USABILITY TESTING OF MEDICAL DEVICES

SECOND EDITION

MICHAEL WIKLUND
JONATHAN KENDLER
ALLISON STROCHLIC

CRC Press
Taylor & Francis Group

USABILITY TESTING OF MEDICAL DEVICES

SECOND EDITION

USABILITY TESTING OF MEDICAL DEVICES

SECOND EDITION

MICHAEL WIKLUND
JONATHAN KENDLER
ALLISON STROCHLIC

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20151116

International Standard Book Number-13: 978-1-4665-9589-7 (eBook - PDF)

Contents

Acknowledgments

In the course of writing the first edition of this book, we interviewed the following human factors engineering (HFE), and design professionals to help identify the topics we should address in the book (i.e., define the book's "user requirements"):

- Marianne Boschelli
- Jason Bush
- Tony Easty
- Torsten Gruchmann
- Peter Hegi
- Edward Israelski
- Wayne Menzie
- Paul Mohr
- Jennifer Nichols
- Kyle Outlaw
- Gerald Panitz
- Susan Proulx
- J. B. Risk
- Gary Searle
- Mahesh Seetharaman
- Eric Smith
- Anita Stenquist
- Paul Upham
- Matthew Weinger

In the first edition of this book, Stephanie Seraphina, an independent usability engineering consultant, served the essential and demanding role of "alpha" reader, reviewing all of the content with an eye toward enhancing its usefulness and readability as well as giving our opinions a sanity check. Our colleagues Rachel Aronchick and Amy Dexter served in this important role for this second edition.

Our other human factors colleagues, past and present, helped us administer numerous usability tests that ultimately led to the insights that we share in this book.

Many clients granted us permission to use photos appearing in this book from the usability tests we conducted on their behalf (all photos without source lines were provided and are copyrighted by the authors).

Michael Slaughter of CRC Press embraced of our plan to update the book to match the Food and Drug Administration's and other regulators' evolving expectations regarding usability testing of medical devices. For the first and second edition, respectively, Jessica Vakili and Joselyn Banks-Kyle of Taylor & Francis provided us with excellent direction and editorial support.

Last and most important, our beloved families and friends gave us the encouragement and free time to write this book.

Thanks to all of you.

Michael, Jonathan, and Allison

About the Authors

The authors cofounded Wiklund Research & Design Incorporated (Concord, Massachusetts), a consulting firm that provided user research, user interface design, and evaluation services primarily to medical device manufacturers. In 2012, Underwriters Laboratories (UL) acquired Wiklund R&D to expand the renowned safety organization's portfolio of advisory services.

Each author is formally trained in human factors engineering (HFE) and frequently conducts usability tests of medical devices and software to identify opportunities for design improvement and validate their use safety for regulatory approval purposes. In 2008, the authors foresaw the need for a book providing detailed guidance on how to conduct usability tests of medical devices, noting the sharp global increase in the number of companies that have chosen to focus more attention on HFE and are compelled to practice it to meet regulators' expectations. In 2015, the authors revised the book to ensure good alignment with the FDA and other regulators' evolving expectations regarding usability testing of medical devices.

Michael E. Wiklund has worked in the HFE profession for over 30 years as a consultant and educator. He earned his master's degree in engineering design (specializing in HFE) from Tufts University, where he has subsequently taught user interface design for over 25 years. He has a professional engineering license and is a board-certified human factors professional.

He joined the profession in the mid-1980s, a time when microprocessor technology started to change the fundamental nature of medical technologies. Originally trained to make machines safe and user friendly, his early work was focused on "knobs and dials" but soon transitioned to making software user interfaces more comprehensible to users. Today, he helps optimize the design

of hardware, software, and hybrid devices as well as learning tools, such as quick reference guides, user manuals, and online resources.

In 1997, the U.S. Food and Drug Administration (FDA) invited Michael to write a guide to applying HFE in medical device development that was consistent with the (then) new guidance of the agency on the topic. Later, the FDA provided the guide to the Human Factors Engineering Committee of the Association for the Advancement of Medical Instrumentation (AAMI), which used it as a basis for writing AAMI HE74:2001, Human Factors Design Process for Medical Devices Development. AAMI HE74:2001 then became the basis for the current standard of the International Electrotechnical Commission (IEC) on the topic (IEC 62366-1:2015).

In 2005, Michael cofounded Wiklund R&D with the goal of providing comprehensive HFE services to industry—medical device manufacturers in particular. In the ensuing years, the firm has provided user research, user interface development, and usability testing services to over 50 clients in multiple countries. Now, as general manager of HFE at UL-Wiklund, Michael manages the work of HFE specialists in multiple countries while still serving as a technical contributor on key projects.

Michael's books include *Usability in Practice* (editor) (Academic Press, Inc., Cambridge, MA), *Medical Device and Equipment Design* (Interpharm Press, Inc., Buffalo Grove, IL), *Designing Usability into Medical Products* (coauthor) (CRC Press, Boca Raton, FL), and *Handbook of Human Factors in Medical Device Design* (coeditor) (CRC Press, Boca Raton, FL). In 2016, CRC expects to publish Michael's co-authored book titled *Medical Device Use Error—Root Cause Analysis*. He has published over 70 articles in the magazine *Medical Device & Diagnostic Industry* (MD&DI) that promote the application of HFE in medical device development and provide practical tips. He has been an invited speaker at multiple professional conferences and universities, where he has described HFE as an imperative in the medical industry and a path toward ensuring device safety and commercial success owing to its effectiveness, usability, and appeal.

Michael has served as a voting member of the AAMI Human Factors Engineering Committee for over 20 years. He has also served on the Human Factors Committee of the IEC and as chair of the Industrial Designers Society of America, Medical Section.

Jonathan Kendler has worked in the HFE profession since receiving his bachelor of fine arts degree in visual design from the School of the Museum of Fine Arts, Boston. He earned his master's degree in human factors in information design from Bentley College (now Bentley University). Accordingly, he brings a strong artistic sensibility to his HFE work.

Also a cofounder of Wiklund R&D, Jonathan has a strong interest in ensuring the usability of medical technology. As the design director of UL-Wiklund's HFE group, he is routinely involved in developing "clean sheet" user interfaces for medical devices as well as enhancing existing designs that need "refreshing." Clients characterize his user interface designs as intuitive and attractive, bringing attention to critical information and controls. His design portfolio includes medical devices ranging from small, handheld devices to room-size diagnostic scanners.

Virtually all of Jonathan's user interface design work is informed by user research and formative usability testing, which he often conducts personally to get close to the intended users and deeply understand opportunities for design improvement. He believes that this level of active involvement by a user interface designer in evaluating personal work is beneficial but requires absolute discipline to maintain objectivity.

Jonathan has also co-taught applied software user interface design at Tufts University and has delivered HFE workshops to medical and nonmedical clients.

Allison Y. Strochlic earned her bachelor of science degree in HFE from Tufts University. She later earned her master's degree in human factors in information design from Bentley College (now Bentley University). She joined Michael and Jonathan at Wiklund R&D shortly after earning her undergraduate degree, helped build the business during the years preceding the UL acquisition, and now serves as the research director of UL-Wiklund's HFE team.

