

Developing the Most Effective Clinical Scenarios for Usability Testing

Jane Robbins, ICF HFE, Kurt Ruark, VA HFE, Ross Speir, VA HFE, Ashley Cook, CSRA HFE, Danielle Hoover, MD VA, and Abigail Noonan, ArcSource HFE

Background

The range of health information technology (HIT) software at the Veterans Health Administration (VHA) includes websites, Computerized Patient Record System (CPRS) templates, commercial off the shelf (COTS) products, and mobile applications, any of which could be part of an electronic health record system. Usability testing is an effective method used at the VHA to determine how efficiently and effectively users can accomplish tasks with HIT systems. In a usability test, representative users carry out key tasks within the HIT system, so an essential component of the usability test is the sequencing of tasks described within the context of a scenario, or the setting in which the test takes place. Testing with realistic and carefully crafted scenarios is important to evaluate any HIT system, but it's particularly important to test the usability of systems intended for clinicians.

Well-crafted scenarios are critical because user tasks must be clearly understood by clinicians participating in the usability test. Scenarios also need to convey contextual details of the tasks and often the clinical environment in which tasks will be carried out. This paper reviews some lessons learned for creating clinical scenarios from VHA's Human Factors Engineering (HFE) team.

Keeping clinician participants such as doctors, nurses and psychologists motivated to complete tasks during testing is sometimes a challenge, since they are extremely busy and focused on their patient workload. Scenarios must be written broadly enough to cover differing processes at VA Medical Centers (or different ways that individual clinicians practice medicine), yet granular enough to address the choices being tested. Additionally, the task descriptions and data used must be realistic.

Usability tests generally analyze failure rates such as time on task and errors -- metrics obtained to determine whether users are able to interact with the interface. This data tells the researcher how usable the software is in a realistic situation, and forms the basis for recommendations to improve it.

Best practices for writing scenarios include setting the context of the clinical setting, stating the patient's situation briefly but completely, and explaining the treatment plan action and tasks in a logical order. Scenarios should be written in plain unambiguous language without extraneous information, unless including distractions or clinical alerts are the specific intent of the testing. Researchers should follow best practices to develop effective scenarios for their clinician-facing usability studies.

Considerations for Crafting Scenarios

In this paper we will explore the criteria for effective scenarios and apply these criteria to several best practice examples. Additionally, scenarios that were not as effective are shown, along with scenario-related issues that were noticed.

Worksheet Job Aid. Vanderbilt University researchers (Reale, Anders & Weinger, 2016) developed a worksheet to allow researchers to narrow down different aspects of a clinical scenario. This is an excellent starting point to determine the questions to ask about the system being tested, and ultimately develop scenarios based on those questions. This can also promote discussion between the usability researcher, clinical Subject Matter Experts (SMEs) and business owners to ensure they are all in agreement about the focus of the research.

Clinical Scenario Worksheet: Developing a Usability Testing Scenario	
Define the clinical goal(s) and user needs of relevance to this scenario.	
What type(s) of user testing will this scenario support?	
Which user group(s) are the primary users? What are their key characteristics?	
What are the critical tasks (high risk or high cost) that are tested in this scenario?	
What are the common tasks tested in this scenario (i.e., those that support use efficiency and effectiveness)?	
What aspects of the clinical task have the highest potential impact on patient safety?	
What aspects of the UI design pose the highest risk for patient safety-related errors?	
What aspects of the UI design pose the highest risk for use errors?	
What elements of the UI design or system functionality raise usability concerns among our team?	
What deviations from or interruptions in the ideal workflow will users typically encounter when performing these tasks?	
What elements of the UI design are most important for user satisfaction?	

Granularity. In writing usability testing task scenarios, provide the proper level of context and detail for the task for every type of participant, especially for clinicians, who typically make decisions based on a complex set of patient data. If not enough information is given in the task description, participants may fall into clinical decision-making mode, rather than pay attention to the task that they've been asked to perform. But if there is too much information about the task or patient involved, the participant may be distracted by their medical opinion for a patient.

Plain Language. Directly instruct the clinician to disregard clinical decision-making and instead focus on how they would perform the task. Straightforwardly stating tasks in plain language and highlighting that the task is not a medical decision point are important. Direct language is ideal, such as "You have a 45-year-old male patient with Type 1 diabetes. Task: Use the template to order an X test." This also reinforces the need to perform the task within the scenario and not make a clinical diagnosis.

