); VA Palliative Care Quality Improvement Resource Center; and a Veterans Health Administration (VHA) Systems Improvement Capability Grant. Views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

References

[1] R.L. Wears, M. Berg, Computer technology and clinical work: still waiting for Godot, JAMA 293 (10) (2005) 1261–1263.
[2] D.F. Sittig, E. Belmont, H. Singh, Improving the safety of health information technology requires shared responsibility: it is time we all step up, Healthc (Amst) 2017.
[3] J.D. Halamka, M. Tripathi, The HITECH Era in Retrospect, N. Engl. J. Med. 377 (10) (2017) 907–909.
[4] R. Koppel, J.P. Metlay, A. Cohen, et al., Role of computerized physician order entry systems in facilitating medication errors, JAMA 293 (10) (2005) 1197–1203.
[5] L. Rosenbaum, Transitional chaos or enduring harm? The EHR and the disruption of medicine, N. Engl. J. Med. 373 (17) (2015) 1585–1588.
[6] Testing and Test Methods: 2014 ed. Draft test scenarios. Office of the National Coordinator (ONC) for Health Information Technology. Retrieved August 11, 2016, from: <https://www.healthit.gov/policy-researchers-implementers/2014-edition-draft-test-scenarios>.
[7] R.M. Schumacher, S.Z. Lowry, (NISTIR 7741) NIST Guide to the Processes Approach for Improving the Usability of Electronic Health Records: National Institute of Standards and Technology., 2010.
[8] T. Tullis, B. Albert, Issues-based metrics, section 5.7: bias in identifying usability issues. measuring the user experience: Morgan Kaufmann Publishers, 2008 pp. 116–17.
[9] C.J. Vincent, A. Blandford, The challenges of delivering validated personas for medical equipment design, Appl. Ergon 45 (4) (2014) 1097–1105.
[10] B. Shneiderman, C. Plaisant, Section 3.7: scenario development, in: M. Hirsch (Ed.), Designing the User Interface: Strategies for Effective Human-Computer Interaction, fifth ed: Addison-Wesley, 2010, pp. 116–119.
[11] S.Z. Lowry, M.T. Quinn, M. Ramaiah, et al., (NISTIR 7804) Technical Evaluation, National Institute of Standards and Technology, Testing and Validation of the Usability of Electronic Health Records, 2012.
[12] C.J. Vincent, A. Blandford, Usability standards meet scenario-based design: challenges and opportunities, J. Biomed. Inform. 53 (2015) 243–250.
[13] R.M. Ratwani, N.C. Benda, A.Z. Hettinger, et al., Electronic health record vendor adherence to usability certification requirements and testing standards, JAMA 314 (10) (2015) 1070–1071.
[14] Office of the National Coordinator (ONC) for Health Information Technology, Department of Health and Human Services, 2015 ed. Health Information Technology (Health IT) Certification Criteria, 2015 ed., Base Electronic Health Record (EHR) Definition, and ONC Health IT Certification Program Modifications, Final rule, Fed Regist 80 (200) (2015) 62601–759.
[15] 2014 ed., Test Scenarios Development Overview, slide presentation: The Office of the National Coordinator for Health Information Technology, Sept 11, 2013.
[16] 2014 ed. Test Scenarios Development Overview Companion Narrative: the Office of the National Coordinator for Health Information Technology, Sept 11, 2013.
[17] Better EHR: usability, workflow,and cognitive support in electronic health records, in: J. Zhang, M. Walji (eds.), The National Center for Cognitive Informatics and Decision Making in Healthcare, SHARPC, 2014 403.
[18] Testing and Test Methods: 2015 ed., Draft Test Scenarios. Office of the National Coordinator (ONC) for Health Information Technology. Retrieved December 18, 2017, from: <https://www.healthit.gov/policy-researchers-implementers/2015-edition-test-method>.
[19] R.M. Schumacher, S.Z. Lowry, Customized Common Industry Format Template for Electronic Health Record Usability Testing: National Institute of Standards and Technology, U.S. Department of Commerce, 2010, p. 37 Report No. 7742.