Allison has accumulated thousands of usability testing hours, most involving

medical devices. She has a passion for making participants feel at ease during a usability test, enabling them to perform tasks as naturally as possible so that she and her colleagues can identify the strengths and shortcomings of a medical device, revealing opportunities for design improvement. Her usability testing projects have taken her all across the United States as well as to Europe and Asia. As such, she has become particularly adept at extracting useful findings from test sessions involving interpreters as well as test sessions conducted remotely (i.e., via telephone or the Web).

In addition to conducting usability tests, as research director, Allison ensures her colleagues conduct usability tests that meet clients' and, as applicable, regulators' needs. She has in-depth knowledge about the FDA's and international regulators' usability testing-related expectations and frequently advises colleagues and clients on effective usability testing methods.

Allison is a board-certified human factors professional and has served as a part-time lecturer in human factors at Tufts University. She is a member of the New England chapter of the Human Factors and Ergonomics Society and User Experience Professionals Association. She is also a member of the AAMI Home Use Environment committee. She has delivered multiple presentations to industry and academic audiences on effective usability testing methods and how to apply HFE throughout medical device development.

How to use this book

This book does not seek to replace other good works on the subject of usability testing, such as Dumas and Redish's *A Practical Guide to Usability Testing* (1999) (Intellect Books, Great Britain [Exeter, UK]), Rubin and Chisnell's *Handbook of Usability Testing* (2008) (Wiley Publishing, Inc., Indianapolis, IN), or the helpful usability Web site of the U.S. government (http://www.usability.gov). Rather, it seeks to help readers take what they might have already learned about usability testing from other resources and tailor it to the evaluation of medical devices software, and learning tools (e.g., quick reference guides, user manuals).

As human factors specialists who have conducted literally thousands of test sessions involving medical devices used by physicians, nurses, therapists, technicians, and patients, we believe we have some important lessons and tips to share. Therefore, we wrote (and have now updated) the kind of book we would have liked to use when we started testing medical devices, which explains why this book has so many illustrations and keeps things simple.

We doubt that many of you will choose to read the book cover to cover in a marathon session, such as one might consume a Danielle Steele or Stephen King novel. There is no protagonist, antagonist, or surprise ending. The book simply tries to answer the myriad questions that medical device manufacturers face when they test the usability of their devices, and we do so in an orderly, readable manner. There is no story to spoil if you want to jump among the topics.

That said, we present the content in a reasonably logical order. It starts with a cursory review of HFE and how usability testing fits in this area. It continues with a review of the government regulations and industry standards that have motivated many medical device manufacturers to conduct usability tests. Then, the book covers the nitty-gritty of planning, conducting, and reporting the results of a usability test.

As you read the book, keep in mind that usability tests are like snowflakes in the sense that each is unique. One hundred usability specialists working independently could take 100 different approaches to testing a dialysis machine, for example. Of course, their methods would have considerable overlap, but there would also be meaningful differences in approach that the practitioners would energetically defend as the best given the circumstances.

So, we suggest drawing as much insight as possible from this book and other resources and confidently approaching usability testing in your own unique way. After all, the point is not to conduct an academically perfect usability test per se. The point is to collect the best possible insights from a usability test so that you and your development team can make your medical device as safe, effective, and appealing as possible.

The limitations of our advice

This book offers our best advice on a wide range of usability testing topics, and *advice* is the key word. This is not a physics textbook replete with provable laws and equations. While force is demonstrably equal to the mass of an object times its acceleration ($F = ma$), the field of human factors lacks an equivalently exact means to calculate usability. Consequently, our advice is hardly the last word on any particular topic. Instead, consider it a starting point or a complement to other usability specialists' opinions and your own opinions and judgment.

The suggestions and recommendations that we offer in this second edition of the book stem from over 50 combined years of usability testing experience. However, we recognize that our professional colleagues might have different experiences and consider some of our advice controversial or even dead wrong. This is the nature of any text that shares knowledge on a substantially subjective topic that has been the focus of decades rather than centuries of study and practice.

Just as we warned in the first edition, please recognize that some of our advice has a limited shelf life. Regulations and accepted practices pertaining to usability testing of medical devices and software are likely to change over time, and some of our advice may become dated. So, please check our recommendations against the most up-to-date requirements. We originally developed this book's content in 2009 and updated it in 2015.

Disclaimers

We include here a few legal statements intended to protect you, us (the authors), and the publisher:

- Readers who choose to use the information and recommendations provided in this book do so at their own risk and discretion.
- The authors and publisher make no warranties, express or implied, regarding the information and recommendations contained in this book.

- Under no circumstances shall the reader hold the authors or publisher responsible for any damage resulting from the application of the information and recommendations contained in this book.

With these disclaimers behind us, we hope you enjoy our book and find its contents helpful, applicable, and thought provoking.

Who could use this book?

This book should be a good resource if you have an interest, need, or direct role in conducting a usability test of a medical device (or system), or if you are presently studying the topic. A list of professionals and role players who might find themselves in such a position includes the following:

- Biomedical engineers, biomedical technicians
- Cultural anthropologists
- Electrical engineers
- Ethnographers
- Human factors engineers, usability specialists, ergonomists
- Industrial designers, product designers
- Industrial engineers, manufacturing engineers
- Instructors and students
- Marketing researchers, marketing managers
- Mechanical engineers
- Medical device inventors
- Medical device regulators
- Program managers, program planners
- Quality assurance specialists
- Purchasing agents, procurement specialists
- Regulatory affairs specialists
- Risk managers
- Software user interface programmers
- Technical writers, technical communications specialists
- User interface designers, user interface experience planners, information architects

chapter one

Introduction

What is usability testing?

Usability testing calls for representative users to perform representative tasks as a means to reveal the interactive strengths and opportunities for improvement of a device. You can think of the activity as pressure testing or debugging the user interface of a device in terms of how it serves the users' needs, a critical need being safe operation. Tests may focus on early design concept models, more advanced prototypes, and even production units. A two-person team usually collaborates to run test sessions with one participant at a time. Good practice calls for preparing a detailed usability test plan and report that can be added to the design history file of a device.

Usability testing is a means to determine whether a given medical device will meet its intended users' needs and preferences. By extension, it is a way to judge if a medical device is more or less vulnerable to dangerous use errors that could lead to user or patient injury or death.

In its classic form, a usability test takes place in a special-purpose facility—a usability test laboratory—where test administrators can direct test activities from within one room while interested parties observe from an adjacent room via a one-way mirror (Figure 1.1). In practice, however, you can conduct a usability test in a wide range of environments, including nurses' lounges, conference rooms, equipment storage rooms, hotel suites, focus group facilities, medical simulators, and actual clinical settings such as an operating room.

The purpose of any usability test is to have test participants perform tasks with the given medical device, be it an early prototype, working model, production-equivalent device, or marketable device. If the medical device were a patient monitor, test participants might connect a simulated patient's sensor leads to the monitor, print an electrocardiogram

Figure 1.1 A conventional usability testing lab equipped with a one-way mirror.

