Usability Testing of Medical Technology

Human Factors Research & Design

Erin Davis
Associate Research Director
Human Factors Research & Design
erin.davis@ul.com

Allison Strochlic
Research Director
Human Factors Research & Design
allison.strochlic@ul.com

EMERGO
by UL

November 2020

Executive Summary

Usability testing of medical products calls for target customers (i.e., the intended users) to perform tasks with a given product to assess its interactive qualities. Desirable interactive qualities include such attributes as ease of learning, ease of use, and ease of avoiding and recovering from a use error. The primary goal of a test is to assess whether users can operate a given device safely and effectively. However, testing may also serve to assess a device's usability and appeal, qualities that might not be prioritized when applying human factors to meet regulators' expectations but that have commercial relevance. Accordingly, it is imperative to conduct usability tests on the full range of medical products (including medical devices, combination products, and in vitro diagnostic devices (IVDs)) to ensure that they are ready for real-world use. Such testing, conducted early and often repeatedly throughout the design process, is a proven route to product acceptance by regulatory authorities and by customers. Notably, regulators frequently defer their approval or even reject applications to market a medical product when a developer does not provide sufficiently positive usability test data. Also, customers seeking a good user experience may turn away from products that are not user-friendly and gravitate toward better alternatives.

Usability tests call upon representative users – people who are, or have a similar background to, the product's intended users – to "test-drive" a given product. Over the course of multiple sessions, usability testing professionals watch for signs that subjects can interact smoothly and efficiently with a product. Test facilitators watch for and record instances when product features appear to induce use errors, cause confusion, or make the users feel frustrated. Of course, designers and engineers try to make their products as easy, safe, and appealing as possible by applying their personal experience and knowledge, along with design best practices derived from industry standards and guidelines.

However, developers cannot ensure that their products truly match users' needs and expectations until the intended users literally get their hands on a prototype. Therefore, it is wise to conduct usability tests that quickly reveal user interface strengths and opportunities for further product improvement. Conversely, skipping early and repeated testing can be perilous to project budgets and product launch schedules. It is also wise and usually necessary to conduct a final usability test of the finished product to be sure that it is ready to go to market with appropriate certifications and approvals. The US Food and Drug Administration (FDA) refers to this final form of evaluation as *human factors validation testing*, which is also known as *summative usability testing*.

The value of usability testing

Simply stated, usability testing gives representative users the chance to interact with a product to see how well it works for themselves. This is an opportunity to see how users go about performing tasks, to see what challenges they face and what mistakes (i.e., use errors) they make, and to collect their opinions about the product. Usability testing can expose a wide range of design flaws that might not have been obvious to the product development team. Here is a sample of interaction problems that we have observed in the course of testing medical products and attributed to user interface design flaws:

- A pharmacist selected the wrong bottle of pills from among three options because each bottle had the same color label, despite that fact that the pill dosages varied
- A patient struggled to follow the correct steps to operate a product because the instructions for use neither clearly delineated the steps nor presented information in a clear hierarchy of importance
- A nurse connected a tube to the wrong port because there was nothing preventing the mistake, such as color, shape, and tactile coding
- A patient misinterpreted an illuminated green light to mean that the current task was complete, rather than still in progress

As exemplified above, usability testing gives the gift of user feedback, which subsequently can help you design a better product. Testing with users is likely to uncover flaws across different portions of the user interface, which then gives you the opportunity to make changes to correct the flaws and mitigate risk. This pattern holds for all kinds of products. In the case of medical devices, the major objective is to ensure that users will use the product correctly and safely once it is on the market. Conducting usability testing and iteratively revising the design until the device is ready to be commercialized has many benefits, such as those presented in the graphic below.

Improved sales
(and lower development cost, due to less re-work)

Satisfied users
(better user experience)

Faster to market
(successful submission)

Simpler training
(and lower demand for customer support, due to increased usability)

Reduced liability
(fewer use errors)

Regulatory requirements

Medical device regulators, such as the US Food and Drug Administration (FDA), have specific expectations for how usability testing should be conducted. When testing a medical device, one is very interested in learning about the overall user experience: what people like and do not like, what makes it easy rather than difficult to use, and how the product can be improved to better meet users' needs. But, the primary goal is ultimately to ensure that the product is safe and effective.

