g) the regulatory requirements, etc.

Selection of training media is one of the major decisions that should be made during the design of training materials. MANUFACTURERS have more choices than ever before in terms of training media, and different forms of media have their strengths and weaknesses in delivering information, depending on the type of USER, the use context of the MEDICAL DEVICE, and the particular information being conveyed during instruction. For example, it is very difficult to teach the PROCEDURE for using an auto-injector to a novice lay USER based on printed instructions alone. To convey adequately the sound and physical actions involved in delivering an injection, video and audio have many advantages over print. Reference [17] provides substantial guidance on the selection of media.

Detailed design guidance is available for the design of specific types of training materials. References [4], [17] and [24] provide detailed USABILITY ENGINEERING design guidance on print and electronic USER manuals as well as quick-start guides and reminder cards. The guidance includes information on the organization of the material as well as the wording and formatting of instructions. Additional detailed design guidance can be found in reference [25]. Finally, USABILITY ENGINEERING guidance on the design of multi-media training and instruction is available in references [17], [26], [27] and [28].

15.5.3 Training

Although the MANUFACTURER might view training as something to be developed once a MEDICAL DEVICE development effort is near completion, it is useful to consider training early in the DEVELOPMENT PROCESS. For example, it is important to determine if all, some, or none of the intended USERS are intended to be trained prior to using the given MEDICAL DEVICE. This determination could have a strong influence on the degree to which the given MEDICAL DEVICE should be easily understandable to a first-time USER or require a predefined level of operational knowledge and skill. The determination of the need for training should be realistic, taking into consideration the MEDICAL DEVICE'S USE SPECIFICATION.

The goal of any MEDICAL DEVICE training is to provide the USERS with sufficient knowledge and skill to be able to use it in a safe and effective way. Trainers strive to move new USERS along the "learning curve," including getting beyond any obstacle that trainees might face if they try to use an unfamiliar MEDICAL DEVICE based only on intuition or related experiences. However, the MANUFACTURER neither assume that trainees will immediately master a given MEDICAL DEVICE, nor that they will recall all of the important details at the time at which they need to use the MEDICAL DEVICE. While the time period between training and actual MEDICAL DEVICE use might be just a few hours in some cases, in other cases the gap could be a few months, during which trainees might forget or confuse portions of what they learned about using the MEDICAL DEVICE.

When defining an appropriate level of training, the MANUFACTURER clarify the following questions.

a) Who among the intended USERS (defined by group or sub-group) are intended to receive training?

b) When in the course of a USER introduction to the MEDICAL DEVICE formal training should occur?

c) How much training will be provided and how many sessions are needed?

d) The training media (e.g. documents, slideshows, videos)?

e) Who will provide the training (e.g. MANUFACTURER-employed nurse educator, salesperson)?

f) What topics training will cover to ensure that all necessary HAZARD-RELATED USE SCENARIOS are addressed?

g) What competency checks (if any) should be performed to confirm that the trainee is prepared to operate the MEDICAL DEVICE safely and effectively?

h) How much time might pass between training and actual MEDICAL DEVICE use (i.e. the period of time during which the learning from training could decay)?

i) Is there a need for recurrent training sessions?

The MANUFACTURER should document the assumptions and intentions listed above in a training plan that can serve as a foundation for writing USER INTERFACE REQUIREMENTS – particularly related to initial MEDICAL DEVICE use and analysis of HAZARD-RELATED USE SCENARIOS. The training plan can subsequently serve as a basis for developing the actual training curriculum. Also, the plan can guide decisions regarding what kind of training (if any) to give USABILITY TEST participants.

MANUFACTURERS should ensure through testing that the training materials, in addition to being effective, do not create additional HAZARDS or HAZARDOUS SITUATIONS to the PATIENT or USER.

15.6 Develop detailed designs

The next step in the MEDICAL DEVICE design PROCESS is to develop detailed and integrated software and hardware USER INTERFACE designs, as well as the training materials and training. These items can be simple or complex. These designs should evolve from preliminary concepts to a refined design by means of iterative development efforts and evaluation efforts (i.e. FORMATIVE EVALUATIONS). Annex J, Clause 13 and Clause 16 provide additional detail for these efforts. [29]

15.7 Verify the design of the USER INTERFACE

Design verification requires the MANUFACTURER to confirm that the MEDICAL DEVICE design conforms to each element of the previously established specifications. Design verification is a requirement of the product realization PROCESS requirements of a quality management system. [9] Since the USER INTERFACE is part of the MEDICAL DEVICE, these PROCESS requirements suggest that the MANUFACTURER verifies the USER INTERFACE against the USER INTERFACE SPECIFICATION. This verification is not part of the USABILITY ENGINEERING PROCESS requirements of IEC 62366-1:2015.

