Healthcare and Medical Devices, Vol. 130, 2024, 140–150
https://doi.org/10.54941/ahfe1004847

Human Factors Validation of Collaborative Medical Workflows Through Multi-User Workflow Simulation: A Case Study in Interventional Radiology

Benedikt Janny[1], Robert Klank[1,3], Julian Glandorf[3], Frank Wacker[2,3], Bennet Hensen[2,3], and Julia Moritz[1]

[1]USE-Ing. GmbH, Stuttgart, 70178, Germany
[2]Department of Diagnostic and Interventional Radiology, Hannover Medical School, Carl-Neuberg-Straße 1, 30625 Hannover, Germany
[3]Research Campus STIMULATE, Otto-von-Guericke University, Magdeburg, Germany

ABSTRACT

This paper challenges the traditional focus on individual 1on1 sessions during workflow simulations and usability testing, which often fail to capture the collaborative nature of medical workflows. To overcome this, a human factors and workflow simulation lab was developed within the research campus STIMULATE. The paper describes the conception, development, and operational capabilities of the lab particularly focusing on collaborative human factors assessment. A novel methodology, based on hierarchical task analysis, is introduced. It breaks down complex workflows into subtasks and assigns them to specific user groups, such as Radiologists and Radiologic Technologists, capturing the intricacies of user interactions in specialized medical environments. The lab's first validation study is presented using the example of the simulation of an image-guided interventional liver biopsy. The study demonstrates the lab's ability to accurately replicate a high percentage of tasks performed by medical professionals in complex procedures, thereby confirming its effectiveness in modelling collaborative medical workflows. It emphasizes the importance of detailed task-level workflow segmentation for analysing human-machine and human-human interactions and introduces specific metrics for measuring usability dimensions like effectiveness, efficiency, and satisfaction.

Keywords: Workflow simulation, Multi-user, Medical, Human factors engineering, Usability

INTRODUCTION

In the dynamic field of user-centered medical technology development, ergonomic considerations have become increasingly paramount. These are predominantly addressed through usability evaluations, conforming to industry standards like IEC62366-1 (IEC 2015) or FDA-2011-D-0469 (FDA 2016). Such evaluations, however, are traditionally conducted in individual sessions with representative users. This approach, while effective in certain

contexts, fails to comprehensively represent medical workflows that frequently involve collaborative tasks among multiple users. This discrepancy between testing methodologies and real-world application poses significant challenges in accurately assessing and optimizing user interactions within medical environments. To bridge this gap, the human factors engineering company USE-Ing. developed a state-of-the-art human factors laboratory dedicated to simulating multi-user workflows together with partners from the STIMULATE research initiative. The study described below served as the initial evaluation of the human factors laboratory.

LITERATURE RESEARCH ON COLLABORATIVE MEDICAL WORKFLOW ANALYSES

A workflow is an umbrella term for many individual work steps or tasks that can run sequentially or simultaneously. Usually, task analysis methodology is used to break down complex workflows into tasks. Among the most popular task analysis methods is the Hierarchical Task Analysis (HTA). HTA is a top-down approach which decomposes workflows into a set of tasks and subtasks. It involves detailed study through interviews, observing users, and examining existing resources like manuals and documentation. HTA outlines tasks using three core elements: the tasks themselves, their hierarchical structure, and the plans for carrying them out. It systematically breaks down tasks into smaller subtasks until these subtasks can be assigned to either the user or the user interface, making them observable and manageable (Stanton, 2006). Workflow analyses have been becoming increasingly popular in the medical context throughout the last decades. They are used to optimize clinical processes, i.e. to shorten treatment times, improve patient throughput and determine staff requirements (Boese & Grote, 2010). In most cases, larger processes and several people are involved in the workflow for these purposes, so that one could speak of organizational workflows here. However, workflow analyses are also used to improve treatment processes through standardization or to determine requirements for instruments, devices and assistance systems (Neumuth et al., 2010). Neumuth et al., 2006 developed a scheme that allows the surgical procedure to be recorded in a standardized way. The basic idea is that an operation is converted into an operation model, the so-called Surgical Process Model (SPM). For this purpose, a distinction is made between three different types of information (activities, state transitions and events). Pfeffer (2017) describes the visualization of workflows and expands this descriptive representation with an evaluation component using various metrics.

