Human Factors in Healthcare 5 (2024) 100070

Contents lists available at ScienceDirect

Human Factors in Healthcare

journal homepage: www.elsevier.com/locate/hfh

Case Report

An academician's approach to the application of human factors standards: A case study on a liver support system

Jeremy Opie [a,b,*], Euan McCann [b], Harris Qureshi [b], Thomas Brookshaw [c], Eloy Erro [c], Clare Selden [c], Ann Blandford [a,b]

[a] Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London, London, United Kingdom
[b] University College London Interaction Centre (UCLIC), University College London, London, United Kingdom
[c] Institute for Liver and Digestive Health, Division of Medicine, Faculty of Medical Sciences, University College London, Royal Free Hospital Campus, London, United Kingdom

ARTICLE INFO

Keywords:
Human factors
Regulatory standards
Healthcare
Normal use

ABSTRACT

Until recently, academia typically did not engage with the development of medical technology but instead developed techniques, processes, and lab-evaluated studies that were picked up by industry. With the introduction of rapid prototyping and access to information via the internet, more and more medical technology is being developed within academia from conception to product. However, academia still suffers from siloed research groups, access to partial knowledge, and anticipating, at the outset, the requirements of a final product necessary to engage all the required multidisciplinary input. Support for the design and development of the technology can often be provided by other academic departments; however, the current environment of isolated research, often due to funding remit restrictions and limited resources, limits this. In this paper, we highlight lessons learnt within academia about using regulatory standards within a multidisciplinary research project to assess the usability of a novel medical device; specifically, the current guidance is not sufficient for new practitioners to apply regulatory processes straightforwardly. We report on the assessment of a novel medical device being evaluated to regulatory standards, using the system under '*normal use*', and discuss how using ecological validity as a guide can help ensure that the device is effective, efficient, and most importantly safe.

1. Introduction

Currently, there are over half a million medical devices on the market (Campbell, 2007), which are involved in a significant number of adverse events (AAMI-FDA, 2010). The most recent report from the Medicines and Healthcare products Regulatory Agency (MHRA) indicated that 14,189 adverse event reports were submitted between 2011-2013, which included 4,955 serious injuries and 309 deaths (MHRA, 2014). Poor usability has been identified as contributing to many of these adverse events (Tase et al., 2022, Wiklund, 2022) resulting in harm to both users and patients (Fairbanks & Caplan, 2004), and in death (Thimbleby et al., 2015).

Historically, there has been a clear distinction between medical device development in academia and industry. Specifically, academia focused on the science behind improving medical treatment processes and industry adopted the science into practice (Gelijns et al., 1989). This was due to experts only working within their specific fields, as it was difficult and costly for academics to engage in translational development. However, in recent years, as rapid prototyping and access to information via the Internet have become more prevalent, more academic projects are being developed into medical devices in-house. However, academia still suffers from siloed research, funding restrictions, and difficulty in finding other academic departments that can support development during the grant-writing phase. Additionally, academics now have access to information about all requirements necessary to develop a product from conception to manufacturing and the regulatory standards required to fulfil. However, as most academics have had little to do with regulatory documents and potentially have never had to adhere to them in the past, this can be a daunting task.

To improve the safety of medical technology, human factors regulatory processes have been developed. They help to minimise risks by identifying processes and features that are difficult to understand, learn, and use to ensure the device is efficient, effective, safe, and

* Corresponding author at: Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University College London, London, United Kingdom.
E-mail address: j.opie@ucl.ac.uk (J. Opie).

https://doi.org/10.1016/j.hfh.2024.100070
Received 4 April 2023; Received in revised form 17 November 2023; Accepted 13 March 2024
Available online 26 March 2024

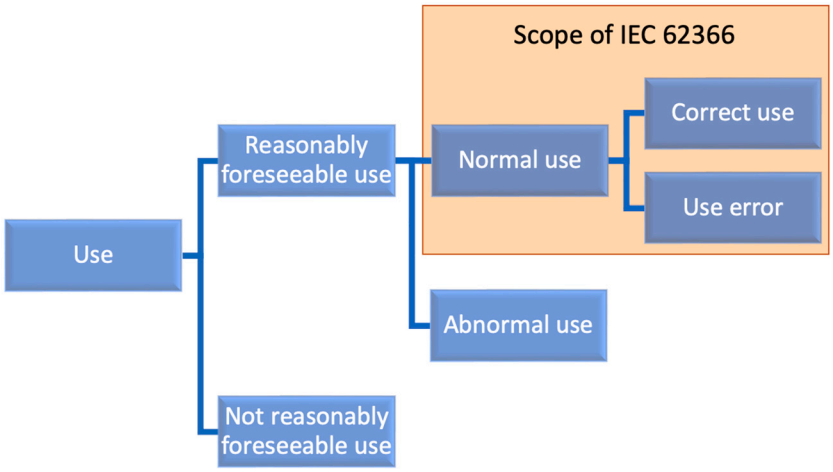

Fig. 1. Types of use as described in IEC:62366.

provide user-satisfaction (62366-1:2015, 2015). These regulations have also led to an increased understanding of human factors in medical device development (Pelayo et al., 2021). The standards and guidelines that have been developed, most notably by IEC (e.g., 62366) and the FDA / AAMI (e.g., HE75), inform or govern the design and evaluation of interactive medical devices from a user-centred perspective. Medical device developers are required to create a usability report throughout the product engineering lifecycle (62366-1:2015, 2015) that measures effectiveness, efficiency, safety, and user satisfaction within specific scenarios (9241-11:1998, 1998). The scenarios are designed to represent realistic environments in which the end-user would be required to perform selected tasks and objectives (Petrie & Bevan, 2009), and can elicit insights to improve product design (Blandford et al., 2007).

However, within the current regulatory framework, the device manufacturers determine the acceptable level of risk and what is considered 'normal use' (Tase et al., 2022), which is defined by the instructions for use (IFU) (62366-1:2015, 2015).

'Normal use', as described in IEC:62366-1 is the '*operation, including routine inspection and adjustments by any USER, and stand-by, according to the instructions for use or in accordance with generally accepted practice for those MEDICAL DEVICES provided without instructions for use.*' (62366-1:2015, 2015). IEC:62366-1 expands on this further by identifying three possible outcomes when using a device: 'correct use', 'use errors', and 'abnormal use' (62366-1:2015, 2015), see Fig. 1. Abnormal use is defined as a conscious or deliberate act or omission that compromises normal use. Normal use is the operation of the device as prescribed by the manufacturers leading to correct use, and use errors are errors that occur when a different result occurs than that intended by the manufacturer or the user, leading to adverse events (62366-1:2015, 2015). Therefore, as normal use is defined by the manufacturers via the IFU, or in accordance with standard practices when no IFU is required (62366-1:2015, 2015), it is important that the IFU is properly evaluated. In the case of novel medical devices, where there is no standard practice on which to base assumptions, the IFU is even more important for ensuring effective, efficient, safe, and satisfactory use of the device, as poor instructions can lead to use errors (Burlington, 1996). However, the core requirement of IEC:62366 is to identify user errors in relation to safety, and although efficiency, effectiveness, and user satisfaction are important and using the methods described in the standards can help to identify and improve these characteristics, they are out of scope of the standard.