Design verification of the USER INTERFACE involves confirming that USER INTERFACE REQUIREMENTS—for example, overall MEDICAL DEVICE dimensions, display parameters such as font size or luminance, control parameters such a button resistance and computer-interface parameters such as response time—are met. It is restricted to those verifiable parameters that do not require USABILITY TESTS to determine that they are met, which, in practice, means that they do not entail behavioural specifications, such as the time required to learn a PROCEDURE.

Design verification of the USER INTERFACE is typically integrated into design verification activities of the product realization PROCESS. This should be done, when possible, to support a consistent and proper level of effort.

Verification of the USER INTERFACE also can be a part of the verification of EFFECTIVENESS or implementation of RISK CONTROL measures as required by ISO 14971 (e.g. verification of a maximum limit of force, implemented to reduce a use-related RISK). Note that verification of EFFECTIVENESS of a RISK CONTROL measure can also involve USABILITY TESTS, which is outside the scope of the design verification.

16 Perform FORMATIVE EVALUATIONS

16.1 Conduct multiple FORMATIVE EVALUATIONS

A FORMATIVE EVALUATION seeks to evaluate USER INTERFACE designs during their development (i.e. during their “formation”) rather than when they are considered complete. A FORMATIVE EVALUATION can be a simple activity, noting that the goal is to learn about design solutions’ strengths and opportunities for improvement. FORMATIVE EVALUATIONS usually takes the form of USABILITY TESTS (see 16.2.4), cognitive walkthroughs (see 16.2.3), expert reviews (see E.7), and other evaluation techniques. FORMATIVE EVALUATION can support MEDICAL DEVICE concept

development, refinement and inform various types of design decisions. FORMATIVE EVALUATION is most beneficial when conducted iteratively throughout the development of the MEDICAL DEVICE.

FORMATIVE EVALUATIONS are completed prior to the SUMMATIVE EVALUATION and should be initiated early in the MEDICAL DEVICE research and development cycle. At an early stage of USER INTERFACE design, FORMATIVE EVALUATION serves to identify design strengths and opportunities for improvement. At the latter stage of USER INTERFACE design, FORMATIVE EVALUATION enables the MANUFACTURER to determine whether the MEDICAL DEVICE meets SAFETY, *USABILITY, USER and business needs* and ultimately supports successful SUMMATIVE EVALUATION of the MEDICAL DEVICE.

It is expected that FORMATIVE EVALUATIONS occur iteratively so that the MANUFACTURER can identify USER interaction problems and implement effective solutions prior to the SUMMATIVE EVALUATION.

FORMATIVE EVALUATIONS usually include TASKS or HAZARD-RELATED USE SCENARIOS in which USE ERRORS could occur and help to determine if the RISK CONTROLS designed into the MEDICAL DEVICE have been successful.

FORMATIVE EVALUATION data can include:

a) customer preference survey responses;

b) focus group participants' inputs (i.e. comments);

c) USABILITY TEST participants' comments, made while performing hands-on TASKS as well as upon reflection on their TASK performance afterward; and

d) USABILITY TEST participants' ratings and rankings pertaining to hands-on TASKS, specific MEDICAL DEVICE characteristics and the MEDICAL DEVICE in general.

The MANUFACTURER is required by IEC 62366-1:2015 to establish and maintain a USER INTERFACE EVALUATION plan to guide FORMATIVE EVALUATIONS. Results should be documented in a test report supported by raw and processed data sets (e.g. a spread sheet containing TASK performance data) and, if collected, video recordings and photographs of the test sessions. Design shortcomings identified during testing should be formally tracked to ensure they are resolved and re-evaluated as needed.

It is a best practice for MANUFACTURERS to conduct enough FORMATIVE EVALUATIONS prior to a SUMMATIVE EVALUATION to minimize the likelihood of discovering new problems. The goal is to conduct FORMATIVE EVALUATIONS at a time in the development PROCESS when they can have a greater level of influence on the USER INTERFACE design. Performing only the minimum possible amount of USABILITY ENGINEERING at the end of the development PROCESS, a time when designs are relatively inflexible and MANUFACTURERS are hesitant to change a USER INTERFACE, increases the likelihood that the SUMMATIVE EVALUATION will discover that the use-related RISKS have not been adequately controlled.