There are some nuanced differences regarding how to implement the human factors engineering (HFE) process in accordance with the expectations of different regulatory bodies. Although the FDA has its own guidance that manufacturers should follow when seeking to market their products in the US, the International Electrotechnical Commission (IEC) provides a general, non-nation-specific blueprint for applying HFE to medical devices in IEC 62366-1, which is widely accepted by other regulatory bodies, such as the European Medicines Agency (EMA) and the UK's Medicines and Healthcare products Regulatory Agency (MHRA).

In the guidances from the aforementioned sources, medical device manufacturers are strongly encouraged, but not officially required, to conduct formative usability tests. And, despite formative tests not being required *per se*, it is rare that a medical device can "pass" an HF validation test if it has not undergone one or more rounds of formative testing and iterative design improvement. HF validation tests, on the other hand, are generally required in cases where medical device use could lead to harm for end-users or others (e.g., patients, in the case of a device used by clinicians). Regulatory bodies and standards call upon manufacturers to manage use-related risk, and conduct usability tests – HF validation tests in particular – to confirm that the manufacturer has reduced risk to an acceptable level.

The FDA tends to uphold the highest and most rigorous expectations for HFE in general, and especially for a pivotal HF validation test. As such, if you conduct usability testing in a way that aligns with FDA's expectations, you will likely be in "good shape" in terms of developing a safe and effective device. Furthermore, you can probably leverage the data from the US-based usability tests to support regulatory submissions outside of the US.

Formative versus HF validation (a.k.a. summative) usability tests

There are two types of usability tests performed most frequently during medical device development: (1) formative usability tests, and (2) human factors (HF) validation tests. Notably, the latter test type – HF validation – was known as "summative" usability testing for many years and is still referred to as "summative evaluation" by IEC 62366. However, as indicated in **its final HFE guidance** released in 2016, the FDA considers the term "HF validation" more appropriate.

A formative test's primary goals are to identify a design's strengths and shortcomings (more charitably called opportunities for improvement), and to identify ways to optimize the design. Formative tests are usually conducted multiple times throughout development while the design is evolving. At least three rounds are common. These tests almost always yield valuable findings and insights at various stages of development, including early in development when you might have multiple alternative concepts to evaluate, and later when you have a near-final design that has been refined based on previous evaluations.

HF validation testing, which comes at the very end of the product development process, is quite different from formative testing. During HF validation testing, the focus is on observing users' interactions with the product and determining whether any of their interactions could lead to harm. At this stage, all aspects of a product's user interface, including software, hardware, labeling, and packaging, need to be "production-equivalent," with rare exceptions. The product is no longer evolving, but instead is final or near-final. The manufacturer is decidedly past the point of wanting to – or perhaps even being able to – make design changes. Rather, the manufacturer considers the product ready to be marketed. At this point, the goal is to generate evidence for regulatory body review that the product can indeed be used safely and effectively as intended.

What is considered a device's "user interface?"

When hearing the word "user interface," many people immediately think of control panels and software displays that present information to users. However, the term "user interface" actually refers to every product touchpoint (i.e., interaction point); all of the parts of a medical device that users see, hear, touch, smell, or taste comprise the user interface. Accordingly, the user interface may include hardware components, software screens, labeling, soundscapes, aromas (odors), and tastes. Many medical devices include several types of user interface elements. For example, an anesthesia workstation features rotary knobs (hardware), a display (software), a printed operator's manual (labeling), and anesthetic agent vapor (aroma).