(ECG) tracing, "shoot" a cardiac output measurement, and adjust the systolic and diastolic blood pressure alarm limits. If the medical device was an endoscope, test participants might place the endoscope into a simulated digestive tract, move the scope through the esophagus and into the stomach and over to the pyloric sphincter (valve), and then place the scope in a retrograde orientation to visualize the lower esophageal sphincter. If the medical device was an insulin pump, test participants might program a basal rate profile calling for different insulin delivery rates at each hour of the day, look up the carbohydrate content of a baked potato, deliver an eight-unit bolus before mealtime, and upload a month's worth of data to a computer for subsequent trend analysis. Importantly, the insulin pump would not be attached to the test participant (as it otherwise would be to an end user, who is using the device to administer insulin). Rather, tasks involving insulin delivery would be simulated, and if the participant needed to fill the device with insulin, inactive fluid (i.e., placebo), such as saline or plain water, would typically be used in its place. As suggested by the examples, usability testing of medical devices typically does not involve actual patients receiving treatment or taking active medications.

While test participants perform tasks, test personnel—typically a test administrator and note taker (e.g., data logger, data analyst)—observe intensively to determine how the medical device facilitates or hinders task completion. In addition to documenting observed use errors, test personnel might record data such as task times, test participants' comments, and various subjective design attribute ratings, such as ease and speed of use (Figure 1.2) (see "What data should you collect?" in Chapter 14).

If you are testing a fairly simple device, test sessions might breeze by in as little as 30 minutes. However, most test sessions last between 1 and 2 hours, providing enough time to properly orient the test participant to the test environment, purposes, and ground rules; to perform hands-on tasks; and to interview the test participant about the strengths and opportunities for improvement of the design, for example. A half-day test session is not an unreasonable duration if the device under evaluation requires one individual to perform an extensive number of tasks (e.g., unpacking, assembling, calibrating, operating [in multiple modes], and servicing). (See "What is the proper duration of a test session?" in Chapter 5 for more information about determining the appropriate test session length.)

Usability specialists (or allied professionals responsible for conducting the test) write detailed test plans to guide effective, consistent, and objective design assessments. After completing a test, analyzing the data, and developing findings, the test administrator reports his or her findings with the required level of detail and formality. A sometimes-lengthy narrative test report that describes the purpose, approach, and participants of

Figure 1.2 Scenes from usability tests of various medical devices.

the test and presents an analysis of the data, findings, and recommendations is a common final product that medical device developers can add to their design history file and submit to regulators.

Medical device developers are well served to conduct formative usability tests "early and often" during device development to assess design alternatives and identify opportunities for design improvement. Later in the design process, developers are essentially required to conduct a summative usability test to demonstrate that their medical devices are safe to use from an interaction design standpoint. During either type of test, users' interactions with the given medical device might proceed smoothly, suggesting that the design is on the right track or even ready for market introduction. Conversely, testing might reveal usability problems that could, should, or must be corrected prior to the release of the device.

Usability tests usually involve a small number of test participants as compared to market research studies and clinical trials, for example. An informal test involving just a few test participants can be productive. However, sample sizes in the range of 8–25 test participants are the norm (see "What is an appropriate sample size?" in Chapter 8), the mode being around 12–15. That said, final (also called validation or summative)

usability tests can include a large number of participants to ensure reliable results and meet regulators expectations. To address the latter need, sample sizes that are multiples of 15 are common. For example, a final usability test of a device used by six different, distinct group of users would call for a sample of 90 participants.

No matter the population sample size, the key is to get the right test participants. This means recruiting a sample of test participants who represent a good cross section of the people who will actually use the given medical device. Usability specialists sometimes expand the sample so that it includes an above-average proportion of people with limitations (i.e., impairments) that could affect users' ability to use the device. Expanding the sample in this way helps usability specialists detect potentially hazardous use errors that unimpaired users might not necessarily commit. Moreover, taking such an approach helps to determine the accessibility and usability of a medical device by people with impairments.

All sorts of usability problems can arise during a usability test (see "What kinds of usability problems arise during a usability test?" in Chapter 12). For example, it is not unusual to see test participants go down the wrong path within a software screen hierarchy because menu options are poorly worded or because information and controls of interest are oddly placed (Figure 1.3). Sometimes, test participants get stuck on a

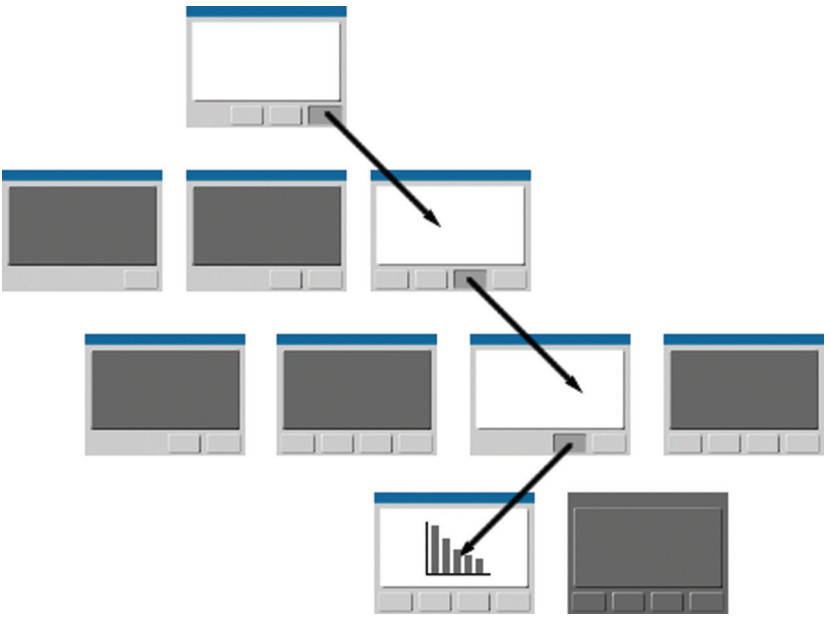

Figure 1.3 A sample user interface structure with a task sequence shown.

task because on-screen or printed instructions are incomplete, incorrect, or unclear. Also, test participants might press the wrong button because they misinterpreted the icon serving as the button's only label, or because the button was small and too close to other buttons.

Plenty of good things can happen during a usability test as well. For example, test participants might correctly set up a device for use on their first try without training—a harbinger of good usability across the spectrum of possible hands-on tasks. They might execute a therapeutic procedure in the exact order prescribed by the on-screen prompts. And, referring to a quick reference guide, test participants might properly interpret an on-screen and audible alarm and quickly perform the troubleshooting steps required to resolve the underlying problem.

Accordingly, usability testing is about discovering the good and bad (i.e., flawed) aspects of a user interface for the purposes of design refinement and validation. Programmers might think of usability testing as a method of debugging a user interface from a user interaction standpoint. Mechanical engineers might liken usability testing to pressure testing or metaphorically dropping a user interface onto a concrete floor from a considerable height. And, begging your pardon for one more comparison, we liken usability testing a user interface to a doctor giving a patient a physical—an inspection that usually shows most things are normal (i.e., in order) but highlights a few areas for improvement.

What is a medical device?

A medical device is a product used to diagnose, treat, or monitor a medical condition. Given this broad definition, regulators group medical devices into different classes based on the complexity and inherent potential of a given device to cause patient harm. Depending on the class of a given medical device, more or less human factors engineering (HFE) will be warranted.

We all have a general understanding of the term *medical device*. A medical device is something that physicians, doctors, nurses, technicians, and even laypersons use to diagnose, treat, or monitor a medical condition. Moreover, we think of a device as a physical item that might also incorporate a software user interface. Medical devices vary widely in terms of their size and purpose (Figure 1.4).