FORMATIVE EVALUATION also can focus on any aspect of USER interaction with a MEDICAL DEVICE that concerns a MANUFACTURER, including interactions influencing USER satisfaction and those that could affect a MEDICAL DEVICE'S commercial success.

16.2 Recommended methods for FORMATIVE EVALUATION

16.2.1 General

There are number of methods available to conduct FORMATIVE EVALUATIONS. The most commonly used methods are:

a) various types of reviews, such as
   – expert reviews (E.7),

  - standards reviews (E.17), and
  - heuristic analyses (16.2.2 and E.11);

b) cognitive walkthroughs (16.2.3 and E.4); and

c) USABILITY TESTS (16.2.4).

16.2.2 Conduct heuristic analysis

During design development, the MANUFACTURER intermittently assess (i.e. inspect or audit) the evolving USER INTERFACE design based on established design principles. The assessment (the aforementioned USER INTERFACE inspections) can be conducted in either a simple or elaborate manner, and be based on established USABILITY ENGINEERING principles and MEDICAL DEVICE-specific USER INTERFACE REQUIREMENTS. Such assessments are an effective way to detect design shortcomings at a stage when it is relatively easy and inexpensive to fix them as compared to fixing them when a design is presumably complete.

16.2.3 Conduct cognitive walkthrough

A cognitive walkthrough can be the first step taken to obtain USER feedback on a MEDICAL DEVICE'S USER INTERFACE. The technique calls for a MANUFACTURER to present its early design solution to a relatively small number of people, one at a time in sessions that might be brief or extended, noting that an hour-long session is not uncommon when only a lower-FIDELITY USER INTERFACE prototype (e.g. model) is available. The early design solution might take the form of a storyboard (e.g. a series of printed screens) or computer-based SIMULATION, perhaps complemented by a physical model. The technique depends on research participants, representing USERS, thinking through and verbalizing their thoughts, reactions and imagined actions based on static or marginally interactive representations of the early design solution. In place of touching a physical control, the participant would describe the control action and the test moderator would describe the MEDICAL DEVICE'S response, or perhaps swap one drawing for another one that depicts the MEDICAL DEVICE'S new state.

16.2.4 Conduct USABILITY TESTS

USABILITY TESTS involve observing USERS while they perform TASKS with the MEDICAL DEVICE.

USABILITY TESTS involve recruiting USERS of a specific USER GROUP and asking those USERS to complete a set of TASKS. The test moderator conducts the USABILITY TEST via a test script. The session can be recorded through audio and video to enable later review to confirm or supplement data collected during the test session.

USABILITY TESTS are usually conducted with representative USERS performing specific TASKS of interest or following TASK-based USE SCENARIOS that involve important MEDICAL DEVICE functions. USABILITY TESTS are normally conducted in simulated-use conditions that could affect the USERS' interactions with the MEDICAL DEVICE. For some USABILITY TESTS, USERS need to have specific domain, product or application-specific knowledge and experience. For example, when testing a diabetes management software app, it can be informative to use participants who have been using paper-based RECORDS to manage their diabetes for many years.

Choosing an appropriate sample size is a key consideration when planning FORMATIVE EVALUATIONS and SUMMATIVE EVALUATIONS (i.e. USABILITY TESTS). USABILITY TESTS for FORMATIVE EVALUATIONS can be beneficial using a small sample (e.g. 5-8) of test participants representing the entire USER population. Many USABILITY SPECIALISTS recommend small sample sizes when conducting FORMATIVE EVALUATIONS because it is usually sufficient to uncover major USER INTERFACE design issues. Sample size is more thoroughly discussed in Annex A of AAMI HE-75:2009 [4] and reference [30]. Standard practice and supporting research studies suggest that after five participants are tested, the law of diminishing returns applies, where participants will identify the same design shortcomings with increasing little additional USABILITY information gained from each additional participant. Annex K contains additional information regarding sample size.

A USABILITY TEST can be conducted on one or more prototypes with varying degrees of FIDELITY such as paper sketches, wireframes, hardware or software mock-ups, a functional prototype or a completed MEDICAL DEVICE. A MANUFACTURER can also conduct USABILITY TESTS on similar MEDICAL DEVICES on the market to understand their strengths and weaknesses. Additional information on USABILITY TESTS of MEDICAL DEVICES is provided in reference [31].

