Healthcare and Medical Devices, Vol. 79, 2023, 36–42
https://doi.org/10.54941/ahfe1003471

Usability Validation of Complex Medical Systems

M Smyth and A C B Medeiros

Kinneir Dufort Design Ltd. Bristol, BS1 5BU, UK

ABSTRACT

The end goal of the application of Human Factors to the development of a safety-critical medical device is to validate the device's safety and effectiveness in the hands of intended users. Generally, the complexity of validation studies varies according to the device intricacy. In the case of complex medical systems, such as a robotic surgical system which can be used for prolonged periods by multiple user groups at the same time and may interact with other devices in the operating room, the planning and execution of such studies require much more thinking, organization, and resources. Bearing that in mind, we have grouped the effort involved in validating complex medical systems into four categories for an optimized, practical approach. This paper discusses these four categories and provides essential guidance, based on our experience.

Keywords: Usability validation, Human factors, Complex medical devices

INTRODUCTION

In a complex medical system, the use process extends in opposite directions to include interactions, beyond what is common with simple devices, before and after system operation: from specialist transportation, storage, and installation on one side, through to reprocessing of reusable parts, maintenance and repair, and adequate disposal of the system or its parts at the end of their safe, useful life. Some of these activities are carried out by different user groups and take place in a number of use environments. The focus of this paper, however, is on setup and tear down during system operation, and to some extent reprocessing.

The end goal of the application of Human Factors to the development of a medical device is to ensure the device's safety and effectiveness in the hands of intended users, in the expected use environments. Apart from low-risk devices and a few exceptions, a usability validation study, performed towards the end of the device development process, will be necessary for regulatory submission and approval. Usability testing can involve varying degrees of structure, complexity, and realism, with more formal evaluations required for validation purposes (ANSI/AAMI HE75-2009). Generally, the complexity of validation studies varies according to the device intricacy. In the case of complex medical systems, such as a robotic surgical system which can be used for prolonged periods by multiple user groups at the same time and may interact with other devices, the planning and execution of such studies

require much more thinking, organization, and resources. Bearing that in mind, we have grouped the effort involved in validating complex medical systems into four categories for an optimized, practical approach. This paper discusses these four categories and provides essential guidance, based on our experience.

LEARN FROM THE EXPERTS

In the lead up to the study, it is sensible to invest time in understanding the different user groups' workflows and tasks responsibilities so that task performance can be rated against the correct user, and use problems are not duplicated in the dataset. Learning from clinical experts in the field can provide valuable insights into hospital policies, the use environment, as well as the dynamics of how multiple users and other devices may interact in a complex system scenario. When working in teams, there might be specific collective behaviors and unwritten rules relevant to the study, which can only be observed in the environment and context of use. A range of methods can be used to gather information, from simple online surveys to full immersion in the use environment, with contextual inquiry usually striking a reasonable balance between invested resources and usefulness of results.

Knowledge on the steps of particular surgical procedures, as well as best-practice clinical judgment can be gained from professionals who are not necessarily users of the medical system. A preceptor, for example, is someone who is experienced and competent in their field of expertise i.e., nursing or surgery. They can be assigned to guide the professional journey of a student or trainee when entering the workforce or learning a new skill, ensuring novice users become confident and competent enough to deliver quality care in their chosen field. In a validation study, preceptors can be particularly helpful during training sessions, when the study participants may be asking medical questions which are outside of the study team's field of expertise. In addition, during simulated use, Human Factors professionals may not be equipped to determine if participants' decision on how to perform a task was clinically acceptable because most of us are not clinical experts. In those instances, when the medical professional is working within their own clinical judgement, but we are not sure of what they are doing, a preceptor can help clarify what is observed. Build in time to liaise with the preceptor before and after the study sessions and make it clear that their role is not to tell the participant what to do although, in some instances, it may be acceptable for them to provide some necessary information to participants. The level of involvement the preceptor is to have should be fully defined by the study team, to avoid biasing the study.

Sterility is another topic for discussion with experts. A sterile field is defined for any invasive surgical procedure. Anything below the waist is commonly believed to be unsterile. The defined field must remain sterile throughout the procedure – unsterile items do not enter the sterile field. Hence the need to have scrub and circulating nurse roles to collaborate in passing items to and from the sterile field in an aseptic manner. Remind study participants that the simulation should mimic a real surgical procedure

as much as possible, including observing sterility throughout the procedure. However, let them guide you when appropriate – as a surgical team, they may have specific rules that they work towards. For example, an item such as an endoscope may be re-draped or bumping into an object with someone's back might not be considered a break in sterility. Allow the participants to discuss and ensure everyone is aligned on what to consider sterile at the start of the study. This is their bread and butter, and they should respect sterility. During the study session, if unsure whether sterility has been broken, the follow-up interview always provides an opportunity to investigate the root cause just in case.