This definition of 'normal use' raises various issues when it comes to the design and – particularly – the testing of medical devices, with the colloquial understanding of 'normal use' concerning the realities of clinical practice. This subtle difference between the two definitions of 'normal use' are difficult to comprehend when first introduced, although, HFEs who have prior experience with regulatory standards and practitioners commonly refer to this as 'intended use'.

Although there are a small number of studies evaluating the usability of IFUs (Vincent & Blandford, 2015, Escalada-Hernández et al., 2019) there are none, to our knowledge, that evaluate how well they align with clinical practice. Additionally, accommodating all healthcare contexts is difficult, as each hospital has its own guidelines, policies and protocols that affect clinical practice (Borsci et al., 2018); therefore, it is important that clinicians, human factors specialists, and engineers work closely to ensure medical devices are safe, effective, and efficient (Catchpole & Alfred, 2018).

Another key aspect of IEC:62366 is to provide medical device developers with appropriate resources so that all medical devices are properly assessed to ensure they are safe for use. To assist with comprehension, IEC:62366 was modified in 2015 by splitting the document into two documents, IEC:62366-1 and IEC:62366-2. Part 1 of the standard was updated to include current concepts of usability engineering whilst also providing a clearer understanding of the nature of the standard as having a focus on safety. This allowed Part 2 to become a technical report containing tutorial information to support compliance with Part 1, as well as providing guidance and detail of usability engineering methods that go beyond safety, incorporating accuracy, completeness, efficiency, and user satisfaction, which although important are not essential to meet the requirements of the standard. However, when reviewing the documents together it becomes apparent that there is minimal cross-referencing from Part 1 to Part 2. This could lead to some aspects of the usability evaluation being missed as some may ignore Part 2, regarding it as an optional support document, rather than cross-referencing all elements of Part 2 that are mentioned in Part 1 or adding additional information to Part 1 stating that further information can be found in Part 2 with an accompanying cross-reference. For example, in Part 1 Section 3.27 describes 'User Interface Evaluation' as the '*process by which the manufacturer explores or assesses the user interactions with the user interface*' (62366-2:2016, 2016). To clarify, it has a note listing various methods by which this can be accomplished; however, there is no cross-reference to Part 2 where these methods are described in more detail.

Despite the detail provided in Part 2, it provides a caveat in Section 1.2 (Purpose) that the document is not intended to be the only source of guidance, and that it should not be used to substitute human factors experts, but rather to provide a general understanding. Additionally, it states that the document '*does not describe a specific set of usability engineering activities that suit all design projects*'. Therefore, the activities used to assess a medical device are left to the design team, specifically, the usability specialist, who '*should have relevant, appropriate training*

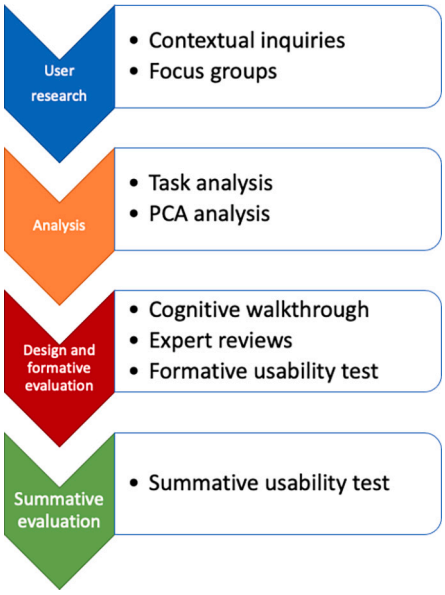

Fig. 2. Example of a usability engineering project, highlighting the phases and methods within each phase.

and have appropriate medical device domain knowledge' (62366-2:2016, 2016).

An example project with the accompanying usability engineering methods is provided in section 7 (Overview of the usability engineering process) of Part 2, illustrated in Fig. 2, which highlights four main phases: user research, analysis, design and formative evaluation, and summative evaluation. To accompany this, a list of potential methods is prescribed in Annex E of Part 2, indicating which method is suited to which phase of the process.

To analyse data from the human factors methods, the perception, cognition, and action (PCA) model (Food and Drug Administration, 2016) is recommended in IEC:62366-2 Section E.15 and is proposed as one of the best methods for understanding the root cause of use errors and the kind of interaction that is required by the user for each task. Perception errors are those that involve a user's ability to see or hear information clearly. Cognition errors can be one of three types: memory failure, rule-based failure, and knowledge-based failure. Memory failure is when the user is unable to recall previous knowledge or skips a step in the procedure. Rule-based failure is when a generally accepted rule is misapplied. A knowledge-based failure is when users have an incorrect mental model of the device or process they are trying to perform. Action errors occur when a participant performs an action incorrectly with the knowledge of what action they should do or they are unable to complete the correct action.

Most of the literature on evaluating novel medical devices reports on controlled 'laboratory' user studies or simulation labs. However, there is growing literature on the importance of the ecological validity of studies. van Berkel et al. (2020) describe seven dimensions of ecological validity that should be taken into consideration when assessing technology. They are: user roles, environment, training, scenario, patient involvement, software, and hardware. If the ecological validity of the evaluation is not taken into consideration during the study design, researchers may not distinguish between real use errors and those occurring due to poor study design. Additionally, some real use errors may be masked by unrealistic environments or low fidelity software or hardware. Therefore, it is important that this constraint is well understood prior to carrying out studies.

Below are two examples taken from real world experiences (personal communications with manufacturers' representatives) that highlight the importance of ecological validity during product assessment:

A device manufacturer was running a user study for a re-designed medical device. The study recruited qualified practitioners from several hospitals (representative users) as study participants. When setting up the device, one participant entered a different number from that specified in the test scenario. When asked about this in the debrief, the participant responded that they had entered the number that they would always enter in the circumstances defined in the scenario in their hospital. In other words: the scenario was not ecologically valid, as far as the participant was concerned, and they behaved in line with their usual practice. The ensuing debate between the manufacturer and the regulator focused on the question of whether this was a use error that brought the device safety into question.

In a situated observational study of devices being used in a hospital, a health professional (device user) reported that they did not feel that they had had sufficient training on the device to be able to use it to its full effect. When this was raised with a representative of the manufacturer, their response was that all users were expected to have received full training on the use of the device, and to maintain their competence through further training. This was specified in the instructions for use. In practice, there was insufficient capacity to comply with this IFU, making 'normal use' impossible to achieve in clinical practice.

This research came about as the initial research team, made up of biomedical scientists, had developed a new method to treat patients suffering from liver failure. To provide this treatment they began the process of developing a novel medical device for clinical and commercial use, using resources from the internet and discussions with other researchers as a guide to regulatory requirements. In doing so, they ensured that the device was able to do what it was designed to do. However, as the prototype became more mature, they discovered that there were specific regulatory guidelines for usability that needed to be complied with. In an attempt to comply, the research team developed IFUs, conducted a risk analysis, and identified intended user groups, eventually leading the research team to be unsure of how to proceed to ensure compliance with the standards. They identified some previous work carried out by the human factors (HF) research group in the medical technology domain and made a connection. However, even though the HF group were aware of the usability standards and had significant domain knowledge of the methods mentioned in the standards, up until this point no one in the group had conducted any formative evaluations with the intention of meeting regulatory standards. In this study, the regulatory standard IEC:62366 was used, as it is the regulatory standard used within the UK where this study was conducted.