METHODOLOGY

For the present evaluation, a workflow simulation of an MR-guided liver biopsy and structured interviews regarding the simulation were performed with medical professionals. Therefore, relevant use scenarios were simulated via test cases.

Task Analysis

To define the test cases, a hierarchical task analysis was done based on discussions with experts and empirical data gathered in clinical observations and interviews. For the task analysis three hierarchical levels were defined. The **first level** represented the phase and is the most general level. Here, the whole procedure is split up into three phases: "Prepare intervention, perform intervention, follow up intervention". The **second hierarchical level** represented the goals. At this level, tasks are defined using the "Generate result" scheme. Tasks are a collection of subtasks that form a coherent process, for instance the preparation of the contrast media injector was one of the goals or dressing in sterile clothes. At the **third level**, the subtasks were defined. Subtasks are differentiated based on the change in the context of use. This can vary according to the dimensions of user group, location and interface element. Subtasks at this level can be assigned to the Perception, Cognition or Action categories. To get an overview of the workflow, we visualized the goals with the user groups involved, the use environment and the resources or interfaces the user interacted with. For this evaluation, we focused on the user groups interventional radiologist and radiologic technologist, as these two user groups are always required in this type of intervention. As use environment, we differentiated between three locations: the scanner room, the control room and the anteroom. The resources and interfaces were named on a rather general level such as imaging software, keyboard, or mouse for instance. A closer definition did not seem to be beneficial as the focus was not on a single device and its improvement but on the observation of the entire process. Figure 1 shows an excerpt of the task analysis.

Figure 1: Example for the hierarchical task analysis.

Procedure and Test Persons

The test cases derived from the hierarchical task analysis were performed with test persons representing the user groups *Radiologic Technologist* and *Interventional Radiologists*. The final sample consisted of five test persons (3 male radiologists, 2 female radiologic technologists) with a mean age of 36.2 years (range: 31–48 years) who were recruited from two university hospitals. They were assigned to the sessions based on their availability; thus

pairing of radiologist and radiologic technologist was random. All test persons had several years of experience in magnetic resonance imaging (range: 2.5 to 19 years), and all of them have already been part of a team who executed MR-guided interventions. Due to the limited availability of radiologic technologists with interventional experience and the complex study design, it was not possible to find an appropriate time slot for participants of both user groups in all three sessions. In two of the three sessions, a radiologic technologist and a radiologist performed their simulation together, in one session only a radiologist participated, and the radiologic technologist was substituted by the test leader. The test sessions were conducted with a team of three human factors experts. The experts were responsible of leading the test persons through the test, interviewing them, observing and documenting their performance and comments, as well as managing the technical systems. The distribution of tasks among the experts was as follows:

- **Test leader:** Instruct both test persons, observe performance of both test persons for root cause analysis, discuss root cause with both test persons, interview radiologic technologist.
- **Observer 1:** Observe and document performance of radiologist, interview radiologist.
- **Observer 2:** Observe and document performance of radiologic technologist, manage technical systems.

The test leader handed over each of the test cases to the two test persons who were present within one session. Test cases were always given to both test persons at the same time. The test persons were responsible for the execution of subtasks (i.e. the distribution of subtasks). They were expected to coordinate who performs the subtasks. For example, a test case that was thought to be done by both test persons together was as follows: "You have already prepared the room and scanner. The patient will now be brought by the transport service in an MR-compatible bed. You will now prepare the patient for the upcoming liver biopsy together so that the next step is that the planning sequences can be recorded next. The intended puncture site is so well positioned that you can position the arms at the side." In total, six test cases were performed that represented the whole interventional procedure. The test persons interacted with the prototypes and mock-ups without help from the test leader. Each observer observed one test person and analysed performance by completing a test protocol. After each test case, a follow-up interview was conducted regarding use problems encountered during the tasks. To assess the realism of the workflow simulation, the test persons completed a questionnaire regarding the realism of the simulated scenarios and missing subtasks. For each task, the test persons were asked to rate the realism on a unipolar four point scale as "very realistic", "rather realistic", "rather not realistic" or "not realistic at all". Finally, the test persons discussed their assessment with the test leader.