Provide Context. Quickly provide context of use within the scenario. Ideally, each scenario should be no more than a few sentences long, to enable clinicians to retain the information as they complete tasks. However, for electronic healthcare record (EHR) testing, sometimes it is necessary to provide a more detailed scenario. Providing contextual information for a task that is medically consistent and/or a likely patient issue is essential, to prevent the clinician from focusing on the inconsistency rather than on the task.

Range of Scenarios. Consider developing scenarios that cover "not only common use scenarios, but also uncommon ones (i.e., edge cases), especially if those use scenarios are safety-related." (Wiklun, Kendler, Hochberg & Weinger, 2015, p. 17). It is important to test the main functionality of the software, but the test should also include one or more scenarios or tasks related to less common items, to more fully exercise the software. Although these recommendations are made for EHRs by the National Institute of Standards and Technology (NIST), they easily apply to most healthcare systems and applications.

The NIST authors also say that "Common use scenarios and their associated details can be identified by speaking with EHR users about how they use their current EHRs or how they currently perform tasks that will eventually be performed

using an EHR under development. Direct observations are useful in both identifying potential use scenarios and in confirming scenarios derived from user interviews.” (Wiklun et al., 2015, p. 18).

Flow & Order. Another consideration to keep in mind is to focus the scenarios and tasks in a logical flow true to the site or app, to keep it more realistic for the clinician. Additionally, put tasks within a scenario which are more likely to be complex and lead to failure at the end of the test, so that the clinician doesn’t get discouraged and acquire a negative bias for other tasks. (User Testing, 2017, p. 9).

Participant Information. NIST suggests to not direct participants to learning aids or other information that would not be available to a user in the real world. (Wiklun et al., 2015, p. 42). For extremely complex scenarios, sending the scenarios and/or tasks ahead of time to participants may be helpful so that they can be printed out and used during the testing. However, this is advisable only for complex scenarios for applications not familiar to the clinician participants.

Workflow & Care Types. Scenarios should test multiple categories of patient care that the clinician is likely to encounter using an application, depending on the application being tested. For an EHR, these categories can include “acute episodes, treatment of chronic conditions, and recommended preventative and health promotion activities.” (Armijo, McDonnell and Werner, 2009, p. 6) Further, NIST recommends investigating users’ workflows, cognitive task requirements and mental models. (Wiklun et al., 2015, p. 110). An example of the environmental aspects of the clinical and patient flow from Armijo, McDonnell and Werner (2009, p. 36) is shown below.

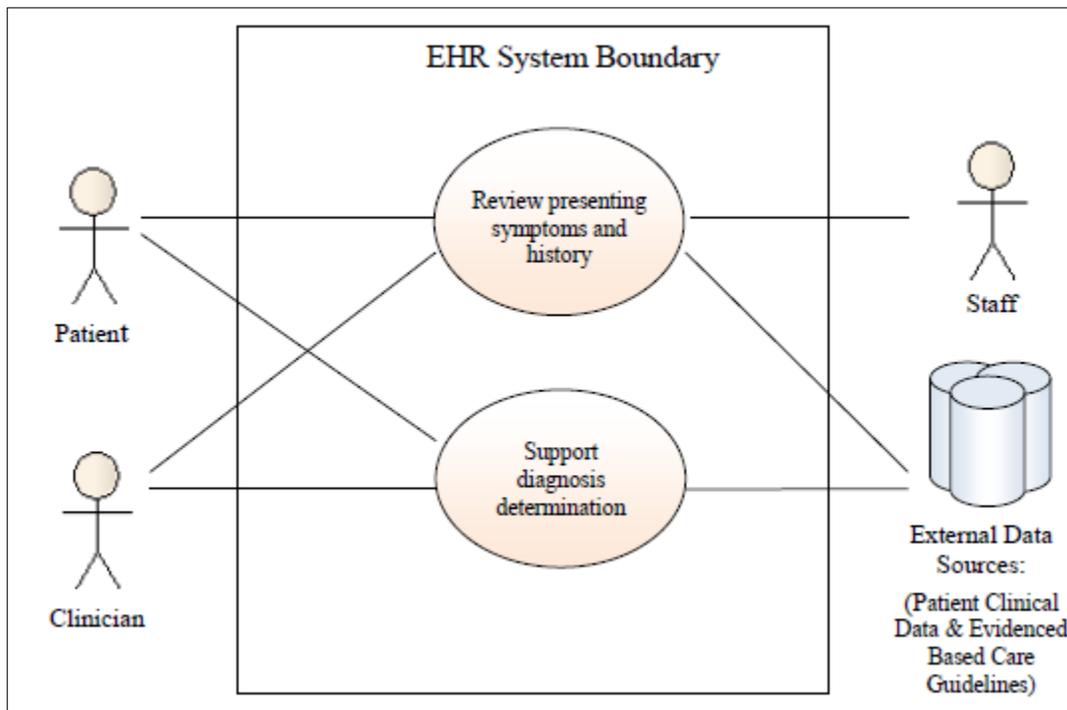


Figure 1: An example of a new or unrelated symptom use case (from Armijo, McDonnell & Werner 2009, p. 36)