16.3 Analysis of FORMATIVE EVALUATION results

Table 6 presents example USE ERRORS that could arise from USER INTERFACE design shortcomings and suboptimal characteristics and be uncovered during FORMATIVE EVALUATION. Certain USE ERRORS that might appear to have been caused by the USER can ultimately be traced to a design shortcoming. Good designs take into consideration and limit the potential effects of human fallibilities, including such common and predictable ones as forgetting a procedural detail or overlooking a visual indication.

Table 6 – USE ERRORS caused by sample USER INTERFACE design shortcomings

USE ERROR	USER INTERFACE design shortcomings
USER presses the wrong button.	Push buttons on a control panel are too closely spaced.
USER misinterprets the icon and selects the wrong function.	Two icons on a software screen look too similar.
USER enters incorrect sequence and fails to initiate therapy.	A USER INTERFACE requires a complex, lengthy, and arbitrary sequence of button pushes to initiate a therapy.
USER repeatedly opens the door and presses the reset key instead of clearing air from the infusion line.	Infusion pump displays misleading “Open Door-Reset” message when air is in the infusion line.
USERS fail to detect a dangerous increase in heart rate because ALARM LIMIT is set too high and USERS do not look at MEDICAL DEVICE display because they are over-reliant on the ALARM SYSTEM.	USER-adjusted high and low ALARM LIMITS on a heart-rate monitor are not continuously displayed.
USER cracks catheter connector during catheter attachment.	Typical USER-applied force exceeds breaking strength of catheter connector.
USER forgot to replace a critical component when reassembling a MEDICAL DEVICE after cleaning it.	The MEDICAL DEVICE could be assembled and powered-up with a critical component missing.
USER ignored a warning label telling the USER to disconnect the PATIENT tube before turning the MEDICAL DEVICE off.	The MEDICAL DEVICE did not require the USER to confirm PATIENT disconnection before powering-off.
USER disregarded a warning symbol and allowed a portable MEDICAL DEVICE to run out of battery power.	The warning symbol was not sufficiently attention-getting.
USER forgot to confirm the new parameter settings.	The MEDICAL DEVICE reset the parameters to the previous settings after “timing-out” without notifying the USER that the new settings had been discarded and the previous ones were in effect or asking the USER to confirm the new settings.

The MANUFACTURER should continue to iterate the design and perform FORMATIVE EVALUATION until it is believed that all use-related RISKS have been adequately controlled, no further refinement is needed and the MEDICAL DEVICE is ready to proceed to SUMMATIVE EVALUATION.

17 Perform SUMMATIVE EVALUATION

17.1 General

The purpose of a SUMMATIVE EVALUATION is to evaluate the USABILITY of the USER INTERFACE as it relates to the successful completion of the TASKS associated with the HAZARD-RELATED USE SCENARIOS. A SUMMATIVE EVALUATION has no testable requirements in the sense used with a laboratory test. It is an evaluation of data that usually includes USABILITY TEST data. The requirement is that the data from the SUMMATIVE EVALUATION allows the MANUFACTURER to

conclude that no further improvement of the USER INTERFACE is necessary or practicable. These results are then transferred to the RISK MANAGEMENT PROCESS to determine whether the RESIDUAL RISK is acceptable.

A SUMMATIVE EVALUATION usually follows one or more FORMATIVE EVALUATIONS. A successful SUMMATIVE EVALUATION demonstrates that a MEDICAL DEVICE is not vulnerable to potentially harmful USE ERRORS. However, a SUMMATIVE EVALUATION might reveal that a MEDICAL DEVICE remains vulnerable to potentially harmful USE ERRORS, either because USABILITY TEST participants committed USE ERRORS on HAZARD-RELATED USE SCENARIOS or because testing revealed a pattern of CLOSE CALLS. Such USABILITY TEST results indicate the need for further USER INTERFACE improvement and re-testing unless RESIDUAL RISKS are deemed to be acceptable in relation to the benefit of using the MEDICAL DEVICE.

SUMMATIVE EVALUATION generally involves performing a USABILITY TEST under conditions of simulated use.

EXAMPLE 1 A USABILITY TEST of a MEDICAL DEVICE using a manikin as the PATIENT.

EXAMPLE 2 A USABILITY TEST on an injection MEDICAL DEVICE that does not contain a needle or any drug.

For some MEDICAL DEVICES, it can be difficult to conduct a USABILITY TEST because it is not practicable to simulate the use and it is unethical to conduct a USABILITY TEST in actual use. In these cases, it can be justifiable to use other evaluation methods.