FAQs: Formative versus HF validation tests		
Question	**Formative test**	**HF validation (aka summative) test**
Do regulators require you to conduct this type of testing?	No. However, US FDA's HFE guidance for medical devices and the prevailing international HFE standard (IEC 62366-1:2015) promote formative usability testing as a beneficial precursor to HF validation.	Often, but not always. If a device has critical tasks (tasks that, if not performed correctly, could lead to harm), then an HF validation test is typically required.
What is the test goal?	Identify product strengths and opportunities for improvement related to usability and use-safety *en route* to an improved design.	Confirm (i.e., validate) that representative users can interact with the given device safely and effectively and without committing dangerous use errors.
When should you conduct this usability test?	Early and often throughout the product development cycle.	When you have developed a final, production-equivalent device and before applying for regulatory clearance.
Can you have participants interact with an early, only semi-functional prototype during testing?	Yes. You can conduct formative testing with almost any design instantiation, including a paper prototype, computer-based prototype, or partially functional device.	No. An HF validation test should involve a production-equivalent device, labeling, and training. However, certain use scenarios might require you to temporarily adapt the device to validate users' ability to manage unusual scenarios, such as alarm conditions.
How many individuals should participate in the test?	It is typical to include five to eight individuals from each distinct user group, though sample sizes can vary.	You must include at least 15 individuals from each distinct user group if seeking FDA approval. Other regulators have not stated a sample size requirement.
Where should you conduct a usability test?	In a usability laboratory, conference room, market research facility, or one of many other convenient environments – even a participant's home or workplace.	It depends on the level of simulation required to mimic the use environment of the device. A usability laboratory or conference room might suffice, but sometimes testing warrants the use of an advanced medical simulator or even an actual use environment (e.g., ambulance). Test must occur in the US if submitting to FDA.
Should you ask participants to think aloud during the test?	Yes. The running commentary provides valuable insights that help you identify the device's interactive strengths and shortcomings.	No. Asking participants to think aloud can interrupt the natural task workflow and distort participants' interactions with the device.

Steps for conducting a usability test

The process for conducting a usability test involves four key steps:

1. Plan the test and develop a test protocol
2. Recruit representative users to serve as test participants
3. Conduct the test sessions
4. Analyze the data and report test findings

Planning the test

If you want a usability test to run smoothly and generate meaningful results, it is helpful to invest sufficient time in planning the test. A thorough test planning approach includes the steps below:

- Familiarize personnel with the device
- Determine the test goals
- Identify test session activities
- Define the test environment
- Define the training approach
- Determine what data to collect
- Develop the test protocol
- Develop the test documents

Familiarize personnel with the device. To ensure that you can define the right test scope, ask the right questions, and detect interaction issues, it is critical that test personnel are familiar with the device. Only once you understand the device's functionality and how to use it can you really plan an effective test and identify activities that will enable you to collect data to meet the test goals.

Determine the test goals. To plan a test effectively, you must define the test goals and the reasons for conducting the test. What are you trying to understand? Are you focusing only on the device's hardware or software components? Or are you also focused on the accompanying instructions and labeling, seeking to ensure they are clear and effective? Whatever your ultimate goal is, you will select your test session activities to serve those goals.

Identify test session activities. You should select test session activities that enable you to meet your test goals. For example, if your goal is to evaluate whether users can utilize an inhaler correctly, one of your test session activities should call upon participants to use the inhaler to deliver a simulated dose of medication. This activity – or use scenario – will enable you to observe whether participants make any mistakes while using the product and collect their feedback on doing so. During formative usability tests, you can include any use scenarios of interest (e.g., frequent tasks, challenging tasks, critical tasks).

When it comes to planning HF validation tests, the use-related risk analysis will be the driving factor behind which use scenarios participants perform during the test, and ultimately the test will focus on evaluating critical tasks. The use-related risk analysis documents all of the use-related risks associated with the device. Each row contains information about a specific use error, including the related task, hazard, hazardous situation, harm, severity, likelihood, and mitigation. Critical tasks are those that, if not performed or performed incorrectly, can lead to serious harm or – for combination products – compromised medical care. For more information on use-related risk analysis, see sidebar on page 11.

Writing use scenario prompts

It is not sufficient to know what the test session activities should be. You also need to determine how to best introduce the activities to the test participants. You have to develop "use scenario prompts," which direct the participant to complete a particular use scenario, and you have to do this without guiding or biasing the ensuing user performance. For example, if the use scenario is a simulated injection, the prompt might read: "Imagine it is time for your next injection. Prepare for and administer a simulated injection."

Note that the prompt does not instruct the participant how to administer the injection. It does not prompt the user to check the expiration date, remove the cap, pinch the skin, etc. Providing such a detailed and step-specific prompt would bias the participant. Instead, select a prompt to enable the test moderator to assess whether the participant could determine independently or with support of the labeling how to go about the task. Having clear, concise, and unbiased use scenario prompts is key to a successful usability test.