[20] Applying Human Factors and Usability Engineering to Medical Devices: Guidance for Industry and Food and Drug Administration Staff: U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health, Office of Device Evaluation, 2016 pp. 1–45.
[21] Medical devices – Part 1: application of usability engineering to medical devices, International Organization for Standardization, 2015 96.
[22] Medical devices – Part 2: guidance on the application of usability engineering to medical devices, International Organization for Standardization, 2016 94.
[23] M.A. Ellsworth, M. Dziadzko, J.C. O'Horo, et al., An appraisal of published usability evaluations of electronic health records via systematic review, J. Am. Med. Inform. Assoc. 24 (1) (2017) 218–226.
[24] K.B. Bayley, L.A. Savitz, T. Maddalone, et al., Evaluation of patient care interventions and recommendations by a transitional care pharmacist, Therap. Clin. Risk

Manage. 3 (4) (2007) 695–703.
[25] T.M. Nester, L.S. Hale, Effectiveness of a pharmacist-acquired medication history in promoting patient safety, Am. J. Health-System Pharm. 59 (2001) 2221–2225.
[26] K.M. Gleason, M.R. McDaniel, J. Feinglass, et al., Results of the medications at transitions and clinical handoffs (MATCH) study: an analysis of medication reconciliation errors and risk factors at hospital admission, J. Gen. Intern. Med. 25 (5) (2010) 441–447.
[27] M.L. Wess, J.J. Saleem, J. Tsevat, et al., Usability of an atrial fibrillation anticoagulation decision-support tool, J. Primary Care Commun. Health 2 (2) (2011) 100–106.
[28] J.J. Saleem, E.S. Patterson, L. Militello, et al., Impact of clinical reminder redesign on learnability, efficiency, usability, and workload for ambulatory clinic nurses, J. Am. Med. Inform. Assoc. 14 (5) (2007) 632–640.
[29] S. Wu, M.R. Lehto, Y. Yih, et al., Impact of clinical reminder redesign on physicians' priority decisions, Appl. Clin. Inform. 1 (4) (2010) 466–485.
[30] J.A. Anderson, K.M. Godwin, J.J. Saleem, et al., Accessibility, usability, and usefulness of a Web-based clinical decision support tool to enhance provider-patient communication around Self-management TO Prevent (STOP) Stroke, Health Inform. J. 20 (4) (2014) 261–274.
[31] D.A. Haggstrom, J.J. Saleem, A.L. Russ, et al., Lessons learned from usability testing of the VA's personal health record, J. Am. Med. Inform. Assoc. 18 (Suppl 1) (2011) i13–i17.
[32] A.L. Russ, A.J. Zillich, B.L. Melton, et al., Applying human factors principles to alert design increases efficiency and reduces prescribing errors in a scenario-based simulation, J. Am. Med. Inform. Assoc. 21 (e2) (2014) e287–e296.
[33] A.L. Russ, et al., Formative Usability Evaluation of a Novel Tool for Medication Reconciliation Human Factors and Ergonomics Society, International Annual Meeting, 2017, Austin, TX.
[34] A. Russ, D. Baker, W. Fahner, et al., A Rapid Usability Evaluation (RUE) method for health information technology, American Medical Informatics Association (AMIA) Symposium 2010 pp. 702–706.
[35] R. Hartson, P.S. Pyla, The UX Book: Process and Guidelines for Ensuring a Quality User Experience, Morgan Kauffman (Elsevier), 2012.
[36] T. Tullis, B. Albert, Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics, Morgan Kaufmann Publishers, 2008.
[37] D. Hix, H.R. Hartson, Developing User Interfaces: Ensuring usability throughout product & process, John Wiley & Sons Inc, New York, 1993.
[38] D.A. Haggstrom, J.J. Saleem, A.L. Russ, et al., Lessons learned from usability testing of the VA's personal health record, J. Am. Med. Inform. Assoc. 18 (Suppl 1) (2011) i13–i17.
[39] A.W. Kushniruk, E.M. Borycki, S. Kuwata, et al., Emerging approaches to usability evaluation of health information systems: towards in-situ analysis of complex healthcare systems and environments, Stud. Health Technol. Inform. 169 (2011) 915–919.