The original aim of this work was to evaluate the usability of a novel medical device, HepatiCan™. Initially, we focussed on assessing the IFUs, as the system design was already mature with little we could physically modify, to ensure they provided the user with all the information they needed to complete their tasks. Then using updated IFUs, we conducted a user study following the regulatory guidelines in IEC:62366. Throughout the period of the analyses, and through further reflection, it became clear that we needed to revisit the concept of 'normal use' and how it relates to the design of IFUs and to ecologically valid evaluation of the device and its IFU. Additionally, we reflected upon our experience of adopting regulatory standards for the first time and highlighted some lessons learnt along the way. Therefore, this paper focuses on the disconnect between what is considered 'normal use' and lessons learnt when developing medical devices within academia.

2. Case study: HepatiCan™ - a liver support system

HepatiCan™ is an innovative technology designed to effectively treat more patients with liver failure than current best practices. This treatment has been designated by the regulatory body (The European Medicines Agency) as a combined advanced therapy medicinal product (ATMP) and medical device. The cell therapy, designated ATMP, delivers liver function to the patient temporarily to address the lack of function in a failing liver. The device not only acts as a delivery de-

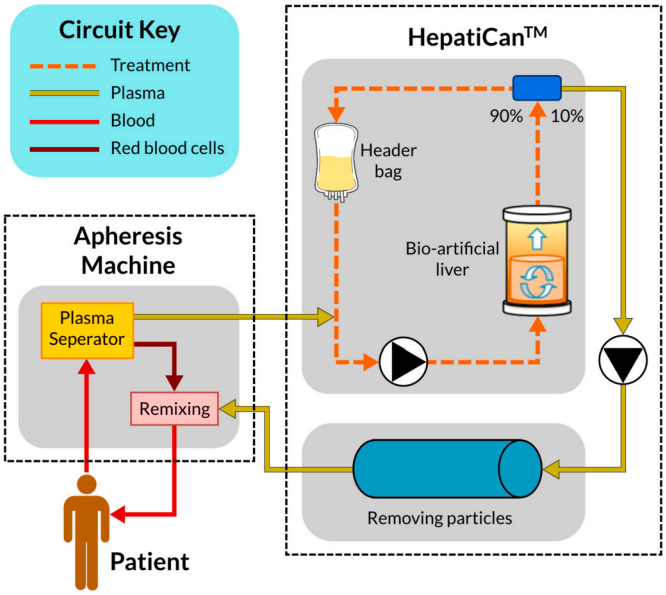

Fig. 3. HepatiCan™ treatment diagram.

vice for cell function, but also provides added functionality and safety in specific elements of treatment. The science underpinning this novel treatment technology had already been generated and both lab and preclinical in-vivo testing had already been conducted (Selden et al., 2017, 2013, Selden & Fuller, 2018). At the time this study was conducted, the team were developing the device element to deliver this life altering process. Technical design was based on an existing research protocol to deliver, temporarily, liver function, with "design requirements for the medical device component", provided to a Design Team, defined to meet the needs of the system for delivery to a patient in an intensive therapy unit (ITU) setting, particularly infection prevention control, whilst also complying with the relevant regulatory standards, functionality, and the practicality of usage in a clinical space. The design team included clinicians, scientists, engineers and, probably most importantly, patients via Patient Public involvement (PPI); the latter enabled a patient centric design from the start. The prototype comprised a physical cabinet housing the components needed for delivery of the ATMP, some additional functional elements, and a digital (touch screen) used to passively read inputs from sensors to alert the user of issues with the device (i.e., pressure buildups, oxygen levels, etc.) and to log data for post-treatment analysis. However, despite the hardware development of the prototype being mature, the software to monitor the device was still in the early stages of development.

HepatiCan™ is an extracorporeal machine; that is: it acts as a bio-artificial liver and is located outside the body. It is a stage-2 device that is not connected directly to a patient but via an apheresis machine, which removes the blood from a patient and separates the blood from the plasma, with the plasma being sent to the device. HepatiCan™ then pumps the patient's plasma through a BioArtificial liver providing time for the patient's liver to repair itself over a 72-hour period. A diagram of the transfer of fluid is shown in Fig. 3.

The device contains two circuits: a fast circuit, with a fast flow rate that mimics portal venous blood flow through the liver, and a slow circuit, mimicking peripheral venous flow used to move the treated plasma through biosafety, back to the patient, at a slower flow rate. The speed of the slow circuit is dictated by the original blood flow from the patient and the patient's ratio of plasma and blood cells (haematocrit). From the slow circuit, the treated plasma is delivered to the apheresis machine where it is restored with the cellular components of blood and returned to the patient.

The system requires users to install tubing and sensors into the cabinet to allow fluids to be connected to the apheresis machine, be treated, and then return to the apheresis machine, which is connected to the patient. During treatment, the GUI is used to monitor the status of the treatment. This is completed in conjunction with controlling pump flow rates at the pump, pressures, temperatures, and the patient status during treatment.

The accompanying documentation, including the IFUs, were all generated by the development team in-line with the construction of the medical device. The team generated three IFUs, with one focused on installing tubing into the device, another priming the system for use, and the third focusing on using the device. Initially, these documents included background information for the device, and for the protocol for each of the three IFUs. Additionally, it included a detailed description of the use and design of various components, from the fast and slow circuits through to the pumps. Large bodies of text detailing the steps required to perform each task were also given, with only a few accompanying images, most of which did not show action, but rather just a screenshot of the graphical user interface (GUI), line drawings of the fast and slow circuits, or still images of the device in various stages of setup. However, from more than 20 tasks, each with a number of intricate sub-tasks, only two sub-tasks from a single task included step-by-step instructions with accompanying images. As this was the first iteration of the IFUs, it allowed the human factors specialists to identify all tasks and work towards improving the IFU.

3. Method

Using standard IEC:62366 as a guide, we conducted a number of human factors analyses to evaluate 'normal use' as prescribed by the IFUs. However, the Human Factors Engineer (HFE) specialists were brought into the project during the design and formative evaluation stage, refer to Fig. 2, and the team who developed the prototype had already conducted a risk analysis and identified intended user groups; thus, some human factors analyses were conducted retrospectively to provide the HFE specialists with the opportunity to understand the functionality of the device. For this study four methods were selected to assess the device and its instructions for use; these were hierarchical task analysis (HTA), cognitive walkthrough (CW), Technical Knowledge Elicitation (TKE) and a user study, which consisted of a survey and a semi-structured interview, see Fig. 4. Further, the PCA model was used to explore the findings from these analyses where relevant. These methods were selected as (a) being recommended within IEC:62366-2 (62366-2:2016, 2016); (b) being suitable at this stage of development, when working under clinical constraints of limited access to representative participants for a user study; and (c) enabling the HFE specialists conducting the study, who did not have a clinical background, to familiarise themselves with the details of the device interaction through their analysis.