Define the test environment. You also need to consider what factors of the actual use environment should be simulated during the test to ensure a representative and rigorous device assessment. Take a product designed for use in a hospital, such as a cardiac ablation system. You do not need to go into an actual hospital's electrophysiology lab (a.k.a. "cath lab") to conduct the test. But, some characteristics of the environment, such as low lighting and space constraints due to the presence of large imaging equipment, could impact a user's interactions with a cardiac ablation system. As such, you should consider simulating these environmental factors in the usability laboratory or conducting the test in a medical simulation center to ensure that the environment is reasonably representative.

Define the training approach. During test planning, you will also need to think about training. In some cases, a user would receive training from a healthcare professional or company representative before using a device independently. In other cases, users are responsible for figuring out how to use a device based on their intuition and/or the accompanying instructions. In some usability tests, the training provided should match how actual users would be trained in the "real world" once the device is marketed. Including representative training is most common during HF validation tests where everything about the device use should be as representative as possible. It can be helpful to withhold training during early-stage formative tests (even if someone would use a device only after receiving training) to facilitate the most rigorous evaluation of the device. Including only untrained users in a formative test will reveal whether individuals can use the device correctly and safely based on the instructions, their intuition, and experience with similar devices.

Determine what data to collect. You should think about the types of data you want to collect during the test, based on the test objectives, and describe this information in the test protocol. For example, you might collect task performance data (i.e., use errors, close calls, and difficulties) and participants' responses to open-ended questions. Knowing about the types of data you want to collect enables you to plan the session and included activities to ensure you meet your research goals.

Develop the test protocol. It is important to document your approach and test methods in a written protocol. Typical protocols might be 10 to 30 pages in length, depending on the scope and formality of the test, and might even expand past 50 pages to delineate how an HF validation test will be performed, especially when there is a high number of critical tasks to evaluate. A protocol should briefly describe the device and its intended use, users, and use environments for context. It should also detail various aspects of the test method, including the test participants and associated recruiting methods, testing materials, environment, staff, activities, data collection and analysis methods, and reporting approach. Ultimately, the protocol serves to guide the test and should be detailed enough to enable the test personnel to run the test and collect the data necessary to meet the test objectives. As such, a final page length need not be predefined or constrained.

Develop the test documents. Before starting a test, you will also need to create some documents to help you record data and guide the test session. Typical test documents include an electronic datasheet (such as an Excel sheet) used by the analyst, a paper-based checklist used by the moderator, and a moderator's guide that includes a scripted introduction and any interview questions.

Submitting a protocol for Institutional Review Board review

Once the protocol is finalized, the conservative next step is to submit the protocol, along with recruiting screeners and informed consent forms, to an Investigational Review Board (IRB) or its equivalent for review and approval.

Generally speaking, research involving human subjects requires pre-approval by an IRB before it may proceed within the US. Pre-approval is particularly important when the research results will be submitted to a government agency such as FDA, which is part of the Department of Health and Human Services (HHS, the agency that established the IRB process).

IRB representatives review usability test protocols to ensure the test personnel are taking appropriate measures to protect human subjects (i.e., the test participants). This includes ensuring that participants are not asked to do anything that makes them uncomfortable and that the test personnel are not placing participants at risk of physical or psychological harm.

Recruiting test participants

Test participants should represent the people who will use the product in real life. If you are testing a peritoneal dialysis machine for use at home, you should certainly include people with kidney failure who are already receiving dialysis. You might also include people in partial kidney failure who might soon need dialysis, caregivers who help the dialysis patient manage their dialysis treatments, and/or healthcare professionals, such as nephrologists who have the ultimate responsibility for dialysis patient care.

Simply put, once you know who you would like to recruit for a test, you need to find them. This can pose quite a challenge in cases when you do not have many appropriate contacts. Usually, the most efficient option is to engage a market research firm to recruit participants on your behalf. Such firms maintain large databases of people with varying characteristics, and they identify people to participate based on criteria you provide. That said, if you are looking for hard-to-find individuals – such as people with a rare medical condition or a super-specialized clinician – you might choose to engage a patient advocacy group or professional association, respectively, to help access the intended users.

Sample size is an important consideration when planning a usability test. Usability test sample sizes are relatively small, particularly compared to the large samples used in clinical trials. This is because developing statistically significant findings is not an explicit goal of a usability test. Rather, the goal is to reveal design flaws that induce interaction problems, and this can be accomplished with a relatively small sample. In fact, research shows that a study involving five to eight participants per user group can identify 80 to 90 percent of usability issues with a given product. Notably, neither FDA nor IEC 62366 provide specific guidance for formative sample size, so you can use your judgment regarding sample sizes.