Data Collection

To evaluate the quality of interaction between the test persons representing medical professional users and the medical devices, various metrics were assessed during the simulated use (i.e. the test cases). These metrics were

defined in discussion with radiologists to identify the aspects that are most critical in the specific tasks. For example, the joint angles are of relevance when the interventionalist is trying to reach the entry point of the needle in the bore, however, it is of no relevance when the technologist is sitting in front of the monitor to start and stop the imaging sequences. The following table gives an overview of the corresponding metrics, the type of data acquisition and data acquisition tools used. For documentation purposes, measurement data, observer notes, and videos of relevant interaction sequences were recorded.

Table 1. Overview of workflow metrics.

Task Specific Metrics	Type of Data Acquisition [Unit]	Data Acquisition Tool
Number of interactions	Observation [count]	Video camera + Eye tracking + observer protocol
Duration (action)	Measurement [hh:mm:ss]	Video camera + data acquisition software
Duration (Perception/cognition)	Measurement [hh:mm:ss]	Video camera + Eye tracking + data acquisition software
Number of use-related problems	Observation [count]	Video camera + observer protocol
Number of simulation-based problems	Observation [count]	Video camera + observer protocol
Joint angle of lower back, cervical spine, shoulder, elbow	Measurement [°]	Motion Tracking (IMU)
Task completion	Observation [binary]	Video camera + observer protocol
Overall Workflow Metrics	**Type of Data Acquisition [Unit]**	**Data Acquisition Tool**
Walking distance	Measurement [m]	Step counter (mobile phone app)
Number of patient translations out of/into bore	Observation [count]	Video camera + observer protocol
Subjective realism ratings	Interview [unipolar four-point scale]	Questionnaire

Laboratory Setup

The human factors laboratory was conceived to enable in-depth ergonomic analyses, circumventing the limitations of field studies in complex operational environments like surgical theaters. The lab's design incorporates flexible mock-ups and prototypes, which facilitate early-stage, cost-effective variations in human-machine interface design. This adaptability is crucial in visualizing and understanding the impact of these interfaces on collaborative medical workflows. The laboratory consists of three rooms: the main test room, the control room and the observer room. **The main room** is where usually medical interventions are simulated. For this study, it contained the physical prototype of an MR in original dimensions in order to simulate realistic work processes. A simplified patient table allowed a patient dummy to be transferred to the simulated MR by the radiologist. Depending on the research focus, the prototype could be adapted, extended or, if necessary, dismantled. If necessary, the main room can be supplemented by additional medical devices, e.g. a mobile patient bench, a monitor, as well as audio and light simulation. In the adjacent **control room**, assistants such as radiologic

technologists can monitor the procedure and simulate the control of the MR scans. In this case, a low fidelity click dummy was used to validate the human-machine interaction with a control software prototype. The two rooms are acoustically shielded and visually connected by a window. The observation room uses mirrored glass windows to provide technical monitoring, documentation and observation by experts concealed from the test subjects. The MR simulation control system and an intercom system for queries are located here. In addition, adjacent rooms and test subjects are monitored using cameras, microphones and sensors. Figure 2 shows a floor plan of the human factors laboratory with the study setup.