EXAMPLE 3 Expert and highly experienced cardiac surgeons can perform an expert review of a very specialized cardiac surgical instrument where an empirical performance based SUMMATIVE EVALUATION by USABILITY TEST of heart surgery success cannot be practically simulated.

Additionally, expert reviews can be considered when the scope of the SUMMATIVE EVALUATION is limited to minor changes to the USER INTERFACE that do not involve HAZARD-RELATED USE SCENARIOS associated with serious HARM or in the case where the MEDICAL DEVICE has no HAZARD-RELATED USE SCENARIOS.

USABILITY TEST participants include appropriately screened representatives of the given MEDICAL DEVICE'S distinct USER GROUPS (e.g. PATIENTS, nurses, and technicians who might all use a home dialysis machine). A typical USABILITY TEST protocol calls for participants to perform hands-on TASKS associated with the selected HAZARD-RELATED USE SCENARIOS. Satisfying the goal of observing USERS interact realistically with a MEDICAL DEVICE, without actually delivering medical care, sometimes requires elaborate USE ENVIRONMENT SIMULATION. Otherwise, USABILITY TESTS can take place in a SIMULATION laboratory or even a conference room.

USABILITY TEST data collected should include:

a) TASK completion (and, where related to SAFETY, time to complete);

b) descriptions of observed USE ERRORS, CLOSE CALLS *and use difficulties*;

c) participants' comments (e.g. anecdotal remarks) about their MEDICAL DEVICE interactions; and

d) participants' reported root causes of their USE ERRORS and CLOSE CALLS.

USABILITY TEST data also can include:

e) *subjective ratings about the USER INTERFACE, if a MANUFACTURER wants to assess MEDICAL DEVICE attributes not related to SAFETY, such as USER satisfaction.*

17.2 Conduct a SUMMATIVE EVALUATION

A SUMMATIVE EVALUATION is a formal activity that follows a USER INTERFACE EVALUATION plan. The testing should follow the plan as precisely as possible. Any deviations from the plan should be cited in the associated test report.

Ultimately, there is one SUMMATIVE EVALUATION. If new problems are found or known problems persist in a SUMMATIVE EVALUATION, the evaluation is redefined to be one more in the series of FORMATIVE EVALUATIONS. In such a case, the MANUFACTURER should make the necessary design refinements and then conduct a SUMMATIVE EVALUATION.

USABILITY TESTS for SUMMATIVE EVALUATIONS are qualitative investigations that can be reported in the form of objective data from observations of USER interactions with the USER INTERFACE and their descriptions of their experiences afterward. These data can be supplemented with statistics, but only simple descriptive statistics (e.g. USE ERROR counts, *TASK times*) rather than inferential statistics (i.e. confidence limits, standard error measurements, statistical significance, Type I or II error rates, etc.).

To evaluate the USABILITY of a MEDICAL DEVICE as it relates to SAFETY, a USABILITY TEST in a SUMMATIVE EVALUATION should have an appropriate probability of observing a USE ERROR caused by a design defect. The number of participants used in the test (sample size) affects the probability of observation. For example, using the methodology of Annex K, assuming for a USER GROUP that a USE ERROR occurs with a probability of 15 % for a single test participant, this USE ERROR would be observed with a probability of 91 % when the sample size is 15 test participants.

To determine the appropriate sample size, the MANUFACTURER should consider the potential consequences of USE ERROR, the complexity of the design and degree of similarity to existing MEDICAL DEVICES as well as the expected heterogeneity of each USER GROUP. Confidence in the test findings of the adequacy of a USER INTERFACE increases when the sample size is increased.

The primary reasons for increasing sample sizes would be to:

a) reveal expectedly subtle USER INTERFACE design shortcomings;

b) involve people with a wider range of secondary selection characteristics; and

c) draw incrementally more reliable conclusions about a design's merits when one expects TASK performance *and USER preference* to vary widely.

Test results should be documented in a report, which can be augmented by raw and processed data sets (e.g. a spreadsheet containing data) and video recordings and photographs of the test sessions, presuming that the MANUFACTURER obtains permission from the test participants to use their images.

17.3 Data collection

17.3.1 General

USABILITY TEST data collection to support evidence that the "MEDICAL DEVICE, as designed, can be used safely and effectively" includes USABILITY TEST participant:

a) performance data (observational); and

b) comments (subjective).