[40] Nielsen Norman Group, Turn User Goals into Task Scenarios for Usability Testing" January 12, 2014. Retrieved July 20, 2017 from: <https://www.nngroup.com/articles/task-scenarios-usability-testing/>.
[41] U.S. Department of Health and Human Services. Usability.gov Usability Basics: Scenarios, Retrieved July 31, 2017 from: <https://www.usability.gov/how-to-and-tools/methods/scenarios.html>.
[42] E.S. Patterson, M.L. Rogers, M.L. Render, A simulation-based embedded probe technique for human-computer interaction evaluation, Cogn. Tech. Work 6 (2004) 197–205.
[43] A.R. Pritchett, R.J. Hansman, Use of testable responses for performance-based measurement of situation awareness, in: M.R. Endsley, D.J. Garland (Eds.), Situation awareness analysis and measurement, 2000 pp. 189–209.
[44] Standards and Operational Guidance for Ethics Review of Health-Related Research with Human Participants, WHO Document Production Services, Geneva, Switzerland, 2011, p. 41.
[45] Lewis JR. Section 2.4.1 Think Aloud in Chapter 49: Usability Testing, in: G. Salvendy (Ed.), Handbook of Human Factors, third ed., Hoboken, NJ: John Wiley & Sons, Inc, 2006 pp. 1282.
[46] J.J. Saleem, E.S. Patterson, L. Militello, et al., Exploring barriers and facilitators to the use of computerized clinical reminders, J. Am. Med. Inform. Assoc. 12 (4) (2005) 438–447.
[47] J.J. Saleem, A.L. Russ, A. Neddo, et al., Paper persistence, workarounds, and communication breakdowns in computerized consultation management, Int. J. Med. Inform. 80 (7) (2011) 466–479.
[48] S.K. Card, T.P. Moran, A. Newell, The Psychology of Human Computer Interaction, LEA, Hillsdale. NJ, 1983.
[49] J.M. Devoge, E.J. Bass, R.M. Sledd, et al., Collaborating with physicians to redesign a sign-out tool: an iterative, multifaceted approach with users – even busy ones – can yield a satisfying and efficient product, Ergon Des. 17 (1) (2009) 20–28.
[50] A.L. Russ, D.A. Baker, W.J. Fahner, et al., A rapid usability evaluation (RUE) method for health information technology, AMIA Annu. Symp. Proc. 2010 (2010) 702–706.
[51] A.L. Russ, A.J. Zillich, M.S. McManus, et al., Prescribers' interactions with medication alerts at the point of prescribing: a multi-method, in situ investigation of the human-computer interaction, Int. J. Med. Inform. 81 (4) (2012) 232–243.
[52] A.L. Russ, S. Chen, B.L. Melton, et al., A novel design for drug-drug interaction alerts improves prescribing efficiency, Jt. Comm. J. Qual Patient Saf. 41 (9) (2015) 396–405.
[53] J.W. Creswell, D.L. Miller, Determining validity in qualitative inquiry, Theory into Pract. 39 (3) (2000) 124–130.
[54] E.H. Bradley, L.A. Curry, K.J. Devers, Qualitative data analysis for health services research: developing taxonomy, themes, and theory, Health Serv. Res. 42 (4) (2007) 1758–1772.
[55] R. Shawahna, D. Masri, R. Al-Gharabeh, et al., Medication administration errors from a nursing viewpoint: a formal consensus of definition and scenarios using a Delphi technique, J. Clin. Nursing 25 (3–4) (2016) 412–423.
[56] M.X. Cicero, L. Brown, F. Overly, et al., Creation and Delphi-method refinement of pediatric disaster triage simulations, Prehosp Emerg. Care 18 (2) (2014) 282–289.
[57] E.J. Bass, et al., A method for investigating real-time distributed weather forecaster-emergency manager interaction. Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference, 2011, pp. 2809–2815.
[58] C.D. Chisholm, A.M. Dornfeld, D.R. Nelson, et al., Work interrupted: a comparison of workplace interruptions in emergency departments and primary care offices, Ann. Emerg. Med. 38 (2) (2001) 146–151.
[59] C.M. Kelley, A.C. McLaughlin, Differences in feedback use for correct and incorrect responses, Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting, 2012, pp. 2427–2431.