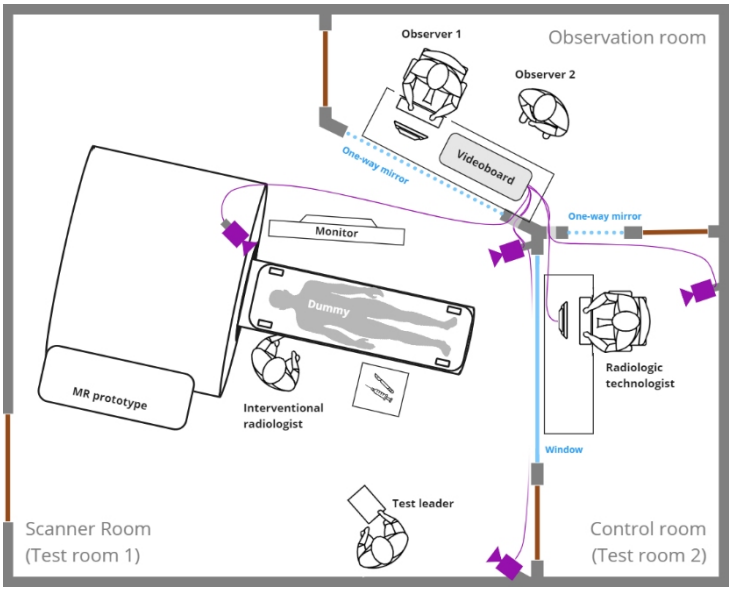

Figure 2: Room layout with test subjects and camera positioning (violet).

Figure 3 shows various views of the main room and the utilized MR mock-up.

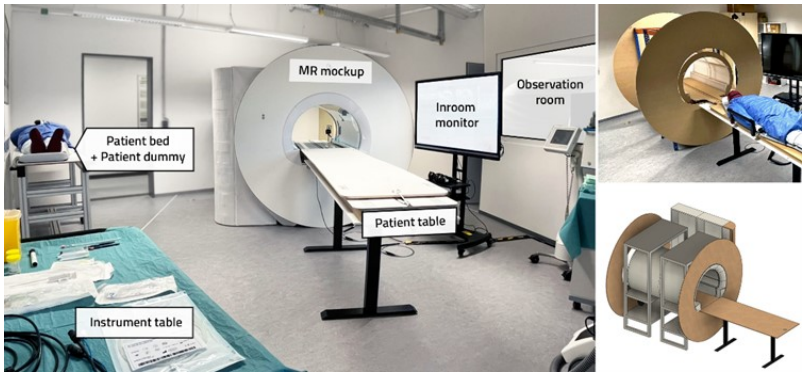

Figure 3: Main test room with evaluation (left). Views of the mock-up (right).

RESULTS

Simulation Realism and Missing Tasks

The focus of this study was the assessment of the experienced realism of the medical workflow by the test persons and the identification of potentially missing subtasks. Therefore, the human factors experts asked the test persons about the deficiencies in our simulation. Figure 4 shows the results for the assessment of realism for user group interventional radiologists. Columns represent the subtasks, the three lowest lines the assessments of the test persons. Dark green is very realistic, light green is rather realistic, light red is rather not realistic, and dark red (not present) is not realistic at all.

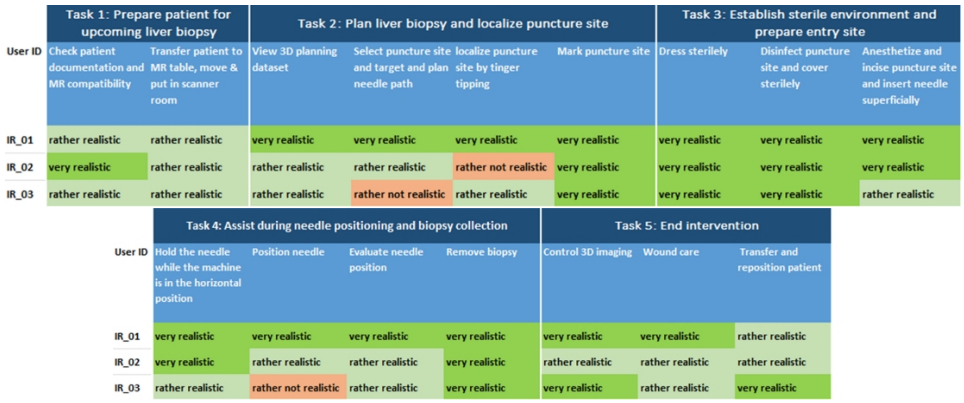

Figure 4: Assessment of realism of simulation for user group interventional radiologists.