A syringe and a magnetic resonance imaging (MRI) scanner are both medical devices. So are exam gloves and cardiopulmonary bypass machines. However, as will be discussed, medical devices fall into different classes. You can conduct a usability test of virtually any medical device, but manufacturers of Class II and Class III devices are likely to invest more efforts into usability testing because their devices have a greater potential to harm someone if operated improperly.

Figure 1.4 Medical devices vary widely in terms of shape, size, function, complexity, and usage. (Photos [clockwise from top-left] courtesy of Industrial Design Consultancy, 3M, David Ivison, BrokenSphere, HEYER Medical AG, and Waisman Laboratory for Brain Imaging and Behavior.)

The Food and Drug Administration (FDA) defines a medical device as follows:

> An instrument, apparatus, implement, machine, contrivance, implant, *in vitro* reagent, or other similar or related article, including a component part, or accessory which is:
>
> - Recognized in the official National Formulary, or the United States Pharmacopoeia, or any supplement to them.
> - Intended for use in the diagnosis of disease or other conditions, or in the cure, mitigation, treatment, or prevention of disease, in man or other animals.
> - Intended to affect the structure or any function of the body of man or other animals, and which does not achieve any of its primary intended purposes through chemical action within or on the body of man or other animals and which is not dependent upon being metabolized for the achievement of any of its primary intended purposes.[1]

In Council Directive 93/42/EEC, the European Union (EU) offers the following definition:

> "Medical device" means any instrument, apparatus, appliance, material or other article, whether used alone or in combination,

> including the software necessary for its proper application intended by the manufacturer to be used for human beings for the purpose of:
>
> - diagnosis, prevention, monitoring, treatment or alleviation of disease
> - diagnosis, monitoring, treatment, alleviation of or compensation for an injury or handicap
> - investigation, replacement or modification of the anatomy or of a physiological process
> - control of conception
>
> and which does not achieve its principal intended action in or on the human body by pharmacological, immunological or metabolic means, but which may be assisted in its function by such means.[2]

The FDA recognizes three medical device classes.[3]

Class I: General controls

Class I devices are subject to the least regulatory control. They present minimal potential for harm to the user and are often simpler in design than Class II or Class III devices. Class I devices are subject to "general controls," as are Class II and Class III devices.

"General controls include:

1. Establishment of registration of companies, which are required to register under 21 *Code of Federal Regulations* (CFR) Part 807.20, such as manufacturers, distributors, repackagers, and relabelers.
2. Medical device listing with FDA of devices to be marketed.
3. Manufacturing devices in accordance with the good manufacturing practices (GMP) in 21 CFR Part 820.
4. Labeling devices in accordance with labeling regulations in 21 CFR Part 801 or 809.
5. Submission of a *premarket notification [510(k)]* before marketing a device.

Examples of Class I devices include elastic bandages, examination gloves, and handheld surgical instruments. Most Class I devices are exempt from the premarket notification and/or the GMP regulation.

Class II: Special controls

Class II devices are those for which general controls alone are insufficient to ensure safety and effectiveness, and existing methods are available to provide such assurances. In addition to complying with general

controls, Class II devices are subject to special controls…. Special controls may include special labeling requirements, mandatory performance standards, and postmarket surveillance.

Examples of Class II devices include powered wheelchairs, infusion pumps, and surgical drapes.

Class III: Premarket approval

Class III is the most stringent regulatory category for devices. Class III devices are those for which insufficient information exists to ensure safety and effectiveness solely through general or special controls.

Class III devices are usually those that support or sustain human life, are of substantial importance in preventing impairment of human health, or that present a potential, unreasonable risk of illness or injury.

Premarket approval is the required process of scientific review to ensure the safety and effectiveness of Class III devices. Not all Class III devices require an approved premarket approval application to be marketed. Class III devices that are equivalent to devices legally marketed before May 28, 1976, may be marketed through the premarket notification [510(k)] process until the FDA has published a requirement for manufacturers of that generic type of device to submit premarket approval data.

Class III devices that require an approved premarket approval application to be marketed are those:

1. Regulated as new devices prior to May 28, 1976, also called transitional devices.
2. Devices found not substantially equivalent to devices marketed prior to May 28, 1976.
3. Class III preamendment devices that, by regulation in 21 CFR, require a premarket approval application.

Examples of Class III devices that require a premarket approval include replacement heart valves, silicone gel-filled breast implants, and implanted cerebella stimulators.

Class III devices that can be marketed with a premarket notification 510(k) are those:

> Postamendment (i.e., introduced to the U.S. market *after* May 28, 1976) Class III devices that are substantially equivalent to preamendment (i.e., introduced into the U.S. market *before* May 28, 1976) Class III devices and for which the regulation calling for the premarket approval application has not been published in 21 CFR.

Examples of Class III devices that currently require a premarket notification include implantable pacemaker pulse generators and endosseous implants.[3]

Why conduct usability tests of medical devices?

Usability testing helps reveal opportunities to make medical devices easier, safer, and more efficient and pleasant to use. These improved interactive qualities benefit nearly everyone associated with a given medical device, especially the manufacturer, end user (i.e., caregiver), and patient. The FDA and other regulators essentially require that medical device manufacturers conduct usability tests to generate evidence that their devices are safe and effective for the intended users, uses, and use environments.

The most profound reason to conduct usability tests of medical devices is to protect people from injury and death due to use errors. Too many people have been injured or killed because someone pressed a wrong button, misread a number, misplaced a component, skipped a step, or overlooked a warning message when using a medical device, for example. And, while usability testing will not catch every design shortcoming that could lead to a dangerous use error, it will catch many of them. Therefore, usability testing should be considered a moral imperative as well as a de facto regulatory requirement. Also, it is usually a commercially advantageous activity.

Usability testing has many beneficiaries:

- **Manufacturers:** Usability testing can lead to user interface design refinements that are likely to increase device sales, engender customer loyalty, reduce the demand for customer support (e.g., calls to a hotline), extend the life span of a device, and reduce the chance of product liability claims. In short, it is good for business.
- **Customers:** Usability testing benefits customers such as hospitals, clinics, private medical practices, and ambulance services in myriad ways. Easy-to-use devices make workers more productive, improve worker satisfaction, reduce training and support costs, and improve patient care.
- **Health care professionals (HCPs):** Usability testing also benefits HCPs such as physicians, nurses, and therapists, as well as technicians and maintainers. Design improvements made as a result of usability testing are likely to make a device easier to learn and use, reduce the need for support, and empower HCPs to do their best work. Usable devices can even speed up work and enable HCPs to go home on time.
- **Patients:** Usability testing benefits patients because they are less likely to be injured or killed by user interface shortcomings that induce

users to err. Sadly, thousands of people die each year due to medical errors involving devices. For example, infusion pump programming errors (e.g., entering the number 80 instead of 8.0) have led to so many deaths that the industry coined the expression "death by decimal."[4] The application of HFE and usability testing in device development helps reduce the use error rate and limit the consequences of use errors that do occur. Increasingly, patients are also more often medical device end users. As such, patients also benefit from usability testing of devices that they might ultimately use themselves.

- **Lay (i.e., nonprofessional) caregivers:** Many medical devices have moved or are transitioning from a clinical to a home environment, where a nonprofessional caregiver, such as the patient's guardian, relative, or friend, will use them. While home health care can allow for greater patient care and convenience, it can also burden the lay users and/or caregivers responsible for operating the medical devices. Usability testing can help ensure that a medical device's user interface is well-suited for such lay caregivers.