Qualitative vs. Quantitative Data. Pay attention to the type of data being collected during testing --quantitative metrics or qualitative feedback. If quantitative data is the prime objective, the scenarios should support a rigorous set of typical and possibly edge tasks. In cases where qualitative information is paramount, a think-aloud protocol is often used, as well as questionnaires to determine participant motives and fit within contextual flow as participants interact with application interface. Many tests gather both types of data, so the practitioner must combine several collection methods.

Personas. Use personas as patients or fellow clinicians. If there are established personas available and they are appropriate to the testing, using them as patients or clinicians (in an interactive situation) is desirable. This allows persona stories to be built out and provides realistic patients for participants to work with, since personas are based on data for the user group.

Stressors. If the application being examined warrants it, the researcher may want to incorporate stressors and alerts into scenarios or tasks. In today's busy clinical environment, "the form and timing of information presentation must respect the risks of break-in-task and loss of activation events that can be caused by introducing competing tasks and distracting information into the already-saturated workflow." (Armijo, McDonnel & Werner, 2009, p. 10) This allows a more realistic environment for completion of clinical tasks, but it may tend to distract participants from their focus on the test scenario. Some applications may need to be tested under stressors, for patient safety reasons inherent to that clinical software that warrant the distraction from the test scenario.

Scenario Reviews

Clinical SMEs: After the usability researcher drafts scenarios, they should be reviewed by one or more clinical SMEs, often the project sponsors for the project. If the SME is not familiar with the project, the researcher must clearly establish and communicate the goals of the testing so the SME can translate them into accurate scenarios. Additionally, the clinical SME should be representative of the test participants, since some terms that are familiar to medical specialists may not be known by primary care providers (PCPs), for example.

Procedurally, if the SME is familiar with the project, it may be optimal for him/her to draft the scenarios first and then send to the human factors researcher. If the project sponsor is not a SME, clinicians must be recruited to assess the scenarios for realism and medical accuracy.

Examples of Good Scenarios

Examples of well-crafted scenarios from recent HFE tests incorporated the considerations noted above.

Scenario (Locate Patient Record): You are Dan's primary care provider at his outpatient clinic. You are preparing to see Dan and want to pull up his record to prepare to document his visit. You've already logged in to the new EHR and have reached the new EHR home page. **Task:** From the new EHR homepage, navigate to Dan's patient record.

The scenario above examined an early version of a new EHR. What stands out about it is its directness in setting the context by telling the participant that they are already logged in to a particular screen. Using the word "You" leaves little doubt about what the participant needs to do. This scenario also utilizes an HFE persona.

Scenario (Review Diabetes Information in the Record): At the end of the prior visit, the plan was to get a fasting blood sugar (BS) in the office, collect a blood specimen to do a lipid panel & HbA1c, have an intake nurse get vital signs, including weight, do a diabetic foot exam, and talk with the patient to get an intervening history. **Task:** Use the system to review documentation related to Dan's diabetes treatment and management. Be sure to review his medications, labs, past vitals readings and any pertinent notes.

This is another good EHR scenario involving a set of five items for provider participants to review, but displayed them in a list-like format that participants could work through to complete the task.

Scenario (Multiple Problem Scenario): You're seeing a new patient who is a 50 year old man with a growth on his nose for 1 year, and a growth on his right ear for 6 months. They bleed when he picks at them. He hasn't been evaluated for these particular lesions before by anyone. He's seen a dermatologist at another VA facility in the last 5 years, and remembers being treated for skin cancer but no details are recalled. There is no family history of melanoma. He has had a lung transplant and is immunosuppressed and on multiple medications, but none started recently. **Task:** You need input from a Dermatologist on this patient to assist with diagnosis and management. Using CPRS and the template that has been opened for you, order a TeleDermatology imaging consult request.