Through each analysis, the insights gained into the usability of the system and proposed changes to either the design or the IFUs were noted. Following the four analyses, the findings were reviewed to further assess what we could learn from them about WAI/WAD and about the expertise of the analysts.

3.1. Participants

During the hierarchical task analysis and cognitive walkthrough no participants were required, as these methods were undertaken by the research team. The technical knowledge elicitation had one participant; due to difficulty in recruiting participants at the time of the study the participant was a member of the research team that developed the device. Even though this participant understood the concept and biology of the device, they had limited prior knowledge about setting up or operating the device.

The user study involved four participants who were recruited as experts in their field. As HepatiCan™ is a novel device, there are no existing user groups to recruit from, so we recruited participants that

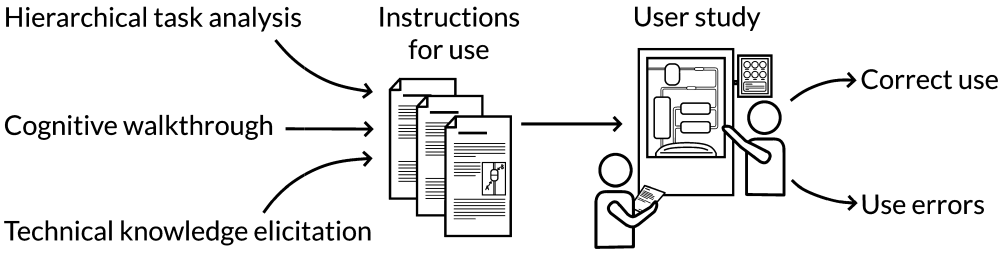

Fig. 4. Graphical representation of the study design.

had been identified by the development team as intended users due to their experience of working with extracorporeal systems. Additionally, due to the constraints of recruiting participants during COVID-19 and the particular difficulty of recruiting clinical personnel, we ensured we covered a full spectrum of intended users for this analysis. These analyses were undertaken with ethical clearance and informed consent was obtained from all participants (UCLIC/ 1819/006/ BlandfordProgrammeEthics).

3.2. Analyses

The focus for the first three analyses (HTA, CW, TKE) was on the design of the instructions for use, and at the conclusion of each analysis the instructions for use were amended in response to findings.

3.2.1. Hierarchical task analysis (HTA)

HTA was conducted retrospectively as the GUI was already developed, but also as a means to identify key tasks and to help the HFE specialists gain a better understanding of the device. The HTA involved deconstructing the user's high-level tasks into subtasks to better understand the workflow, as explained in IEC:62366-2 Section 9.2. This allowed us to (1) confirm with the research team that the tasks and workflow were correct, and (2) explore whether changes could be applied to the workflow to improve the setup and use of the device, of which none were discovered at this point.

3.2.2. Cognitive walkthrough (CW)

The CW involved evaluating the instructions for each task listed in the IFUs by the HFE specialists, which is outlined in IEC:62366-2 Section 16.2.3. However, the interpretation of CW described in this section differs from that defined by Spencer (2000) in which a CW should be conducted by HFE specialists acting as the user as they have the skills to identify cognitive constraints, whereas the regulatory standards reference using representative users performing think-aloud analysis being observed by HFE specialists. Within Appendix E of IEC:62366-2, this is clarified by stating that this is sometimes called a pluralistic evaluation and is more inline with the Technical Knowledge Elicitation we carried out. Therefore, within the context of this analysis we conducted a traditional CW with only HFE specialists. This allowed us to assess whether users would understand what tasks were required, whether a task was available, and whether the user would associate a task with its intended outcomes. It also allowed them to assess the learnability of the device and whether there was sufficient feedback from the system to allow them to move to the next task. It was found that many tasks in the IFUs only consisted of text with no images or diagrams to support users in identifying how to achieve their tasks, see Fig. 6a. Additionally, some instructions lacked detail, which may result in users having to make assumptions to complete tasks with the potential of use errors occurring. Additionally, this analysis highlighted that alarms were only auditory and that there was a lack of detail in the instructions for use when setting pump flow rates, exporting data, and inserting temperature probes. Based on these findings, more detail was added to the IFUs, including supporting images to improve sensemaking. Visual alerts were added to the graphical user interface (GUI) to provide additional methods of identifying alarms rather than being reliant on only audible alerts.

3.2.3. Technical knowledge elicitation (TKE)

To understand whether the IFUs provided the correct information, a technical knowledge elicitation was conducted with a member of the development team. Essentially, this method was similar to that of a contextual inquiry (CI); however, as described in IEC:62366-2 Section 8.4.2, a classical contextual inquiry is a method to learn about prospective users, their tasks, and the environment in which the medical device is used. Ideally, this would have been conducted with an existing expert user, but as this device is novel no such user currently exists. Also, the team member who participated in this analysis is a domain expert in extracorporeal systems and at the time of this analysis had not physically interacted with the device. However, we understand that this may have unintentionally introduced bias into the analysis as they are not an intended user. Therefore, our technical knowledge elicitation facilitated a discussion between the developers and the HFE specialists on design ideas and suggestions for improvement as the user conducted a think-aloud of their actions. It also allowed us to review whether there were any steps performed by the development team that were not captured in the IFU. Although this analysis was initially conducted to determine how accurate and comprehensible the IFUs were, the HFE specialists were able to identify some issues with the existing system processes that could be streamlined to improve the setup and usability of the device. Specifically, we identified that the setup of the device was cumbersome and the GUI lacked user feedback and clear indicators of the device's current state. This led to major changes to the setup of the device, which were added to the corresponding IFU, and to the description of the interaction with the GUI.

In summary, based on the results from the above three analyses, significant changes were made to the IFUs. The initial three IFUs were re-organised to make it easier for participants to follow. One document was redesigned to focus on key background information for the entire system, as previous versions of the IFUs had background information included in the introduction for each task so users could understand the relationship between the biology and the task. Another focussed on the setup and priming of the system, as these steps followed on from each other. The third IFU underwent relatively few changes, focussing on using the device at the patient bedside, but no longer including the background information. Changes to this IFU improved workflow and removed unnecessary and potentially confusing information throughout the instructions prior to beginning user studies.

3.2.4. User study

The user study consisted of four parts: training, using the system, a follow-up survey, and a semi-structured interview. Participants were provided with basic training on the device by a member of the HepatiCan™ development team, immediately prior to the study observation. This consisted of a general overview of the machine, covering necessary background information about how the system works, an overview of the components and their purpose, and a basic introduction to the GUI. Thus, the training provided an understanding of what the device was used for, the biology behind how it worked, and a general overview of the device itself. In practice, to use a novel device such as HepatiCan™ users would be required users to undertake formal training which

would include a more detailed review of the background information accompanied by a demonstration of the device.