Another reason to conduct usability tests of medical devices—closely related to preventing patient injuries and deaths—is to meet the device regulators' expectations. We address this topic extensively in "What is the relationship between usability testing and risk management?" in Chapter 2. For now, we will provide some basic details about the FDA's usability-testing-related expectations.

The FDA recognizes usability testing as one of the methods manufacturers should use to generate design inputs by evaluating the performance of existing products and, moreover, to validate the design of a device. In its HFE guidance (issued in 2011), the FDA identifies usability testing (a.k.a. "simulated use testing") as a primary means of demonstrating that "the intended users of a medical device can safely and effectively perform critical tasks for the intended uses in the expected use environments."[5] This statement refers specifically to summative (i.e., validation) usability testing, but the guidance also identifies usability testing as a productive, formative evaluation technique that facilitates the design of a safe and effective device.

Even before issuing its HFE guidance in 2011,* the FDA dictated that "design validation shall ensure that devices conform to defined user needs and intended uses and shall include testing of production units under actual or simulated use conditions."[6] Another FDA publication, *Medical Device Use-Safety: Incorporating Human Factors Engineering into Risk*

* The HFE guidance FDA released in 2011 was labeled "draft." At the time this book's content was completed, FDA had not yet released an updated (i.e., final) version of its guidance, which has been keenly anticipated for several years. We advise readers to check whether FDA has updated its guidance and to take note of any disparities between our advice and FDA's guidance.

Management,[7] described usability testing as a tool to identify potential use-related hazards. Refer to "Can usability testing identify use-related hazards?" in Chapter 2 for more on this particular topic.

What is a use error?

"Use error" is a jargony term that usability specialists use to describe cases in which usability test participants make a mistake. Importantly, the term is not supposed to place blame on the user per se (otherwise, one might refer to the event as a "user error"). In fact, usability testing shows that most use errors are due to shortcomings in a device's user interface, rather than a user's misstep. Use errors include performing the wrong action or failing to act when necessary.

"Use error" is a term of art used by usability specialists, among others, that describes cases when a device user does the wrong thing, including acts of omission (not performing a necessary action) and commission (performing the wrong action). However, instead of using the term "user error," which suggests the user is to blame, the "r" drops off to more neutrally suggest a user–device interaction problem.

When speaking with usability test participants, we typically use the term "mistake" rather than "use error." When we interview test participants after they complete hands-on tasks, they understand what we mean when we say, "Do you think you made any mistakes?"

Technically speaking, a mistake is the result of erroneous thinking. According to Donald Norman, a mistake occurs when "a person makes a poor decision, misclassifies a situation, or fails to take all the relevant factors into consideration."[8] However, we use the term more broadly, covering all kinds of errors, including those classically described as mental lapses and slips.

Another source—IEC 62366-1:2105—defines a use error as a "user action or lack of user action while using a medical device that leads to a different result than that intended by the manufacturer or expected by the user."[9] The standard includes the following notes as clarifications:

- "Use error includes the inability of the user to complete a task.
- Use errors can result from a mismatch between the characteristics of the user, user interface, task, or use environment.
- Users might be aware or unaware that a use error has occurred.
- An unexpected physiological response of the patient is not by itself considered use error.
- A malfunction of a medical device that causes an unexpected result is not considered a use error."

By the IEC standard definition, a use error occurs when a participant deviates from a prescribed procedure. However, in practice, simply

departing from a formalized procedure does not constitute the kind of use error you would cite in a usability test report as long as the deviation did not affect the participant's ability to achieve the desired outcome.

Here is a sample of use errors (i.e., mistakes).

- Layperson takes a single inhalation through an inhaler's mouthpiece when administering a single dose, but a single dose requires two sequential inhalations.
- Surgeon implants a drug port upside down (facing inside the body instead of outward), making it impossible to inject drug into the port's septum using a needle-tipped syringe.
- Critical care nurse attempts to deliver blood intravenously using an intravenous pump with a disposable tubing set that is not compatible with blood infusions.
- Resident physician attaches a pressurized oxygen line, instead of an IV (intravenous) fluid line, to an IV access.
- Pharmacist selects the wrong epinephrine pen from a shelf stocked with pens containing different volumes of drug (e.g., 0.15 mg for children weighing ≤66 lbs and 0.30 mg for individuals weighing >66 lbs[10]).
- Hospital housekeeper inadvertently adjusts a hospital bed's position while using a wet cloth to clean the bed's safety rail.
- Biomedical technician enters the wrong calibration value into the input field on a dialysis machine's setup screen.

Figure 1.5[11] is from an FDA presentation on use errors that mirrors a similar table in IEC 62366:2007. The illustration presents a logic-driven

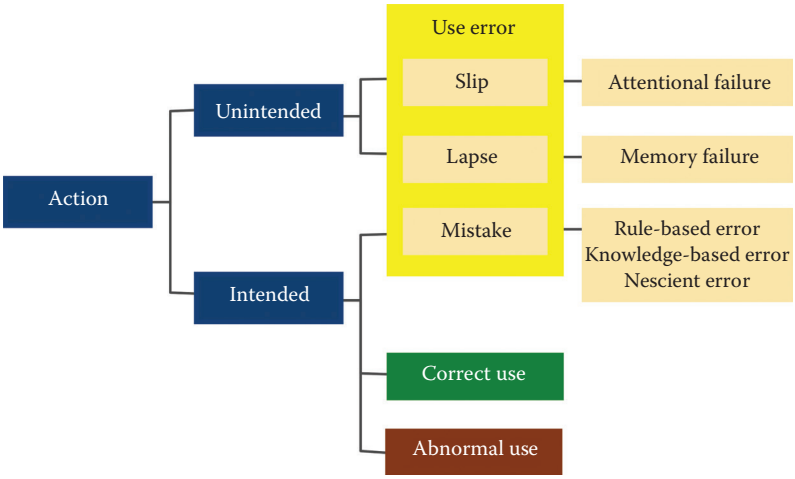

Figure 1.5 Diagram illustrating a logic-driven view of users' interactions with medical devices and how one decides if a use error occurred.

view of users' interactions with medical devices and how one decides if a use error occurred. The scheme accounts for the possibility that the use error might result from an intended or unintended action. The scheme also allows for a manufacturer to treat intended actions constituting abnormal use as outside the bounds of risk control. Abnormal use describes cases when an unqualified individual and/or someone who did not receive mandatory training operates a device and errs. For example, it would be abnormal use for a dermatologist to use an anesthesia machine to anesthetize a patient, or for a surgeon lacking the necessary training to perform robot-assisted surgery.

In 2015, IEC published an updated version of IEC 62366 (IEC 62366-1:2015)[12] that presents a new classification scheme for use errors (see Figure 1.6). The new scheme is similar to the previous one, differentiating normal use from abnormal use. However, it now places use errors into three classes:

- Use error caused by perception error
- Use error caused by cognition error
- Use error caused by action error

The new scheme aligns nicely with the increasingly popular approach to task analysis that focuses on user perceptions (P), cognitive tasks (C), and actions (A); what many usability specialists refer to as "PCA analysis." Given the task of using a glucose meter to test one's blood, and according to this scheme, misreading the expiration date on the test strip container would be a perception error, forgetting to disinfect one's fingertip prior to lancing would be a cognition error, and depositing a blood droplet off-the-mark on the test strip would be an action error.