This scenario from a TeleDermatology study of PCPs that compared two CPRS templates provided the PCP participants just enough information to complete the rather extensive template, but no extraneous data.

Scenario 1 (Phase 1): A fine needle aspiration and biopsy needs to be completed on a 65-year-old female patient who does not have a DNR [Do Not Resuscitate] order and an anesthesia provider will not be present.

This scenario from a Cognitive Walkthrough of a consent software wizard clearly explained the patient's situation in plain language, although it would be better in active voice. The simplicity of combining both the scenario context into the task ensures participant understanding.

Examples of Scenarios That Could Distract Some Clinicians

Some scenarios used in previous HFE studies could have been more effective.

Scenario: A 50 year old male patient without previous colorectal cancer screening (CRCS) is seen in primary care by you, his PCP. FIT [fecal immunochemical test] screening is preferred but the patient is acutely ill with a GI [gastrointestinal] illness. You do not believe the screening should be done now.

In this study examining CPRS Colorectal Cancer screening templates for PCPs, one scenario was seen as confusing by some clinicians who participated in usability testing. They complained that "The diagnosis of abdominal issues complicates the thought process on how to determine this deferral." and "I was thrown by the 4-month time frame [which was the choice on the template]. May [want to] offer a more generic time frame like 'next visit'." Additionally, some PCPs were not familiar with the acronym "FIT" and this term distracted them. Improvements to the scenario could include more specifics on the desired time frame for clinician action (rather than the vague "done now") and a definition of "FIT".

Scenario 1 (Phase 2): A simple procedure that does NOT pose a high risk of serious arrhythmia or cardiopulmonary arrest (e.g., hemodialysis, cardiac catheterization, electrophysiology studies), NOR does it require anesthesia care by anesthesia providers.

The previous scenario from a Cognitive Walkthrough of some consent software includes a double negative, which might confuse some users. It also does not mention a patient and gives a broad range of conditions for the clinician to consider, whereas possibly one or two examples would have sufficed.

Scenario: Please create a three-word note (content is not important) with the title "Primary Care General Note." Please sign the note using the credentials "VK1234!!" Once you feel you've signed the note, please see if there are any unsigned notes, if there are, please sign those as well.

Clinicians were confused in an EHR study, even though they were only asked for a three word answer. The somewhat contradictory language and long series of tasks and even a tangent to check for unsigned notes, caused participants to want to write something clinically realistic, longer than three words.

Tasks and Data

Once scenarios have been developed, the usability testing team must create descriptions that outline the tasks that testing participants perform. Multiple tasks can be written for a single scenario, or each task can have its own scenario.

Tasks. For some usability tests, a single scenario can be used, with multiple tasks underneath it. This has the advantage of setting the scene once and allowing users to focus more on the tasks within that scenario, potentially saving time and reducing cognitive load. Developing tasks to use within a scenario requires a deep understanding of the user's workflow within the context of that scenario. Complex tests, or those where very different functionalities are being evaluated, may necessitate more than one scenario.

Challenge the System. The NIST researchers (Lowry et al. 2012, p. 44) "recommend using test cases that are developed so that they can be evaluated on both a clinical and usability level and include challenging scenario elements such as nonstandard..." and "imprecise, incomplete, and inaccurate data." Additionally, users should focus on "the highest-priority (the most safety-critical and all essential) user tasks." (Lowry et al. 2012, p. 61)

Realistic Data. It is also very important that data described in scenarios and used in tasks are realistic. Clinicians can easily be distracted if the data is not within the normal range expected for patients, and often this precludes them from focusing on the task. For example, in a test of an antibiogram application (where the provider selects the most appropriate antibiotic for an infection) one of the antibiotic names used during the testing is no longer typically used in hospitals, as it can cause fatal reactions. Participants noticed this during the testing, and were distracted by it, commenting on its presence rather than facets of the wireframes.

Conducting Testing

Dry Runs. Setting up a good dry run is extremely important to ensure the success of testing. Dry runs of the testing scenarios and tasks should include SMEs knowledgeable about that clinical area. The dry run is also an opportunity to validate the data and ensure that it fits the scenarios and tasks and can be utilized by the participant.

Actors. In test scenarios where interaction with others is necessary, an actor can be used to act the part of a patient or other person involved with the clinician. Two recent HFE studies have successfully used actors (generally VHA employees) to act as patients or spouses of patients. Incorporating actors into the scenario allowed the clinicians being tested to interact with someone in a more natural way than just listening to or reading instructions about the patient.