[60] A.L. Russ, B.L. Melton, J.K. Daggy, et al., Pilot evaluation of a method to assess prescribers' information processing of medication alerts, J. Biomed. Inform. 66 (2017) 11–18.
[61] T.H. Altom, The Effect of Think-Aloud on Timed Performance in Usability Testing (thesis for Masters of Science), School of Informatics, Indiana University, 2006, p. 54.
[62] M. Hertzum, K.D. Holmegaard, Thinking aloud influences perceived time, Human Factors 57 (2015).
[63] M. Van Den Haak, M De Jong, P. Jan Schellens, Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue, Behav. Inform. Technol. 22 (5) (2003) 339–351.
[64] A.W. Kushniruk, V.L. Patel, Cognitive and usability engineering methods for the evaluation of clinical information systems, J. Biomed. Inform. 37 (1) (2004) 56–76.
[65] J. Trafton, S. Martins, M. Michel, et al., Evaluation of the acceptability and usability of a decision support system to encourage safe and effective use of opioid therapy for chronic, noncancer pain by primary care providers, Pain Med. 11 (4) (2010) 575–585.
[66] A.L. Russ, M. Weiner, S.A. Russell, et al., Design and implementation of a hospital-based usability laboratory: insights from a Department of Veterans Affairs laboratory for health information technology, Jt. Comm. J. Qual Patient Saf. 38 (12) (2012) 531–540.
[67] J. Rubin, Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests, John Wiley & Sons., 1994, p. 26.
[68] J. Nielsen, Usability engineering, Academic Press, Boston, MA, 1993, p. 362.
[69] Healthcare Failure Mode and Effect Analysis (HFMEA), Veterans Affairs National Center for Patient Safety. Retrieved October 2, 2017 from: <https://www.patientsafety.va.gov/professionals/onthejob/hfmea.asp>.
[70] J.J. Saleem, A.L. Russ, P. Sanderson, et al., Current challenges and opportunities for better integration of human factors research with development of clinical information systems, Yearb Med. Inform. (2009) 48–58.
[71] J.J. Saleem, L.G. Militello, A.L. Russ, et al., The need for better integration between applied research and operations to advance health information technology, Healthc (Amst) 4 (2) (2016) 80–83.
[72] R.M. Young, P. Barnard, The use of scenarios in human-computer interaction research: turbocharging the tortoise of cumulative science, CHI '87 Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface. Toronto, Canada, 1987 pp. 291–296.
[73] R. Young, P. Barnard, Signature tasks and paradigm tasks: new wrinkles on the scenarios methodology, People and Computers VI, Proceedings of the HCI '91 Conference August 20, 1991, pp. 20–23.
[74] M.L. Bolton, E.J. Bass, Formally verifying human-automation interaction as part of a system model: limitations and tradeoffs, Innov. Syst. Softw. Eng. 6 (3) (2010) 219–231.
[75] P.Masci, et al., Using PVSio-web and SAPERE for rapid prototyping of user interfaces, in: Integrated Clinical Environments, Workshop on Verification and Assurance, co-located with CAV, 2015.
[76] R. Ruksnas, P. Curzon, A. Blandford, et al., Combining human error verification and timing analysis: a case study on an infusion pump, Formal Aspects Comput. 26 (5) (2014) 1033.
[77] A.R. Pritchett, S.Y. Kim, K.M. Feigh, Measuring human-automation function allocation, J. Cognitive Eng. Decision Making 8 (1) (2014) 52–77.
[78] A.R. Pritchett, S.Y. Kim, K.M. Feigh, Modeling human-automation function allocation, J. Cognitive Eng. Decision Making 8(1) 2014b 33–51.
[79] M.L. Bolton, S. Göknur, E.J. Bass, Framework to support scenario development for human-centered alerting system evaluation, IEEE Trans. Human-machine Systems 43 (6) (2013) 595–607.
[80] M.L. Bolton, N. Jiménez, M.M. Van Paassen, et al., Automatically generating specification properties from task models for the formal verification of human-automation interaction, IEEE Trans. Human-Machine Systems 44 (5) (2014) 561–575.