A use error is not a close call (see "What is a close call?" in this chapter) or a difficulty (see "What is a difficulty?" in this chapter). In the edge case that a participant makes a mistake but immediately detects and corrects it without there being a significant opportunity of harm, we call it a close call. If a test participant struggles for an extended period of time to complete a task, there is certainly a difficulty but not necessarily a use error. If a user cannot complete a task, she or he has failed the task. However, the participant might not have committed a use error, even though task failures usually involve some type of use error. These distinctions are important when identifying and analyzing the use errors that occur during a usability test.

Use errors might be safety-related or not. They are safety-related if they are listed in a comprehensive failure modes and effects analysis (FMEA) or equivalent use-related risk analysis end product, and pose a significant risk of causing injury, death, or perhaps property damage. Use errors posing a low risk are treated as non-safety-related in the context of summative usability testing. One edge case is use errors that do not pose an immediate threat of harm but lead to a delay in therapy that could be harmful.

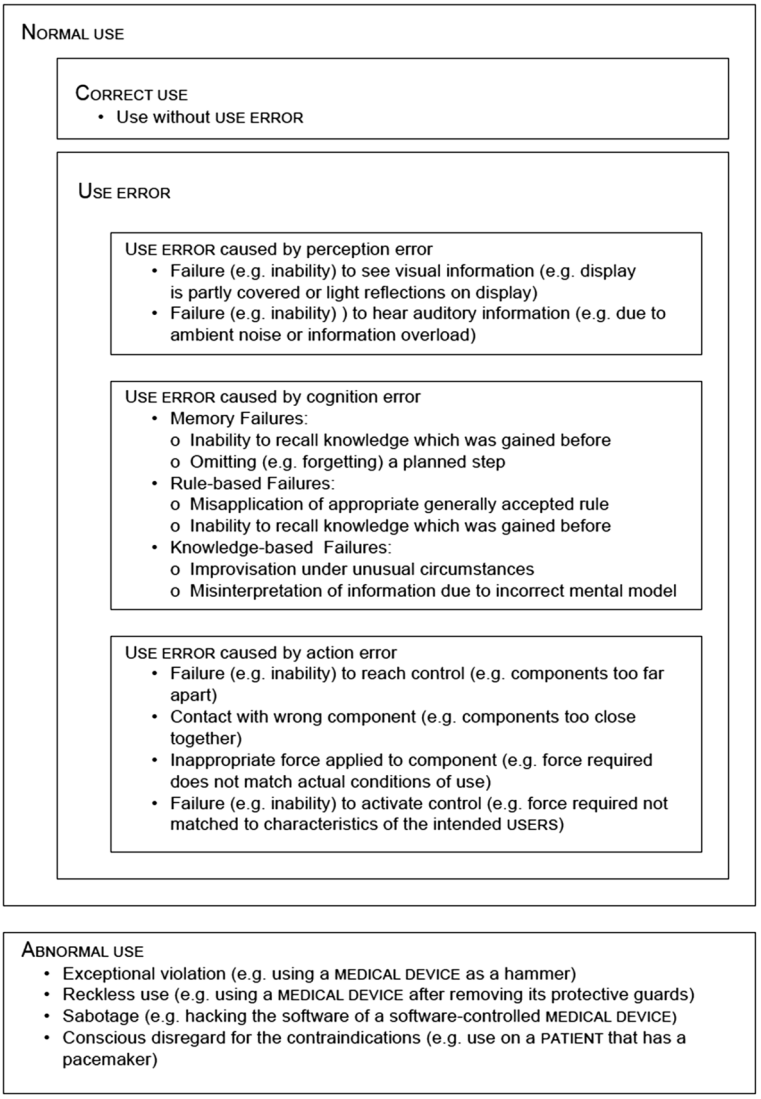

Figure 1.6 Definitions of normal use (correct and erroneous) and abnormal use per IEC 62366-1:2015.* (IEC 62366-1:2015, Table D.1. Copyright © 2015 IEC Geneva, Switzerland. www.iec.ch.)

* The author thanks the International Electrotechnical Commission (IEC) for permission to reproduce Information from its International Standard IEC 62366-1 ed. 1.0 (2015). All such extracts are copyright of IEC, Geneva, Switzerland. All rights reserved. Further information on the IEC is available from www.iec.ch. IEC has no responsibility for the placement and context in which the extracts and contents are reproduced by the author, nor is IEC in any way responsible for the other content or accuracy therein.

Use errors that are not safety-related might impede effective use of a given device—a significant concern—but not pose a significant risk of injury, death, or property damage. For additional information on what constitutes a dangerous use error, see "What is a dangerous use error?" in Chapter 2.

The occurrence of even one safety-related use error during a summative usability test is cause for intensive root cause analysis (see "What role does root cause analysis play?" in Chapter 15) and might necessitate design modifications. This is why formative usability testing is so important. Formative testing seeks to (1) reveal if a given device is vulnerable to use errors and, therefore, needs modifications, (2) identify potential mitigations to prevent the identified use errors, and (3) reduce the chance of use errors occurring before a device undergoes (and tries to pass) a summative usability test.

What is a close call?

During a usability test, participants sometimes come close to making a mistake, or they make a mistake but correct it before any harm could occur. These events are called close calls. Multiple close calls, which suggest a greater chance of an uncorrected use error, can indicate a user interface problem that should be fixed. Cases in which a medical device detects a use error, directs the user to correct the problem, and the user does correct the problem are not technically close calls. Rather, they are cases of a risk control measure working properly.

You probably know the expression "Yikes! That was a close call." One is likely to say it when just avoiding a traffic accident or almost spilling hot coffee on your lap. An airline pilot would experience a close call when over-rotating his or her aircraft and almost scraping its tail on the runway. The expression is used to describe circumstances when something bad almost happened, but did not (Figure 1.7).

In the usability testing business, "close call" is a term of art. It refers to cases in which a test participant comes close to making a mistake but does not. Stretching the definition a bit, the term can also describe a case when a test participant makes a mistake (i.e., commits a use error), but quickly recognizes and corrects it before any harm occurs. Reasonable individuals could argue that the latter is indeed a use error, quickly followed by corrective action. Whichever view you take of such close calls, you need to explain it in usability test reports and be consistent when distinguishing use errors and close calls.

Let us get back to the classic case of a close call, in which a user almost commits a use error. Here are some examples.

- **Nebulizer:** A layperson with chronic obstructive pulmonary disease (COPD) is cleaning a nebulizer's disassembled components one

Figure 1.7 USAF C-5 Galaxy comes close to striking its tail on the runway. (Courtesy of Lucas Ryan Photography.)

by one. She picks up a component containing electronic parts and is about to submerge it in a pan of soapy water. At the last moment, she realizes that the soapy water would damage the component.