Observational and subjective data are complementary inputs to assessing adequacy, strengths, weaknesses, SAFETY and EFFECTIVENESS of the USER INTERFACE. USE ERRORS are investigated and explained such that subjective assessment by USABILITY TEST participants is used to help identify the root cause of each observed USE ERROR. Additional information is found in 17.3.3.

17.3.2 Observational data

During a USABILITY TEST for SUMMATIVE EVALUATION, the test participants are asked to perform the USE SCENARIOS previously selected (see Clause 12).

While they perform each USE SCENARIO, the study moderator should observe the test participants and record their performance on each study TASK and sub-TASK or step as one of the following: CORRECT USE, USE ERROR, CLOSE CALL *or use difficulty*. It is important to collect the observational data at a sufficient level of interaction detail to enable identification of the source of any use problems that occur.

During the SUMMATIVE EVALUATION session, it is important that the moderator not influence the participants' behaviour. The purpose of SUMMATIVE EVALUATION is to approximate realistic use situations so as to learn how USERS are likely to interact with the MEDICAL DEVICE in actual use. The moderator should be neutral to the outcome of the test and seek to ascertain the truth. For example, the moderator should not ask the participant to "think aloud" because this interferes with realistic use; however, the moderator should record any comments the participant makes spontaneously, if any.

Sometimes it is difficult to avoid the occurrence of test participant behaviour that is caused by the artificial nature of simulated use. For example, test participants sometimes fail to check the expiration date of a medication because they do not expect it to have expired or because it does not matter since the medication is not actually delivered to a PATIENT. The MANUFACTURER should seek to minimize the occurrence of such events (sometimes called "test artefacts") but if the events are unavoidable, these aspects of the USER INTERFACE can be assessed through KNOWLEDGE TASK STUDIES (see 17.3.3.6).

Not all observational data is necessarily objective. Some assessments of USER behaviour are a subjective interpretation of an observation based on professional expertise and experience of the USABILITY SPECIALIST.

17.3.3 Subjective data

17.3.3.1 General

Subjective data should be collected through debriefing interviews with the test participants following a USABILITY TEST performed for SUMMATIVE EVALUATION. Simply counting the USE ERRORS does not support understanding of the root cause of a USE ERROR, which can be only understood with clarification derived from the perspective of the test participants involved with the USE ERROR.

Post USABILITY TEST interview data can be used to establish the root causes where USERS were observed to commit a USE ERROR, experience CLOSE CALLS or *have use difficulties* completing important USER TASKS. Post USABILITY TEST interview data is often the best or only available data for assessing USE ERRORS, CLOSE CALLS *or use difficulties* that occurred but were not observed during testing. The purpose of the interview is to identify unobserved use problems and also any errors in perception or cognition that the test participants might have made because such errors are not observable. Obtaining the USER'S perspective in an interview provides information to help determine whether the observed USE ERROR, CLOSE CALL *or use difficulty* might have been caused by an error of perception or cognition. It is also essential for determining whether previously unknown use-related HAZARDS exist in the design of the USER INTERFACE.

Post USABILITY TEST interview data collection should be "active" rather than "passive" such that test participants are asked questions directly by test moderators rather than being simply allowed to comment voluntarily, given rating scale instruments or invited to respond to electronic questionnaires or surveys. Care needs to be taken to ask the questions in an unbiased way so as to not lead the participants.

17.3.3.2 Impression of the overall use of the MEDICAL DEVICE

Test participants should be asked and allowed to respond. This data is valuable because USERS can be aware of specific concerns as well as positive impressions regarding their use of the MEDICAL DEVICE that are valuable for evaluating the SAFETY and EFFECTIVENESS of use as well as ease of use and USER satisfaction.

17.3.3.3 Instances of confusion or difficulty

Test participants should be asked and allowed to respond. This data is valuable because USERS can be aware of specific concerns as well as positive impressions regarding their use of the MEDICAL DEVICE that are valuable for evaluating the SAFETY and EFFECTIVENESS of use as well as ease of use and USER satisfaction.

17.3.3.4 USE ERRORS and CLOSE CALLS observed during simulated use testing

USE ERRORS, which occur during simulated use USABILITY TESTS, should be followed up by collecting subjective data to enable clarification and root cause analysis of the USE ERROR that includes essential experience and insight from test participants. Likewise, USE ERRORS should be similarly followed up as well as to determine if previously unknown use-related HAZARDS exist in the design of the USER INTERFACE.

17.3.3.5 CLOSE CALLS (not observed)

Test participants should be asked whether they experienced CLOSE CALLS since the CLOSE CALL can be "cognitive" and might not have been observable. If test participants report such CLOSE CALLS, the interview should proceed to 17.3.3.4.