The user study we ran differs from that outlined in IEC:62366-2 Section 16.2.4, in that rather than just evaluating users performing set tasks, we also included a survey and interviews to better understand users' perceptions and experiences when using the device. Within IEC:62366-2, surveys are suggested as a means to gather information quickly. However, it does not recommend any particular surveys for evaluating medical devices, so it is up to the HFE specialists to develop or identify surveys to gather information. Based on this we used the system usability scale (SUS) questionnaire (Brooke, 1996) at the end of their interaction with the device to measure participants' satisfaction followed by a semi-structured interview to understand the participant's engagement and thoughts on the device as a whole. As this was the first usability study conducted on the device with intended users, we wanted to not only ensure that users were able to use the device properly, but to review whether the current version was easy to use and identify areas where we could improve user satisfaction.

As HepatiCan™ is a medical device, to properly evaluate the system further it needs to be connected to a patient, which will take place in a clinical trial setting. However, to achieve that, regulatory approval is required to demonstrate that the device meets certain standards, including IEC:62366. Therefore, for this analysis the device is still in a prototype phase and being connected to a real patient is not yet possible. To overcome this, a patient was simulated as a bottle of fluids, which was connected to an apheresis machine ready to be attached to HepatiCan™.

Following the training, participants were asked to perform eight tasks as prescribed in the updated third IFU, which prescribes the tasks related to using the device. The eight tasks chosen related to the fundamental running of the device, and included tasks that end users would be required to conduct when using HepatiCan™ to treat patients. Each participant was asked to think aloud while following the refined IFU for the "device operation" stage of the overall set-up task. Think-aloud requires users to describe what they are thinking when performing tasks, and allows the observer/s to gain insight into how participants are making their decisions and highlight mismatches that occur.

The reason for focusing on the IFU relating to using the device, rather than all of the IFUs, was that most intended users would be interacting with the machine more often than setting up, so it was important that this was prioritised during the prototype stage.

3.3. Analysis

Following the recommendations of IEC:62366-2, the data from the hierarchal task analysis, cognitive walkthrough, and user study were analysed using the PCA method. Additionally, a thematic analysis was conducted on the transcripts of the semi-structured interviews, in line with Braun and Clarke's thematic analysis technique (Braun & Clarke, 2006). This involved coding the transcripts; by grouping the codes we were able to identify themes. On completion, these themes, used in conjunction with the PCA results, identified the effectiveness, efficiency, safety, and user satisfaction with the device.

4. Findings

During the user study, we further explored the usability of the device in relation to its effectiveness, safety, efficiency, and user satisfaction. As the effectiveness and safety of the device were both analysed using the PCA method, these findings have been combined. Similarly, efficiency and satisfaction are also grouped together as the findings from these are closely aligned. The errors described below are those that were perceived as being of higher severity.

4.1. Effectiveness and safety

4.1.1. Perception errors

As HepatiCan™ has two circuits, it is important that the interface between the apheresis machine and the device is correctly set up. The interface consists of two tubes coming from the device, an inflow and an outflow, that need to be connected via its inflow and outflow connectors, with the outflow connected to the device inflow and the inflow connected to the device outflow. In the IFU, a lack of definition of markings on the inflow and outflow to and from the device confused the users. This led to them struggling to properly connect the device to the simulated patient and in some cases it was connected back to front. To rectify this, we recommended providing clearer detail to the IFU to assist users in identifying which tube was which.

Another perception error concerned changing filters on the device. The device has two filters in parallel so that they can be switched mid-treatment, due to the length of treatment and the filters having a finite capacity. To switch the filters, participants are required to clamp off the primary filter and then switch the patient flow to the secondary filter. However, participants found it difficult to correctly identify the components they would be working on, with one clamp, in particular, being difficult to locate as it was obscured by other components. To address this perception difficulty, we recommended additional information and images be added to the IFU, along with labels on the pumps and filters.

4.1.2. Cognition errors

One of the main features of the medical device is a large container filled with the ATMP biomass to treat the patient. The container has several inlets and outlets that can be used, depending on the function it is performing. All the participants were confused at times as to which clamps to clamp and unclamp to perform the required function, which could lead to a build-up of pressure in the machine. We recommended adding coloured arrows across the top of the container where the clamps are connected to indicate the correct flow of fluids.

After being modified based on the results of the previous analyses, the IFU still contained long paragraphs of text to explain to users what they were doing. This led to participants losing their place in the text or accidentally missing a step and then needing to re-read the instructions to complete the task. Additionally, some of the tasks included supplementary tasks that could be completed based on the scenario; sometimes, participants would begin a supplementary task, realise it was not required, and have to reverse steps before continuing with the actual task. To improve this, we recommended providing less detail between steps and focusing on the actions users are required to perform.

To complete certain tasks participants needed to interact with the GUI. However, participants found that this section provided insufficient direction about task order, leading to frustration. We recommended redesigning some aspects of the GUI to improve the task order and updating the IFU to capture the changes.

4.1.3. Action errors

The tasks relating to connecting the device to the simulated patient, effectively a glass bottle of plasma or whole blood, with an entry and exit port connected to a double headed peristaltic pump, described as 'patient bottle', became a point of frustration for the participants. To complete this task, clamps needed to be changed from open to closed, and vice versa. However, if this task is not completed in the correct sequence a level difference and/or pressure difference may occur outside the device with the potential for liquids to leak. In practice, when connected to an apheresis machine fluid flows would be maintained correctly between the patient and HepatiCan™. To mitigate this, the sequence of steps was broken down to better highlight the order clamps positions should be altered, with accompanying images to improve sensemaking.

Additionally, when there was pressure build-up in the device, the system alarms. However, the IFU at that time did not have a trou-

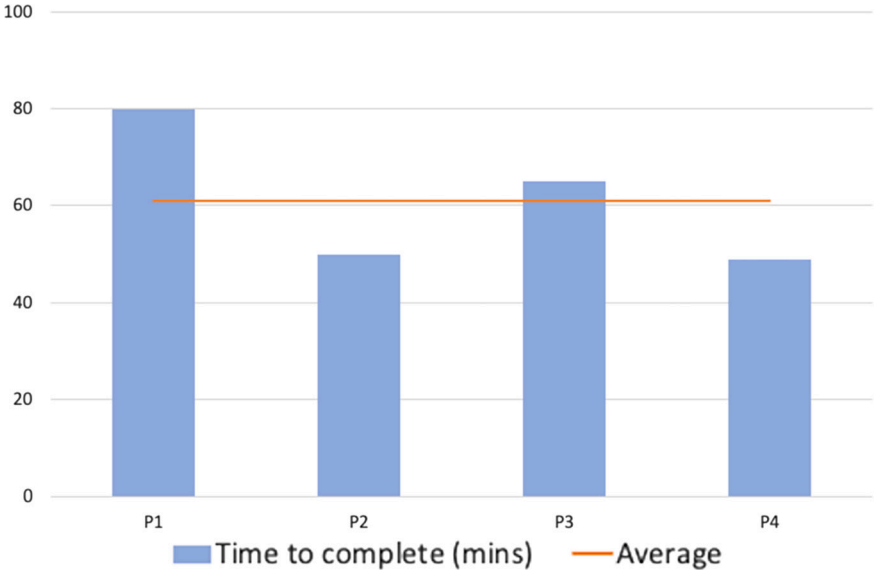

Fig. 5. Time to complete eight tasks during user study.

bleshooting section for participants to quickly identify what fault was occurring and how to resolve the issue. In light of this, we recommend that a troubleshooting guide be written to provide resolutions to all alarms and potential errors.