Figure 5 shows the results for the assessment of realism for user group radiologic technologist. The structure is the same as in Figure 4.

Figure 5: Assessment of realism of simulation for user group radiologic technologist.

As shown in the figures, both user groups assessed most subtasks as very realistic or rather realistic. However, the radiologists rated three subtasks as rather not realistic, the technologists seven subtasks. For the user group interventional radiologists, all subtasks that were rated as "rather not realistic" were subtasks that involved the imaging. The subtask "select puncture site and target and plan needle path" in reality involved 3D imaging and normally requires the user to scroll through the 3D data visualization and rotate the planes. However, in our rather simple software prototype, the users could only navigate through slides and a short video that showed the images. That means, no free navigation through the data was possible for the users. The other two subtasks that were assessed as rather not realistic were "localize puncture site by finger tipping" and "position needle". Both subtasks normally require live imaging to assess the position of the finger on the body surface or the position of the needle in the patient's body. We simulated both with images depicting a finger on the body surface or a needle in the body, however this did not represent the actual position of the test person's hand or the needle progress. The subtasks that were assessed as "rather not realistic" by the radiologic technologists were "prepare contrast medium pump", "connect contrast medium pump", "start contrast pump", "position coil", "record 3D planning dataset and import layer into sequence", "start/stop live imaging", and "perform 3D imaging". As with the radiologists, the imaging subtasks were criticized due to the missing correspondence between the images and the user's actions. Additionally, one technologist criticized the contrast medium pump because it was missing the tubing set. The other technologist criticized that the positioning of the coil was not realistic in the simulation because the radiologist took the coil and positioned it on the patient simulator. This is normally done by the radiologic technologist. Regarding the missing subtasks, the radiologists mentioned three aspects. The first aspect was the missing communication with the patient. They normally talk to the patient and explain the procedure. In our simulation, the patient was simulated by a manikin and the patient communication was not a relevant part of the simulation. The second aspect missing was the fact that imaging normally has to be repeated more often. In our simulation we went through the whole process once without repetitions or complications. It was mentioned that this is almost never the case and is an oversimplification that limited perceived realism. The third aspect related to organizational aspects. Two radiologists mentioned that they write notes for the ward or for the pathologist, the other radiologist mentioned that he normally documents the procedure for billing (including saving the instrument tags). All these tasks were ignored in our simulation, as we thought that these tasks were not of relevance for the intervention. Also, the radiologic technologists mentioned the communication with the patient and the oversimplified process without repetitions as missing tasks. In this case, the technologist mentioned, that normally, more than one biopsy is taken. Additionally, one technologist mentioned that she normally covers the patient table with a sterile sheet, and that she places the coil and headphones on the table during preparation.

Details of Mock-Ups

The click dummy we used was a slide show of the control software of a MR scanner with embedded video sequences. These sequences were meant to simulate scrolling through a 3D dataset and to simulate live imaging. However, the degrees of freedom within this simulation were perceived as too narrow by the radiologists. They did not have the option to configure anything nor did the video show the position of the needle progress. This prototype was suitable to give an impression of the task, but for someone who is experienced with this task, the misfit of action and depicted video was irritating. Another component of the simulation mentioned to be not realistic was the patient table. When the study was conducted, the patient table could only be moved through a device that lifted the table slightly. This device was operated by the test leader. However, it emitted loud noises and was not realistic in its behavior. Additionally, the patient table was built with a table base that allowed the users to place their feet under the table. The real patient table is more like a block that limits the freedom to position the feet close to the table.

Organizational Challenges

Another finding that we want to point out is the increased organizational effort that is linked to group usability tests. The test sessions require that representatives of all involved user groups are available. Especially in the niches where experts are rare, this can result in a whole team that is out of action during the test session. If additional requirements such as unknown collaborators or specific demographics are relevant, it can easily result in immense effort for test person recruiting. Another aspect that should be considered is the start and end of the usability tests. We invited both participants at the same time and started together. However, in the hospital, the interventionalists normally join the technologist later, when the room, the devices and the instruments are prepared. This resulted in our study in the unrealistic scenario that the interventionalists supported the technologists and took over preparation subtasks from them. In the interviews, the radiologists mentioned that they also support the technologists in the hospital in case they are present during this phase, however this is rarely the case.