Before a large procedure, some hospitals may implement group huddles or a similar activity with the same purpose, so that everyone in the surgical team is on the same page. Ask participants if they would want to replicate such activity in the study – it may help in allowing the study team to hear what their plan of action is for those moments during simulated use where we might wonder "what are they doing now?!". Capturing a group huddle or any other such discussion may also provide supporting evidence for regulatory justifications where acceptability is down to 'clinical judgment'.

When assessing a complex system, it will be undoubtably necessary to simulate safety-critical use scenarios – make sure the relevant possibilities are accounted for in the use-related risk assessment. The simulation of alerts and alarms will most likely require a technician or engineer on site to help carrying out these activities, without compromising the medical system, or resolve unexpected technical problems which may occur unintentionally. Human Factors professionals may not have a technician's background or experience, and the appropriate expert should be on hand to attend to any potential technical issues.

AVOID SCOPE CREEP

To make the most of the allocated resources, it may be tempting to execute multiple studies in one – for example, including parallel activities in the main validation study such as a formative study on a specific aspect of the user interface or marketing research. Although combining studies can be an efficient way of running these activities, objectives can become entangled and lose clarity. If that is the case, it is advantageous to, early in the process, split study protocols into individual documents containing a well-specified set of objectives. This should help define each activity to avoid scope creep, as much as possible.

Let us assume that the main activity is the usability validation of a robotic system with a surgical team in an operating room. Usually, that team would consist of four different user groups: 1) a surgeon, who controls the robotic instruments and is positioned outside the sterile field; 2) a first assistant, who works under the supervision of the surgeon and is positioned in the sterile field; 3) a scrub nurse, who supports the first assistant at the operating table; and 4) a circulating nurse, who remains outside the sterile field, passes supplies to the scrub nurse, operates equipment, such as an electrosurgical generator that interacts with the system, and helps tearing down the

system at the end of the procedure. The protocol for the main study must, consequently, include all critical tasks expected to be performed by these four user groups. Now, let us assume that a second activity would be the usability validation of the reprocessing of robotic instruments. Reprocessing is performed by technicians in the hospital central sterile services department or an external service provider. It is a completely different process, performed by a distinct user group in a separate use environment, therefore the need to write independent study protocols. That way, the clarity of the two activities is not lost or entangled. In addition, each protocol should define what is part of the system and what is not. For instance, a vacuum source may interact with a surgical system, although its user interface is not part of the system being assessed. Where validation and formative studies are happening in conjunction, having independent protocols will provide the study team with clarity in the best interview techniques as approaches in Root Cause Analysis may differ depending on study objectives.

Post study, it is recommended to write a separate report for each parallel activity so that results from one activity do not cast a shadow on and compromise results from another. That should, hopefully, make it easier to confirm which specific objectives have been met and which ones, if any, have not been successful.

PAY ATTENTION TO THE DETAILS

There will be a significant number of details to consider when planning for a smooth study – from regular and consistent communication with participants prior to and during the study, to the simplest subtleties that might be easily overlooked such as preparing name tags, remembering to synchronize watches, or ensuring there is a clock visible to everyone involved in the study.

Given that complex studies may run over a number of days with the same users, possibly in different rooms, in order to evaluate different aspects of the system (training, instructions for use, device labelling, physical interface, etc.), it is indispensable to have a plan of communication between users and observers, as well as a detailed schedule listing every activity for the participants to see before they arrive – from greeting participants upon arrival, to coffee breaks, time to change into scrubs, training sessions, simulated use, follow-up interviews and any other essential activity. Allow time for questions and discussion with participants so that they understand what is required from them every step of the way and feel confident they can perform their tasks as they would in real life. Consider how a complex study design can be optimized to ensure all objectives are met. For example, it would be beneficial to have the validation study on a day which is less intensive for the participants to reduce fatigue. Other scheduling considerations could involve building in training decay by planning a formative study after training is concluded.

Multiple observers will be common for complex studies where there are several users carrying out their own sequence of steps. Most often than not, multiple users will require an equal number of observers, co-existing in the same environment with no interference from observers except for guidance

on what comes next upon completion of a task. It is prudent to assign an observer to each user and remind observers to stick to their users. Consider how users can identify their assigned observer and vice-versa when everyone wears scrubs or uniforms. Name tags and a color-coding system can be helpful – for example, blue tags for scrub nurses and their observers, yellow tags for first assistant and their observers, and so on.