Currently, the device is using off-the-shelf pumps as they have already been approved as biotechnology-grade equipment. However, participants found using one particular component a point of major frustration. The front facia of the pumps has icons that look like buttons but they are just indicators, with the buttons being on the top of the device. To improve the usability of the pumps we recommended including clearly labelled images of the pumps in the IFU alongside instructions taken directly from the pump's IFU.

4.2. Efficiency and satisfaction

Efficiency was measured by timing how long it took each participant to complete the eight assigned tasks (e.g., changing filters, fluidisation of the biomass, reading pressure differentials, sampling, etc. On average it took each participant just over 60 minutes to complete the eight tasks, see Fig. 5, with P1 taking significantly longer to complete them. For comparison, an expert user (a member of the research team) can complete these same tasks in approximately 30 minutes as they no longer need to refer to the IFUs. However, when reviewing the data, it became apparent that P1 spent more time reading the instructions prior to performing tasks. This corroborated other findings that some tasks require streamlining to improve understanding and reduce the information in the IFU that experts could be assumed to be familiar with.

Upon reflection, it became apparent that the additional information that was added to the IFUs after performing the HTA, CW, and TKE may not have all been necessary. This came down to the HFE specialists not having enough domain knowledge prior to conducting their evaluations and not having conducted user research themselves to gather insight into what knowledge they could expect from intended users, but it should also not be expected that HFE specialists be domain experts in everything. This was reinforced during the semi-structured interviews with participants suggesting that the IFU included too much information with P3 stating "*it was a heavy read, it's very wordy. . . I sort of had to stop reading it because it was just too much!*" However, P1 acknowledged that this was necessary due to the complexity of the machine stating "*It's quite a tongue twister I think sometimes to read with all the exact terminology, but I can understand if you're training someone, you have to have that.*" Therefore, the complexity was seen as being necessary. However, after assessing all of the information from the user study it became apparent that the IFU still needed refinement, to allow a balance between what information is important to the user at the time they are required to perform actions and providing clear steps on how to enact those actions. This would reduce the likelihood of use errors occurring from mismatching information from the IFU.

Table 1
Participants' SUS scores.

Participant ID	SUS score
P1	70
P2	72.5
P3	82.5
P4	77.5

The SUS is a questionnaire used to measure usability, which consists of 10 questions with five response options (strongly agree to strongly disagree) and results in a percentile ranking from 0-100, with a percentile of over 68 indicating that the device is above average, and anything below this requiring further design iterations. The results from the SUS can be seen in Table 1 which showed that in general participants were satisfied with the device and should be willing to use it again in the future as all values were higher than 68 (Brooke, 1996).

As all of the users were experts with existing extracorporeal devices, through the semi-structured interviews they were able to provide insights into HepatiCan™, particularly in comparison to the functionality of other extracorporeal devices. Some participants stated that it was a significant improvement when compared to similar medical devices that they currently operate, with P2 stating "*This is a vast improvement on what I'm used to.*"

5. Discussion

Throughout this study we used the regulatory standard IEC:62366 as a guide to evaluate the usability of a novel liver support system, HepatiCan™. We applied multiple methods, as prescribed by the standard, to measure the effectiveness, efficiency, safety, and user satisfaction of the device. This study also generated feedback for the HepatiCan™ development team to improve the system usability, thus reducing the likelihood of use errors occurring in later development stages.

However, through reflection on the study, we determined that the standard only provided generalised information and required HFE specialists to develop their own plan to properly evaluate medical devices, which could lead to certain aspects of medical devices not being prop-

erly evaluated if the HFE specialists were unfamiliar with methods that may yield better insights. Additionally, we established that more attention should be placed on the development of IFUs so that they describe 'normal use' in a way that aligns with real practice and evaluate them in an ecologically valid environment.

Prior to conducting these analyses, the research team were familiar with the regulatory standards for medical devices; however, they had no experience in adopting them in practice. Through this study we have identified key areas that new practitioners need to be aware of prior to commencing regulatory usability assessment, specifically those within academia

- Knowing what multidisciplinary research is required at the initial stages of development; within academia this is knowing who to add to grant proposals to ensure adequate funds will be available.
- The current regulatory standards provide insufficient information on which method will provide the best output to inform effective design decisions.
- The IFUs determine *"normal use"* - Although this is well known for HFE specialists familiar with regulatory standards, this concept differs from the colloquial understanding of *"normal use"*.

5.1. Multidisciplinary research

This study highlights the importance of identifying key research fields that can support the transition from clinical viability to product development, especially within the heavily regulated medical domain. Within academia, this requires research groups to identify required specialities during the grant writing phase to ensure that adequate funding will be available. To achieve this, universities need to foster multidisciplinary collaboration so that these groups can be more easily identifiable as it is unfeasible that a single department has all of the required domain knowledge to successfully develop a medical device.

Additionally, when carrying out analyses it is important that all members of the research team are included and not rely solely on HFE specialists to assess usability. This was highlighted by the user study, as through having clinical experts evaluate the device it became evident that the additional information that the HFE specialist deemed necessary was not required and was distracting. Instead, the clinical experts expected more detail in relation to specific connectors and other components within the physical space to ensure correct operation, drawing on clinicians' existing mental models of extracorporeal circuits.

In summary, without early multidisciplinary collaboration research groups will be required to understand and conduct usability evaluations without expert knowledge in human factors engineering. Relying solely on passing knowledge, that is information obtained from the internet or word of mouth, key methods may be skipped or ineffective or insufficient analyses may be carried out leading to inadequate design changes to reduce the likelihood of use errors occurring. The separation of expertise in HFE, device design, and clinical work is necessary but does not always enable cross-disciplinary solutions to be developed. Additionally, solutions that could be developed in only a few iterations may take multiple iterations. Worst case scenario, without adequate HFE training potential use errors may never be identified prior to the device being launched resulting in unexpected costs and potential harm to patients or users. Therefore, there needs to be better support for people to gain the necessary interdisciplinary expertise to conduct evaluation analyses of novel medical devices without relying solely on an "apprenticeship" model that assumes existing sophisticated interdisciplinary expertise. This could include seeking advice from HFE professionals with existing expertise in interpreting and applying regulatory standards or turning to other sources of training beyond the standards.

5.2. Analysis methods

Prior to conducting the user study, three methods (HTA, CW, TKE), were used by the research team to evaluate the IFUs, following the instructions in IEC:62366 as closely as possible. Although the standards are thorough, at times they are difficult to navigate and we did not find the guidance as to which usability method is best at which time clear. For example, the information provided about time and motion analysis, which is a method to identify tasks that facilitate versus hinder users completing tasks, doesn't explain why this method may be more suitable than a usability study in some circumstances. Additionally, while some of the methods are appropriately tied to specific periods of evaluation, others are just added in Annex E of IEC:62366-2 with a look-up table for when these methods may be best incorporated. However, the standard does state that technical report IEC:62366-2 is not intended to be a sole source of information, and that additional resources should be used when in doubt, putting more emphasis on ensuring that the HFE specialists are familiar with multiple methods of evaluation and when they are most appropriate.