- **Ventilator:** An anesthesiologist is preparing to start a new case. The last patient was an adult and the next patient is a young child. The anesthesiologist performs various machine setup tasks but overlooks switching the machine into pediatric mode. When he completes machine setup and is about to begin ventilation, he realizes the machine is still in adult mode and switches it to pediatric mode.
- **Drug port and catheter:** A neurosurgeon attaches a catheter to an external device's port, sliding the rubbery catheter tip over a thin tube on the port's side. When nearly finished with the task, she realizes that she did not apply a clip to secure the connection and proceeds to correct her oversight.
- **Glucose meter:** An individual with diabetes inserts a test strip upside down into his glucose meter. He immediately recognizes his mistake, removes the strip from the meter, and inserts the strip in the correct orientation. (This example is a case of immediate correction of a use error.)

In these examples, each user ultimately completed the task correctly. No harm occurred, but each user experienced a close call.

Sometimes, usability test specialists can detect a close call simply through observation. For example, it is usually easy to see when someone inserts a test strip in the wrong orientation into a glucose meter and then

corrects the problem by reorienting the test strip. However, it is sometimes necessary to ask participants if they experienced any close calls during a task. You can pose the question after each task or after the participant performs all tasks.

Why do we care about identifying close calls? For the same reason we would care about an automobile driver almost running through red lights. A close call indicates an increased potential for an actual mistake (i.e., use error) that could lead to harm. Also, a close call could delay task completion and/or cause user dissatisfaction.

Regulators care about close calls primarily because they indicate an increased potential for harmful use errors and an unwelcome delay in delivering potentially critical therapy. This is quite sensible if you consider the hypothetical close call involving the aforementioned ventilator. What if a usability test showed that all 11 anesthesiologist participants properly switched the ventilator from adult to pediatric mode before treating a two-year-old and four participants experienced close calls? This outcome suggests a heightened risk of a mode selection error occurring when the device is in actual use. Accordingly, one would want to identify the root cause of the pattern of close calls and introduce one or more new risk control measures to further reduce the likelihood of a user forgetting to set and confirm the correct mode.

Now, let's examine an event that might seem like a close call but really is not, at least from the perspective of one of the FDA's HFE specialists. We are talking about times when the user makes an initial mistake, but the medical device detects the problem and directs the user to correct it and the user does so. For example, when attaching a disposable blood tubing set to a dialysis machine and starting treatment, the user might forget to open a particular clamp that enables fluid flow. Upon blood pump activation, the machine might detect an abnormal fluid pressure that indicates a closed clamp and present a warning message with troubleshooting instructions to check and open the clamp. Yes, there was an initial use error. However, the user subsequently corrected it before there was a significant chance of. The FDA typically regards this kind of event as a case of a risk mitigation working properly. Therefore, the event does not warrant reporting as a use error or a close call. Usability test specialists need to use their judgment to decide if the event warrants reporting as a difficulty (see "What is a difficulty?" in this chapter).

In our experience, there are fewer—perhaps half as many—close calls than use errors during a usability test. This reduced frequency is likely due to the fact that users often are unaware that they have erred, and as a result, they do not take corrective action.

The FDA and other regulators ask manufacturers to look for patterns of close calls as well as difficulties. What constitutes a pattern? By definition, a single close call does not constitute a pattern. However, a couple

close calls of the same type constitute a pattern, particularly if there are a relatively small number of test participants. It indicates that a particular close call was not an anomaly, but rather indicates a possible user interface design shortcoming that might warrant further risk control. Having identified a pattern of close calls (two or more is our standard of care), the next step is to determine their root cause (see "What role does root cause analysis play?" in Chapter 15).

Documenting User Performance

Molly Story, a former member of the FDA's HFE team, stated during a Regulatory Affairs Professional Society (RAPS) presentation that usability test specialists should "observe and note all use errors, failures, and difficulties, including details about performance, e.g., task success or failure, use error, close call, reference to instructions for use (IFU), need for assistance, evidence of difficulty or confusion, unsolicited comments."[13]

What is a difficulty?

A difficulty is a hindrance to performing a task. A difficulty might arise when a usability test participant interacts with a device containing components that are small and difficult to manipulate, tries to remember the exact sequence of actions to properly calibrate a sensor, or searches for a desired option in a series of cascading software menus. A difficulty does not have to "stop the show"; a participant might complete a task without making a mistake, but sense that things could have gone easier.

The third member of the "trinity" that includes use errors and close calls is difficulties. By process of elimination, user–device interaction problems that do not fit the definition of the first two terms (see "What is a use error?" and "What is a close call?" in this chapter) are considered difficulties. Difficulties are cases in which a participant struggles in some way to perform a task. Difficulties might be viewed as lowest on the hierarchy of interaction problems, but difficulties can hobble a medical device and possibly lead users to reject it even if it is safe and clinically effective. Accordingly, the presence or absence of user interaction difficulties should be a major concern to marketers. A device's commercial success might ride on minimizing difficulties (i.e., making the device usable).

When testing reveals that many participants struggle to perform a given task (i.e., have difficulties), it might suggest an increased chance of use errors and close calls. It also suggests that users might take longer than ideal to complete a task, if they complete it at all, which raises concerns about a device's effectiveness.

Here are some examples of difficulties.

- A nurse takes more time than expected to open a kit (i.e., tray) containing a catheter and accessories because he initially cannot find its visually indistinct pull tab, and then he repeatedly loses his grip on the tab when trying to pull it with a gloved hand.
- An anesthesiologist tries to insert a block-like carbon dioxide filter into a plastic housing in various orientations until discovering the proper orientation on the fourth try—one that keys properly into the housing.
- A layperson tries to power-on a nebulizer by pressing and holding the on/off button, but the device does not activate. She tries again without success. Frustrated, she presses and releases the button quickly during her third attempt, and the device activates. In the first two attempts, she held the button down too long and the device ignored her input.
- A pharmacy technician reads a set of instructions repeatedly, trying to determine the correct amount of diluent to inject into a vial of lyophilized drug (i.e., powder) to produce the correct volume and concentration of fluid drug for an intravenous injection.
- An ophthalmologist struggles to properly align two components of a surgical device that penetrates the cornea, finding that the components become easily misaligned with the slightest errant hand motion.
- A patient transporter cannot determine how to lower a hospital bed's safety rail, having tried pulling and pushing on various bits and pieces until he finally pulls the release bar, which is difficult to see from his standing position.
- A physician tries to access an alarm log screen on a patient monitor by pressing the monitor's menu button and then touching various pop-up menu options until she finally locates the correct option.
- A layperson tries to follow printed instructions to attach a needle to a pen injector, but she is initially unable to do so. After several attempts, she realizes that she misinterpreted the arrows in the instructions' graphic and had previously twisted the needle counter clockwise rather than clockwise.

As suggested earlier, such difficulties might be more of a concern to device manufacturers than to regulators, presuming that any difficulties resulting in slow or unsuccessful task performance are not safety-related. This is because difficulties are a root cause of dissatisfaction that might drive away customers. Indeed, well before HFE became a regulatory imperative, some medical companies practiced it chiefly to improve their devices' ease of use in the quest for commercial advantage through design excellence.

Meanwhile, keep in mind that initial difficulties might lead to use errors and close calls. In such cases, there is no need to report difficulties separately and redundantly in a usability test report. The related use error and close call descriptions should provide sufficient detail to understand the interaction problem.

People are prone to experience more difficulty using a medical device for the first time, particularly if they have not been trained to use it. Often, difficulties subside when people use a device for the second time. However, difficulties might persist if training is needed to convey information that is essential to performing tasks, but the information is not self-evident or discoverable.