When it comes to the study physical space, take into account where in the different environments an observer may sit for best view of their assigned user. It should not differ from usual studies, nonetheless it is extra important in assessing complex medical systems because there is likely to be multiple observers who need to find their own best spot, as well as a moderator orchestrating the sessions, technicians, clients, and research facility staff running around.

When it comes to data filling, have a plan. Make sure everyone is aware of any naming conventions prior to the study for consistent filling – it will be much easier to find a specific file should you need to double check a piece of information. Every observer should be responsible for logging their own data capture sheets, video or audio files. Alternatively, assigning this role to one person in particular could work, which would allow another team member to carry out study set up and tear down. Whichever way is chosen to divide the work, confirm everyone's role is clearly communicated and the workload is equally balanced throughout the team to prevent overloading a team member and assignments from piling up.

As outlined by the FDA, usability validation testing should be designed so that, among other aspects, all critical tasks are performed during the test by participants who represent the intended users of the device. Critical tasks are defined as tasks, that if performed incorrectly or not performed at all, would or could cause serious harm. Hence, an essential component to usability validation is the identification and categorization of user tasks, which should result in a list of critical tasks (FDA, 2016). This is usually achieved by applying a use failure mode effects analysis (uFMEA) approach. Once critical tasks have been identified and categorized, it is useful laying out the acceptance criteria for each individual task in preparation for the study. This will guarantee task performance can be consistently rated against such criteria, thus creating a robust dataset, and avoiding unnecessary reviews of the data and delays to the analysis and reporting phases.

BE FLEXIBLE

Last but not the least, having some flexibility regarding the planning and running of the study is crucial for a positive outcome. Be aware flexibility comes in many shades.

Before the study, agree as a team on who is responsible for doing or overseeing each activity, including the most mundane ones such as ordering equipment and stationery, printing name tags, sending pre-study emails, letting recruiters know when participants have landed. No one wants to panic at the start of a validation study because materials or participants have not arrived in time.

Other not-so-trivial activities may include preparing cadaver sheets and planning for the types of bodies needed for the study including age, sex, height and weight, any history of previous surgeries and so on – it is not possible to perform a hysterectomy in a male body, or in a female cadaver who already had her uterus removed. What if things go wrong, as sometimes they do? Be flexible and think of ways to improvise to at least simulate the critical tasks of the intended operation, while still respecting the body which has been donated to science.

During the study sessions, there will be plenty going on. Being flexible to a certain extent can ensure all relevant data is captured and is of good quality. For instance, if an observer needs a quick break, they could ask for help from another observer, who might be momentarily idle because their assigned user is on standby. At other times during the session, there may be several tasks happening all at once. Remember that people work differently – a flexible data capture document which allows observers to write everything down in a stream of consciousness may prove itself more valuable than a document with limited space to write or pre-set lists of options to choose from.

In addition to the above, you may wish to consider the flexibility of the expected use process, especially taking into account the complex nature of the operating room and the likely complex accompanying uFMEA. It is important to remember there can be flexibility in these workflows, and it may not be an error for a user to assist their team members in task completion. That said, there are instances were there can be no flexibility, such as the pre-defined approach to sterility – for example, a circulating nurse carrying out a task for a scrub nurse in the sterile field would be regarded as a use error due to the break in sterility. Where flexibility is applicable, keep in mind that the task performance rating can be amended later after the follow-up interview, when there is more time and information to figure out exactly what happened in the execution of specific tasks. For data integrity, save the data capture document as a PDF immediately after the study session as version 01, then re-save the file as version 02 after consulting with the team and clarifying what use problems belong to what tasks and checking acceptability against the pre-defined criteria with the preceptor.

Finally, it is worth considering how flexibility can be applied to safeguard the wellbeing of your study team. Validation and formative studies of any nature can be challenging and taking steps to support the physical and mental health of your team will lead to a more successful, data-rich study. Consider including the team schedule and how this can be optimized to reduce the risk of study fatigue, having other team members as backups to assist if needed, and regular check-ins to keep moral high as the team work long hours away from home.

ACKNOWLEDGMENT

We would like to thank Kinneir Dufort for supporting this work. Kinneir Dufort are a user-centered innovation and product development consultancy, combining creative, technical and user experience expertise to deliver world changing products and services for medical, consumer and industrial clients.

REFERENCES

ANSI/AAMI HE75-2009: Human factors engineering – Design of medical devices. Arlington, VA: Association for the Advancement of Medical Instrumentation, 2009.

US Food and Drug Administration. Applying human factors and usability engineering to medical devices. guidance for industry and food and drug and administration staff, 2016. Available: https://www.fda.gov/media/80481/download [Accessed Jan. 2023].