Analyzing Testing Results

Metrics. The analysis of task metrics is one of the most important quantitative indicators of issues that may exist for the application. Usability researchers can code each step taken by the participant in order to break down the choices made, as well as to record deviations from the optimal path for the task.

Among the test metrics – besides traditional ones like task completion and success rates – are identification of qualitative workarounds, redundancies or clinician exhaustion noted while performing the tasks. (Lowry et al., 2012, p. 18) Analysis of the data also allows the researchers to determine “errors of omission as well as commission.” The example given concerns a clinician not noticing vital information (omission) or selecting the wrong drug due to ambiguous abbreviations (commission). (Lowry et al. 2012, p. 27)

Data Summary. A useful tool for analysis is provided in the Common Industry Format for Usability Test Reports (Lowry et al., 2012, page 56). Researchers can use a table like this to summarize issues found in testing for each task. Researchers need to devise analysis methods that make sense for their study. Note that time may or may not be relevant, depending on the type of testing or if comparison against a known baseline is a goal.

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

Figure 2: Example of tabular data reporting of validation study results. (From NISTIR 7742)

Questionnaires. NIST research also suggests that researchers include a question for participants to offer what they perceive as “required and missing information” after finishing the task. If the participant did not enter required information, a validation error is triggered in most applications. And the usability test should include a prompt of participants for any missing or confusing information related to the scenario or task description. (Wiklun et al., 2015, p. 68)

Comments. Examining comments provided by participants is often helpful in determining how they approached the scenarios, if they understood them, and if there were any distracting elements. Comments may be solicited by using a think-aloud protocol and/or via questionnaires after task completion. Almost every usability test reveals surprises about the mis-steps that participants make outside of the optimal path, as well as observations that were not previously noticed about the scenarios, tasks, or system being tested. The sequence that led to the missteps and the participant's logic, as seen in their remarks, are useful in identifying potential usability issues.

Conclusions

Realistic usability test scenarios are essential to evaluating how a proposed health IT system design will support clinicians in carrying out tasks effectively and efficiently. A well-crafted scenario will capture and maintain participants' focus, enable the participant to understand the goal and underlying tasks to achieve that goal, allow usability engineers to measure task performance, and help the project team improve the usability of health IT designs. Using the scenario best practices described in this paper should greatly improve the quality of scenarios for clinicians and thereby provide insightful usability data about the application being tested.

References

- Armijo, D., McDonnel, C. and Werner, K. (2009). *Electronic Health Record Usability. Evaluation and Use Case Framework*. AHRQ Publication No. 09(10)-0091-1-EF.
- HIMSS. *EMR Usability Evaluation Guide For Clinicians' Practices. Sample Usability Task Scenarios*. (undated).
- Lowry, S. et al. (2012). *Technical Evaluation, Testing and the Validation of the Usability of Electronic Health Records*. NISTIR 7804. National Institute of Standards and Technology.
- Johnson, C. et al. (2011). *EHR Usability Toolkit: A Background Report on Usability and Electronic Health Records*. AHRQ Publication No. 11-0084-EF.
- Reale, C., Anders, S., Weinger, M. (2016). *eHMP UX Guide Recommendations Clinical Scenarios*. Center for Research and Innovation in Systems Safety, Vanderbilt University.
- Taylor, K. T. and Herout, J. (2015). *eHMP Workflow Study. User Personas and Scenarios*. VHA HFE.
- UserTesting.com. (2017). *The complete guide to usertesting websites, apps and prototypes*. http://downloads.usertesting.com/white_papers/UT_Complete-Guide-to-User-Testing.pdf. UserTesting.com.
- Wiklun, M., Kendler, J., Hochberg, L, Weinger, M., (2015). *Technical Basis for User Interface Design of Health IT*. NIST GCR 15-996. National Institute of Standards and Technology.

Glossary

Primary and secondary users: The primary users of an application or website are those for whom the app or site was designed. Secondary users are others who may use the system on occasion, but they generally may have less experience with the system and their complete needs are not always taken into consideration in system design. An example is an app designed for PCPs that is also accessed on occasion by nurses.

Critical tasks: Tasks that are necessary to perform the main goals of the app or site, especially those involving patient safety. For example, an app may be designed to allow clinicians to view patient records, but also allow entry of contact data by the clinician. The critical task in this example is to view the records but a non-critical task would be entry of contact data.

Scenarios: The setting and framework under which a usability test takes place. Usability testing depends on scenarios to be clearly explained to test participants, who then carry out tasks using the context of the scenario.