One of the main struggles when performing the HTA for HepatiCan™ is that as a novel medical device it is difficult to know whether the flow of tasks is correct or whether unforeseen tasks are still required, as until it can be evaluated in its intended use environment, some issues may be unidentifiable. Moreover, as no one on the team was from a clinical background as HepatiCan™ evolved from a biomedical study, we relied heavily on engagement with healthcare professionals, which is why we ensured that intended users were recruited for the user study. Additionally, as the device is novel, assumptions were being made as to how this device may work, with the design being based on existing extracorporeal systems. This too may result in tasks unintentionally being missed or given at an inappropriate time at this stage of prototype development.

Within the standards, CW was listed as a method to be conducted during formative evaluation at various stages of development as it is primarily focused on obtaining feedback on the device's user interface. In our analysis, we used the CW to evaluate the IFU to verify that it contained all the relevant information for a potential user to complete their tasks and to ensure that we were developing a solid framework for 'normal use'. One of the main limitations we had with this method was that neither of the HFE specialists that conducted the CWs had a clinical background. IEC:62366-2 highlights that not only should those conducting usability engineering have appropriate training in that area, but also have appropriate medical device domain knowledge. This led to additional information being added to the IFUs that was later found unnecessary as the intended users already had sufficient understanding of some of the added areas. This led to the IFUs becoming cluttered and more difficult for potential users to understand, see Fig. 6 for an example of the modifications to one of the IFUs over time. This highlights the importance of conducting user research prior to other analyses so that HFE specialists can understand the intended user groups' work practices and prior knowledge even at the early stages of product development.

Additionally, CW is a method designed for "walk up and use" devices; since HepatiCan™ is not such a device, and would only be used after suitable training, it is questionable whether this was an appropriate method to use. However, getting access to training or upkeep of training can be difficult. And, even after receiving training, staff may still not feel confident in using the device, essentially changing what should be an "only use if trained" device into a "walk up and use" device. This is also reflected in the second example provided in the introduction of this paper of the situated observational study, which highlighted that in clinical practice there is typically a limited time for formal training. If user research is conducted early in the development of medical devices to identify intended users and to understand what domain knowledge they have, training can be tailored to their needs. Thus, training should be used to cover the users' gaps in knowledge that the novel device would exhibit, and the IFUs should provide a complete

Note: pressure transducers can only be added in correct orientation with respect to tubing but should also be placed in correct orientation of PT holders where available. Their position with respect of other components of the Giving Set is also important (e.g. T connectors, sampling ports…)

Attach a pressure transducer (PT2) to luer next to position 24.

Connect a PT (PT3) between positions 34 and 35.

Connect a PT (PT4) between positions 44 and 45.

Connect a PT (PT5) between the 60mm line after the T connector after position 44 and the 60mm line before the Y connector before position 56.

Connect a PT (PT6) between positions 61 and 62.

(a)

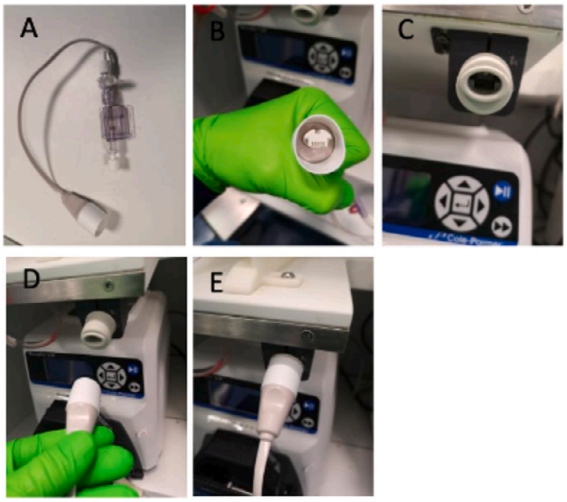

Figure 31. A: Pressure transducer. B: PT1 cable connection. C: Workstation PT1 connection. D: orientation of the PT1 cable for connection. E: PT connected.

(b)

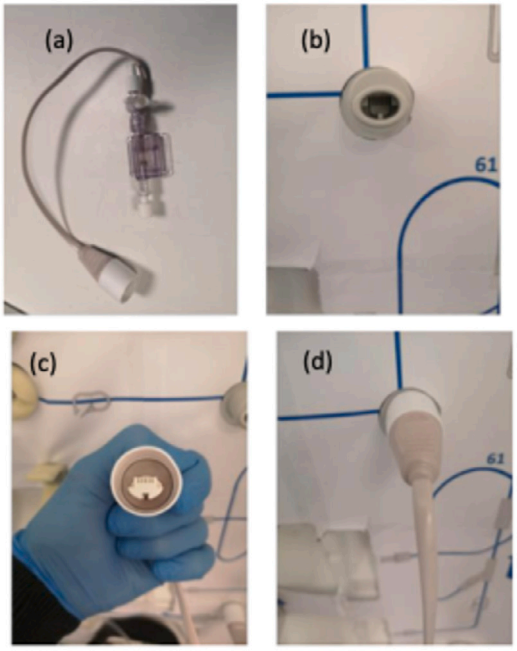

Figure 22. (a) Pressure transducer, (b) Workstation PT connection, (c) PT cable connection highlighting orientation of the connector, and (d) PT connected. *Note: PT connection to the Workstation for PT2 to PT6. PT1 is mounted in the opposite orientation to that in this protocol.*

(c)

Fig. 6. These images are all various versions of the instructions for use (IFU) used to inform the user how to connect a pressure transducer with (a) the original IFU, (b) the version after the HTA, CW, TKE (repeated for each transducer), and (c) the version after the usability study.

understanding of the device, including its intended use, operating procedures, troubleshooting guides, etc., to provide the intended users with all the necessary information to use and operate the device safely and effectively.

As stated previously, during this study we used the Technical Knowledge Elicitation (TKE) method, which is not described in the regulatory standards. When developing novel technologies, it is easy for those who develop the technology to inadvertently miss steps when developing IFUs. This is caused by their familiarity with the system and from the potential multiple iterations of the device, which can cause some tasks to not be amended in the documentation. To identify these types of unintentionally missed or redundant steps, we propose that a TKE be run with a working prototype with the developers of the system prior to user studies to identify how they envisage the system being used and to highlight any undocumented tasks to be added to IFUs. This allows the developers to step back from the medical device and be led by HFE specialists to re-visit the IFUs to ensure their completeness. In our study, we focused on how the development team used the IFU, which highlighted instances where the team would conduct tasks not in the IFU or in a different order from the IFU as they are familiar with the technology.