17.3.3.6 KNOWLEDGE TASK STUDY data

Some HAZARD-RELATED USE SCENARIOS cannot be evaluated by only using observation if TASKS involve important knowledge USERS need to operate the MEDICAL DEVICE SAFELY and effectively. KNOWLEDGE TASK STUDIES assess the content of the ACCOMPANYING DOCUMENTATION as it would be typically used by USERS during actual use and the knowledge that is necessary for USERS to enable safe and effective use of the MEDICAL DEVICE.

17.4 Data analysis

It is common for SUMMATIVE EVALUATIONS to result in the occurrence of some USE ERRORS, CLOSE CALLS *and use difficulties* suggesting that no MEDICAL DEVICE or its USERS are perfect. USE ERRORS, CLOSE CALLS *and use difficulties* can reflect USER INTERFACE design shortcomings (i.e. flaws). Sometimes, they reflect shortcomings in the test participants' behaviour, such as conscious disregard for the instructions for use, which are not necessarily related to the USER INTERFACE design or within the MEDICAL DEVICE MANUFACTURER'S control.

Although human beings are imperfect, it is inappropriate to blame the USER when problems occur during SUMMATIVE EVALUATION. The key in any analysis of USE ERRORS, CLOSE CALLS *or use difficulties* is to intensely search for a design-based root cause before attributing the USE ERROR to the USER.

USABILITY TESTS are mostly qualitative rather than a statistically based activity. Any and all USABILITY problems uncovered, particularly those found in a SUMMATIVE EVALUATION, should be thoroughly analysed to determine root causes, and their impact on HAZARD-RELATED USE SCENARIOS should be carefully considered. Regardless of the root cause(s), the MANUFACTURER should conduct a follow-up RISK ANALYSIS of all USE ERRORS, CLOSE CALLS *and use difficulties* that arise during a SUMMATIVE EVALUATION. Root cause analysis of MEDICAL DEVICE USE ERRORS is discussed in detail in reference [32].

The MANUFACTURER should look for any new USE ERRORS or interaction difficulties that would suggest the need for a design change. If new HAZARDS, HAZARDOUS SITUATIONS or HAZARD-RELATED USE SCENARIOS are discovered or improvement is necessary and practicable, then IEC 62366-1:2015 instructs the MANUFACTURER to perform additional USABILITY ENGINEERING effort.

Alternatively, this analysis might determine that no improvement is necessary and practicable for the tested MEDICAL DEVICE. IEC 62366-1:2015 then instructs the MANUFACTURER to perform a RESIDUAL RISK EVALUATION according to ISO 14971:2007.

Modifications of the USER INTERFACE implemented after a SUMMATIVE EVALUATION require follow-up USABILITY EVALUATION. If the change is minor, a desktop analysis might be a sufficient means of confirmation, but only if the modification does not increase the use-related RISK *and does not create the potential for new use difficulties*.

EXAMPLE 1 Revising an on-screen prompt's wording.

EXAMPLE 2 Graphically enhancing a warning by capitalizing the signal word "WARNING".

EXAMPLE 3 Change in logo or branding.

However, even such minor modifications might warrant follow-up USABILITY TESTS, particularly to confirm that a previously detected USER interaction problem has been resolved. For example, if an initial USABILITY TEST showed that USERS misread a button label, the MANUFACTURER would probably need to conduct a follow-up test to demonstrate that they could reliably read the new button labels.

More often, and particularly regarding major design modifications, the best way to confirm that use-related RISKS have been adequately controlled is to conduct a follow-up SUMMATIVE EVALUATION. RISK CONTROLS with far-reaching effects on USER interactions might warrant conducting a complete SUMMATIVE EVALUATION, essentially repeating the previous SUMMATIVE EVALUATION that is redefined as a FORMATIVE EVALUATION. For small modifications with limited effects on USER interaction, a smaller scale, supplemental USABILITY TEST involving fewer test participants and perhaps fewer TASKS than the initial SUMMATIVE EVALUATION can be sufficient.

In unusual cases, the MANUFACTURER might need to study the MEDICAL DEVICE in actual use. Such studies are likely to involve unobtrusive observation of USER-MEDICAL DEVICE interactions, and possibly follow-up interviews with the MEDICAL DEVICE USERS, over a longer period of MEDICAL DEVICE use than is common in simulated-use testing. Annex F contains additional information.