Technical Findings

As shown in Table 1, besides subjective data collection a variety of objective measurement data was planned to be recorded during the workflow simulation. This included, e.g. joint angle metrics for the movement of the test persons or walking distance metrics measured by a step counter. However, the wirelessly transmitted data from various systems were disturbed by interference, which may be related to structural conditions (real MRI in the building) or parallel used radio systems (microphones, Wi-Fi, mobile phones). This led partially to data corruption. Future laboratory optimizations will focus on these problems to ensure reliable data recordings during human factors evaluations.

CONCLUSION

In conclusion it can be stated that the developed human factors laboratory achieved a high degree of realism. The results of this study demonstrate that the Workflow Lab was able to accurately replicate 95.8% of the work tasks of radiologic technologists and 94.1% of the tasks of interventional radiologists in the context of an interventional liver biopsy (rated as "rather realistic" or "very realistic"). This remarkable level of task replication fidelity underscores the lab's effectiveness in realistically modeling complex, collaborative medical workflows. Apart from that, several areas for improvement could be identified. For the simulation, we focussed on the intervention and ignored that radiologists and radiologic technologists work in a hospital that requires them to communicate with the wards and other functional areas such as the pathology. The radiologists mentioned that communication is of high relevance. Assuming that interventions are done in general anaesthesia, we used a patient dummy that did not simulate any communication. As a learning for future studies, we will equip the patient dummy with a speaker and simulate the communication with help of one of the observers. Another aspect of improvement related to the click dummy used. This dummy did not simulate live images. For future studies, such click dummy should offer the option to select different settings and it should offer the option to navigate through a 3D dataset. Based on the identified learnings, the human factors laboratory will be improved, and further evaluations will be conducted simulating different image-guided interventions. The implications extend to the design and assessment of medical devices, potentially improving workflow efficiency, user satisfaction, and overall patient care quality.

ACKNOWLEDGMENT

The authors would like to acknowledge the research campus STIMULATE and the Institute for Diagnostic and Interventional Radiology at Hannover Medical School for their support. This research was funded by the German Federal Ministry of Education and Research (grant numbers 13GW0473D and 13GW0473B).

REFERENCES

Boese, A., & Grote, K.-H. (2010). Workflow analysis as tool for development of medical devices, a white paper. In Proceedings of The World Congress on Engineering (Vol. 3, pp. 2149–2152).

FDA (2016). Applying Human Factors and Usability Engineering to Medical Devices -Guidance for Industry and Food and Drug Administration Staff. FDA-2011-D-0469.

IEC 62366–1 (2015). Medical devices - Part 1: Application of usability engineering to medical devices (IEC 62366-1:2015 + COR1:2016 + A1:2020); German version EN 62366-1:2015 + AC:2015 + A1:2020, Beuth publishing.

Neumuth, T., Durstewitz, N., Fischer, M., Strauß, G., Dietz, A., Meixensberger, J., Jannin, P., Cleary, K., Lemke, H. U. & Burgert, O. (2006). Structured recording of intraoperative surgical workflows. In Medical Imaging 2006: PACS and Imaging Informatics (Vol. 6145, p. 61450A). International Society for Optics and Photonics.

Neumuth, T., Kaschek, B., Neumuth, D., Ceschia, M., Meixensberger, J., Strauss, G., & Burgert, O. (2010). An observation support system with an adaptive ontology-driven user interface for the modeling of complex behaviors during surgical interventions. Behavior research methods, 42(4), pp. 1049–1058.

Pfeffer (2017). Development of a modular method for the evaluation of human-machine-interaction with application for medical devices, Dissertation, Institut für Konstruktionstechnik und Technisches Design, Universität Stuttgart, ISBN 978-3-922823-97-1

Stanton, N. A. (2006) "Hierarchical task analysis: Developments, applications, and extensions", in: Applied Ergonomics, Volume 37, Issue 1, pp. 55–79, ISSN 0003–6870.