When it comes to HF validation testing, sample sizes are larger and somewhat regulated. The FDA's guidance calls for 15 participants per user group, which often leads to tests including a sample size of 30 to 60 participants – or even more – depending on the number of distinct user groups. IEC 62366 does not require a specific sample size. Many manufacturers include 15 participants per user group to ensure that they meet both FDA and IEC expectations.

Manufacturers who do not expect to market their product in the United States (and, therefore, are not subject to FDA expectations) sometimes choose to include fewer than 15 participants per user group to reduce cost. In such cases, manufacturers should avoid samples (including all distinct user groups) smaller than perhaps 20 to 30 participants in total.

Table B-1. Percentage of Total Known Usability Problems Found in 100 Analysis Samples (Faulkner, 2003)

No. users	Min. % Found	Mean % Found	SD	SE
5	55	85.55	9.2957	.9295
10	82	94.69	3.2187	.3218
15	90	97.05	2.1207	.2121
20	95	98.4	1.6080	.1608
30	97	99.0	1.1343	.1051

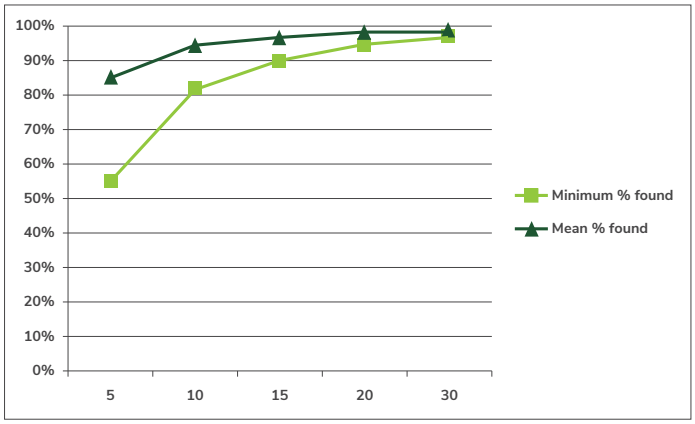

Figure B-1. Percentage of Total Known Usability Problems Found in 100 Analysis Samples (Faulkner, 2003)

Source: US Food and Drug Administration

Defining distinct user groups

When defining user groups, we suggest considering three key questions.

1. Who are the device's users?
2. What tasks do each of these users perform? Do all users perform the same tasks or, for example, do nurses perform device setup and physicians administer the treatment?
3. Are there distinct experiential, educational, or demographic differences among your users?

In some cases, these differences in tasks performed and background can be an indication that the broad set of users includes more than one distinct user group. Deciding what constitutes a distinct user group can be quite nuanced and is subject to input from regulatory bodies. The US FDA in particular has helpful suggestions regarding "distinct user groups" in their human factors guidance documents. As such, it can be helpful to seek input from experts and/or directly from the FDA when defining user groups, especially before conducting an HF validation test.

Conducting test sessions

Once you have planned the test and recruited test participants, the time has come to conduct the test sessions. We recommend conducting the test sessions in person, perhaps in a usability test lab that enables additional observers beyond the test personnel to observe unobtrusively from behind a one-way mirror. That said, there are various methods for conducting usability tests remotely, such as via a video conference and/or by sending test items to a participant's home, if in-person testing is not feasible.

In most cases, it is beneficial to engage a two-person testing team consisting of a moderator and an analyst to conduct each session. The moderator typically leads the sessions, asks interview questions, presents the test activities, and records key findings. The analyst observes the session, records detailed notes, and asks some follow-up questions.

Regardless of the test session's goals or focus, a typical test "agenda" includes the following:

- Greeting the participant and asking him/her to review an informed consent form and sign the form once all questions have been answered
- Providing an introduction describing the test session's goals and the basic activities the participant will perform
- Reviewing the participant's demographics and relevant background and experience (e.g., diagnosis, experience with a given type of product)
- Administering use scenarios to evaluate the participant's hands-on interactions with the device
- Conducting a final interview to collect subjective feedback about the participant's impressions of the device, and/or gather root causes of interaction issues
- Thanking, compensating, and dismissing the participant

Analyze data and report test findings

After completing all the test sessions, it is time to analyze the data collected and summarize it in a meaningful way. The type of analysis you perform and the type of report you will develop depends heavily on the type of test you conducted and the test's goals.