Usability test specialists should be able to detect when a participant is experiencing difficulty while performing a task. Signs of difficulty include the following:

- A series of unsuccessful actions (e.g., trying to connect two components by incorrectly pushing them together rather than screwing them together, as in the case of Luer connectors)
- Facial expressions indicating frustration or confusion (e.g., a grimace)
- Spontaneous comments, such as:
  - "Hmmmm…I'm not sure what to do at this point."
  - "This is a bit fiddly, isn't it!" (Expression common to UK residents.)
  - "This is like nailing Jell-O to the wall."
  - "Not sure where this is going."
  - "Am I missing something?"

Other ways to identify difficulties include (1) asking a participant to rate the ease or difficulty of performing specific tasks, and/or (2) asking a participant to comment on the ease or difficulty of performing specific tasks.

What are common regulator comments on summative (i.e., validation) test plans?

Regulators encourage medical device manufacturers to conduct usability tests, and, therefore, prepare test plans, that focus on the riskiest hands-on tasks. From a regulatory perspective, the ideal test plan will raise confidence that the ensuing usability test will reveal user interface design flaws that could lead to dangerous user errors, if any exist. Test plans that effectively link usability testing and risk management instill such confidence. Testing activities that are important but do not relate directly to device safety, such as evaluations focused chiefly on usability and appeal, should be marked as such.

Medical device manufacturers might choose to seek feedback on their usability test plans from regulators before proceeding with a summative

usability test. For example, the FDA might review a usability test plan on request and provide official comments via teleconference and letter. Undoubtedly, responding appropriately to the feedback increases the chance that the regulatory agency will accept the revised usability testing approach. Of course, accepting the testing approach has little to do with accepting the test findings as evidence that the design is valid (i.e., safe for use).

Following is a sample of the feedback that manufacturers have received over the past few years via discussions with and letters from regulators. Note that we have commingled comments on test plans and reports because they really address the same methodology issues in either a prospective versus retrospective manner, respectively.

Caveat: We have paraphrased and, in some cases, expanded the feedback for clarity's sake. As such, the feedback is indirect and should not be regarded as regulatory policy. Moreover, various regulators might have different views on the issues addressed. Therefore, you should regard the feedback presented as one of many possible inputs influencing your usability testing approach.

- **Finding new use errors:** Hypothesize the use errors that might occur during each task and consolidate them into a checklist the test personnel can use to evaluate participants' interactions during the usability test. Include the checklist as an appendix in the test plan, and be sure to indicate that test personnel will also document any unanticipated (i.e., unexpected) use errors that occur.
- **Prioritizing tasks:** Identify and prioritize hands-on tasks based on risk analysis results. We address this topic extensively in "How do you prioritize tasks?" in Chapter 11.
- **Relating tasks to risk analysis results:** Create a table delineating the identified risks and associated tasks to show that usability test participants will perform the riskiest tasks (i.e., tasks subject to use errors that are most likely to cause harm). Also demonstrate that participants will perform tasks that serve to assess the effectiveness of risk mitigations such as protective design features, labels, warnings, and IFU.
- **Including tasks related to "essential performance":** Testing should include tasks central to the device's primary functions, even if those tasks are not associated with highly rated risks.
- **Including secondary tasks:** Testing should include tasks such as cleaning, maintaining, and storing a device if these tasks are pertinent to the device's safe use.
- **Simulating use:** Describe how you will evaluate the critical aspects of user interactions without having participants actually deliver or receive treatment using the device.

- **Involving representative users:** Describe how you will recruit a sufficiently diverse sample of prospective users, including marginally trained or even untrained users who might choose or be directed to use the device, and users with certain impairments. As appropriate, the test participant sample should include individuals with varying levels of clinical experience, education, and experience with predecessor devices, for example.
- **Involving "low functioning" users:** Include "low functioning" individuals in the user population sample. Recruiting only "high functioning" individuals will not produce a representative sample of the intended user population.
- **Involving people with low language proficiency:** Include individuals who are less proficient in the device's selected language (e.g., English), noting that some devices, especially those sold over the counter without a prescription, might be used by individuals who have low proficiency in the selected language.
- **Company employees serving as test participants:** Do not use company employees as participants in the usability test.
- **Providing training:** Fully explain the need for and nature of any training that you plan to deliver to test participants. Make sure the training provided during testing matches the real-world training you expect to be available upon market launch.
- **Providing training materials/learning tools:** Indicate that participants will have access to any device training materials and learning tools (including user documentation) that would normally be available in an actual use scenario.
- **Allowing training benefits to decay:** There should be a delay between training and testing that might, in a realistic manner, result in some "decay" in the knowledge and skills attained during training. The length of the delay should be based on real-world use scenarios, and should be at least 30 or 60 minutes.
- **Including a sufficient sample size:** Include an appropriate size sample from each distinct user group—at least 15 people per group for a summative usability test. Regulators seem less concerned about the total test sample size, although a minimum of 15–25 participants appears to be a good working number, subject to increase if the intended user population has segments with widely differing capabilities and use the given device in distinctive ways (see "What is an appropriate sample size?" in Chapter 8 for more information about selecting an appropriate sample size). Be sure your plan includes a sample size rationale.
- **Identifying outliers:** Establish criteria for declaring a test participant as an "outlier" (see "How do you handle outliers?" in Chapter 15) whose data should be excluded from post-test analyses. If providing

participants with training before the usability test, establish criteria for disqualifying a test participant from participating in the subsequent usability test if he or she is unable to use the given medical device. For example, if a nurse-trainer determined that, based on a preestablished checklist of core competencies, a current home dialysis patient—a candidate usability test participant—could not safely use a dialysis machine at home, such an individual would not be an appropriate participant for a test of such a device. Importantly, usability test trainers should only administer a "competency assessment" if doing so is part of the training that will be provided in the "real world." Manufacturers should not develop or implement such an assessment for the sole purpose of identifying ideal (from the standpoint of passing the test) usability test participants.

- **Collecting subjective assessments:** Indicate that you will ask open-ended questions to collect "subjective assessments"—participants' feedback regarding the device's use-safety, labeling clarity, and any use errors' potential root causes. Regulators consider this subjective feedback to be an essential supplement to objective task performance data.
- **Collecting data unrelated to use-safety:** Delineate the type of data you plan to collect and how you will analyze it to draw conclusions regarding the use-safety of a given device. Be sure to differentiate between data you are collecting for the sake of validation (e.g., observed use errors, reported root causes, participants' subjective assessments of device use-safety) and data to serve commercial interests (e.g., subjective ease of use and satisfaction ratings). Explain that you will prioritize the reporting of primary, use-safety-related data and related analyses, and cover secondary data and related analyses in a manner that does not obscure validation-related data (e.g., in a report appendix). Purely usability-related data and related analyses can be excluded from summative usability test reports except in cases that usability shortcomings could lead to a critical delay in therapy, which by definition would make the shortcomings safety-related.
- **Predefining safety-related use errors and potential harms:** List the safety-related use errors that could occur during each hands-on task, including any actions or failures to act that might lead to harm. Concisely describe each use error and associated harms, referencing the source risk analysis document.
- **Tracking and analyzing close calls and difficulties:** In addition to describing how you will detect and document use errors, describe how you will detect and document close calls (cases in which participants almost committed a use error) and difficulties (cases in which participants struggle to perform a particular step or task). Describe