Sample USE ERRORS and possible root causes for those USE ERRORS are listed in Table 7. While in this example the USE ERRORS are described briefly and in a generic manner, actual USE ERROR descriptions should be described in as much detail as possible.

Table 7 – Sample USE ERRORS and their root causes

Sample USE ERROR	Sample root cause for the USE ERROR
1 participant (1 nurse) did not properly secure the syringe in its holder	The syringe clamp required relatively high force to secure the syringe. The nurse tried to engage the clamp, but was not able to apply enough force.
3 participants (1 nurse, 2 PATIENTS) stopped the treatment rather than pausing the treatment	All three USERS drew upon prior experience using a similar MEDICAL DEVICE to operate the new MEDICAL DEVICE, but the new MEDICAL DEVICE did not work the same way (i.e. there was negative transfer).
1 participant (1 technician) programmed ten times the intended dose because he did not add a decimal point when entering the prescribed flow rate	Small text on the display was illegible to the USER who had minor vision impairment (mild cataract). He didn't realize the decimal point was missing.
1 participant (1 PATIENT) did not detect (i.e. notice) that the MEDICAL DEVICE stopped even though the MEDICAL DEVICE repeatedly presented a high-frequency ALARM SIGNAL	ALARM SIGNAL frequency too high to be heard by individual with high frequency hearing loss (presbycusis).
2 participants (1 physician, 1 nurse) did not connect the line to the port	USABILITY TEST artefact: The test participant misunderstood the TASK posed by the test moderator. NOTE USABILITY TEST artefacts are actions induced by an artificiality that would not be present in an actual USE SCENARIO.

18 Document the USABILITY ENGINEERING project

A USABILITY ENGINEERING report can be created to summarize the USABILITY ENGINEERING project for the purposes of communicating with internal and external stakeholders. Importantly, a USABILITY ENGINEERING report is not the same as a SUMMATIVE EVALUATION report, but a USABILITY ENGINEERING report should cite SUMMATIVE EVALUATION results. Such a report should include:

a) *an executive summary;*

b) *a summary of the USE SPECIFICATION;*

c) *a description of the USER INTERFACE;*

d) *a summary of known use problems;*

e) *a description of the HAZARD-RELATED USE SCENARIOS evaluated and why they were chosen;*

f) *a summary of FORMATIVE EVALUATIONS;*

g) *a summary of the SUMMATIVE EVALUATION; and*

h) *a conclusion.*

Annex D provides additional information regarding USABILITY ENGINEERING project end products.

19 POST-PRODUCTION review and analysis

According to ISO 14971, POST-PRODUCTION surveillance is required by the RISK MANAGEMENT PROCESS. This includes evaluating data related to USE ERROR in order to identify strengths and shortcomings of the USER INTERFACE. However, the PROCESS of POST-PRODUCTION surveillance is not included in IEC 62366-1. The design and development PROCESS in IEC 62366-1 for the design and development of a MEDICAL DEVICE ends prior to the POST-PRODUCTION stage.

The POST-PRODUCTION surveillance PROCESS can provide a rich pool of customer complaint data that can be used to support USABILITY ENGINEERING activities. MANUFACTURERS can use these data to identify use-related problems, including those related to USE ERROR. To most MANUFACTURERS, this PROCESS is not new, but it can often be enhanced by collecting more details about events that occur.

In general, USE ERRORS in the field are underreported. This can in part be attributed to the workload involved for RESPONSIBLE ORGANIZATIONS to file reports and the effort needed by a lay USER to file a complaint. MANUFACTURERS should include in the instructions for use contact information for USERS to report ADVERSE EVENTS and complaints.

According to IEC 62366-1, all USE ERRORS should be identified in USABILITY EVALUATION during development. Despite this effort, a MANUFACTURER can attempt to identify previously unidentified USE ERRORS after placing the MEDICAL DEVICE on the market. An example is test market evaluation of new products. This evaluation is usually a limited launch with very tight control of where and to whom the product is initially provided. This effort can allow early intervention on these USE ERRORS, before they cause HARM.

Preparing the specific tools for collecting and managing USE ERROR will be unique for each MANUFACTURER, driven by the RISK associated with use of the MEDICAL DEVICE and the PROCESSES and systems that the MANUFACTURER has available for collecting information.

To fully capture all of the necessary information related to a use-related event, the MANUFACTURER should collect answers to the following questions:

a) *What happened (i.e. what was the unexpected or unwanted result)?*

b) *Was there PATIENT or USER HARM?*