At a high level, a test report typically includes the following sections:

- An introduction to the test's purpose and objectives
- A brief summary of the test method, including the sample size, test session activities, and data collected
- A description of the results and findings (usually the longest section)
- A summary of the participants' background information

Regarding formative tests, we suggest focusing reporting efforts on describing high-level trends, observations, and behaviors. What seemed to go well? Where did participants struggle? Notably, a formative test report should not only describe the use-related issues. Rather, it should also include potential solutions to the issues. Human factors engineers and designers should reflect on the test findings, including participant feedback and test personnel observations, and provide actionable recommendations to help address use-related issues and improve the product's design.

HF validation test reports are quite different than formative test reports. As a reminder, an HF validation test involves a production-equivalent product, so you will be past the point of evaluating usability and identifying general trends. Instead, the goal is to generate evidence that the device can be used safely and effectively. To serve this goal, an HF validation report should describe and include an analysis of each interaction issue (e.g., use error, close call, difficulty) that occurred during the HF validation test. The report should include a description of the issue and describe the issue's root causes.

Regardless of whether you are writing a formative or HF validation test report, it is important to present findings in a clear and concise manner. Great writing and clear communication help those who did not observe the test sessions first-hand, such as FDA reviewers, get a true sense for how people interacted with and reacted to the device during the usability test.

Usability testing and use-related risk

Analyzing use-related risk is a critical part of the human factors engineering process, and an essential input to planning an effective usability test of a medical device. FDA and other regulators are not especially concerned about use issues that lead to minor or negligible harm. However, they pay close attention to interaction problems that could lead to serious harm or even death. Therefore, it is important to think about the relative level of risk associated with a given use error by referencing the use-related risk analysis. For example, injecting medication before it has warmed to room temperature might result in relatively minor harm, such as moderate pain caused by the cold medication's increased viscosity. But the potential harms associated with administering the same drug in a much more concentrated formulation than intended are much greater, and perhaps even fatal. So, during an HF validation test, participants will perform tasks in which there is a potential for serious harm to occur in a real-use scenario, enabling you to evaluate whether the device is designed in a way that facilitates safe and effective use. Of course, you will put protections in place to fully safeguard test participants from actual harm. For example, an automated external defibrillator under evaluation would be modified so that it could not deliver a significant shock during a usability test.

Conclusion

Medical device and combination product manufacturers conduct usability tests to ensure that users have safe, effective, and satisfying interactions with the medical devices they develop. Formative usability tests serve the important purpose of helping manufacturers uncover use-related issues and opportunities to improve the design during development. An HF validation test is conducted at the end of the development effort as a means to demonstrate to regulators – and confirm internally – that the intended users can interact with the device safely and effectively.

For more information about Human Factors Research & Design, visit us at HumanFactors.EmergobyUL.com.

About the authors

Erin Davis is an Associate Research Director with Emergo by UL's Human Factors Research & Design (HFR&D) team. She has been with the team since 2012. She has experience delivering HFE services to the medical device, pharmaceutical, scientific instrument, and laboratory equipment industries. A board-certified human factors professional, Erin leads and oversees research activities such as early-stage user research and usability testing. Furthermore, she helps clients develop key HFE documents for their design history files, including use-related risk analyses, and advises clients on how to apply HFE during product development to meet regulators' expectations. Erin is co-author of *Medical Device Use Error – Root Cause Analysis* and has served as President of the Human Factors and Ergonomics Society's New England Chapter. She holds a B.S. in Biomedical Engineering from Marquette University and an M.S. in Human Factors from Tufts University.

Allison Strochlic has been part of the team since its inception in 2005. She received her B.S. in Human Factors from Tufts University and went on to receive her M.S. in Human Factors in Information Design from Bentley University. A board-certified human factors professional, Allison serves as one of Emergo by UL's Research Directors. In this role, she contributes to and manages research projects such as usability testing, contextual inquiry, and interviews. Allison also contributes to expert reviews (i.e., critiques), helps clients develop key HFE documents for their design history files, and advises clients how to apply HFE during device development to meet FDA and other regulators' expectations. Allison is a co-author of *Usability Testing of Medical Devices* and has co-authored several articles on the topic of applying HFE during medical device and combination product development.

EMERGO

by UL