In summary, whilst all of these methods had value in improving the IFUs prior to conducting the user study, they also had limitations. For ease, we have summarised the strengths and limitations of each method in this context, see Table 2. Ideally, these analyses should be conducted by an interdisciplinary team and not rely solely on HFE specialists. This would allow each domain expert to identify potential use errors or points of interest that require further exploration to improve

Table 2
Strengths and Limitations of methods used in the analyses of HepatiCan™.

	Strengths	Limitations
Hierarchical Task Analysis (HTA)	Defines all of the necessary tasks to achieve set goals.	For medical devices still in the prototype phase their task analysis may still be in flux and require re-evaluation of certain goals.
	Provides the team with information so they can review their entire workflow and re-order tasks if necessary.	May be difficult to highlight tasks that are missing without having a strong clinical background.
	Can highlight tasks that frustrate users in achieving their goal.	
Cognitive Walkthrough (CW)	Quick to run and does not require recruitment.	Potential to add in information that inhibits normal practice due to a lack of clinical background.
	Helps to identify potential instances of mismatches and vague instructions.	Difficult for HFE experts to put themselves in the shoes of the intended users.
Technical Knowledge Elicitation (TKE)	Quick to perform and can provide opportunities for clarification of context in the moment.	Difficult to conduct within the constraints of healthcare.
	Promotes that the user is the expert and the researcher is trying to understand how they perform their tasks.	For novel medical devices, there are no established expert users with clinical practices still being ratified.
User Study	Provides unique insight into the efficiency, effectiveness, safety, and user satisfaction of devices.	Take a long time to design and recruit participants that are representative of intended users.
	Allows for engagement with representative users and the opportunity to explore concepts together.	Difficult to engage participants in user studies outside their work hours.
	Early identification of issues.	

the usability of the device, especially when developing novel medical devices.

5.3. Designing for context

In design, there is an onus on developers to understand clinical (or other situated) practices so that 'normal use' as defined through IFUs aligns as well as possible with clinical practice. Many novel medical devices introduce new possibilities, and therefore will – by design – disrupt established clinical practice with a view to improving that practice (making it safer, more efficient, etc.) and/or improving patient safety and patient outcomes.

However, by creating IFUs that accommodate all of the nuances that arise due to the different contexts of use, there is potential that they will become cumbersome to the point that they are no longer helpful. Instead, it is important that HFE specialists have a strong understanding of the various contexts of use. Therefore, when developing IFUs, HFEs need to design them so that a user within any anticipated scenario can adapt the existing instructions to meet their individual circumstances.

Through working with clinical experts, especially during the early phases of medical device development, a better understanding of clinical practice can be identified. This can be achieved through the development of personas and scenarios, which are standard HFE tools to support this and are included as methods within the regulatory standards, regardless of whether or not there is an existing device or workflow. In this study, the HFE specialists were brought in at the beginning of the design and formative evaluation phase, and the intended users and scenarios were developed by the development team. However, in retrospect, the HFE specialists should have conducted contextual inquiries with experts on similar extracorporeal devices to gain an understanding of their domain knowledge. Without this understanding, the HFE specialists modified the IFUs, which establishes 'normal use', to be overly complex.

When evaluating the ecological validity of this study, compromises needed to be made. While the three "expert analyses" (HTA, CW, TKE) helped with articulating the design of the software and hardware, and assumptions about other aspects of use, they had low ecological validity, serving more to support reflection on the design than to evaluate it in an ecologically valid way. These analyses helped shape the design of the user study, which was strong on some aspects of ecological validity but not all, which is contrasted with the seven elements established by Van Berkel et al. in Table 3. However, for these formative usability analyses we believe that the ecological validity was adequate, but in some elements, discussed in Table 3, there is room for significant improvement especially as we move towards summative evaluation closer to deployment into clinical practice.

Through conducting an ecological validity assessment, we were able to critically determine how our user study performed in relation to the seven dimensions, which in turn allowed us to understand how close our user study was to clinical practice. In our user study, we observed that the patient bottle was a major constraint that caused the device to react abnormally as the pumps used to simulate delivery from and return to the patient bottle did not remain synchronised with the pumps within HepatiCan™. This would not be the case when attached to a patient via an apheresis machine. Additionally, with the software of the device still a work in progress some features were not yet properly implemented. The ecological validity assessment allowed us to identify which factors require further attention, and which factors were potentially causing use errors that would most likely not occur when used in practice. In hindsight, if this assessment had taken place prior to conducting the analyses some of the issues that arose in relation to the user study design may have been mitigated, producing a closer representation of clinical practice.

However, when assessing ecological validity as a whole, the research by van Berkel et al. (2020) did not distinguish clearly between training materials and instructions for use. The analyses reported here highlighted the need to also consider the instructions for use, alongside software and hardware, as a key component of the device. This is essential to evaluate the device against 'normal use' as defined in IEC:62366 and other regulatory documents. Doing so allows for potential use errors to be identified during the early phases of development so they can be properly mitigated during product development rather than identifying them only when used in practice.

To improve on this, we provide three lines of thought for those who are unfamiliar with the regulatory standards: (1) ensuring that tools such as personas and scenarios, regardless of whether there are existing devices or workflows that can be mimicked, are incorporated early on during the development stage. (2) During the development of the IFUs the research team should use ecological validity as a guide, considering questions such as: Who are the intended users? In what context will this device be used? What are the tasks that need to be performed? What training do users require to be able to perform these tasks? What prior knowledge do they have that can simplify the IFUs? etc. Essentially, understanding that in relation to the regulatory standards "normal use" is defined by the IFUs. Lastly, (3) Running user studies that evaluate both the IFUs and the medical devices in a manner that maximises ecological validity. This is dependent on the current state of the medical device and the extent to which the user tasks and contexts of use are already defined. In relation to medical devices similar to HepatiCan™,

Table 3
Analyses contrasted with the seven elements of ecological validity.

User roles	The four participants in the user study were representative of the target user groups in that they all had experience with extracorporeal circuits and/or kidney dialysis machines – the most similar type of procedure – but as this liver support procedure had never been performed before there was further understanding required by the users. To improve future analyses more participants would be recruited to capture a wider understanding of user interactions.
Environment	The environment in which the analyses were undertaken was a lab clean room with limited space, where in actuality the procedure of using the device would take place at the patient's bedside and most likely in a specialised ward, e.g., intensive care unit or high dependency unit. For future studies we would endeavour to use mock intensive care wards rather than clean laboratory areas.
Training	The training was somewhat basic, but a good basis for future training once deployed. It allowed us to understand the minimum training required to orientate users to the device, where detail could be removed due to users' prior knowledge, and areas that needed further explanation as they are specific to the device.
Scenario	The scenario in the user study was to understand how users new to the device used and understood the IFU and to identify use errors. However, as this involved a simulated patient, the tasks conducted were rigid with a set start and end position with no context. In reality, these tasks would be performed based on specific events occurring which could not be replicated in this user study. In future studies we would create specific scenarios in which we can more fully test the device: i.e., pressure build-ups, pump failures, drop in oxygenation levels, etc.
Patient involvement	The simulated patient was adequate from a scientific (medical) perspective, but completely unrealistic from all other perspectives; this is an inevitable consequence of testing a novel medical device prior to clinical trial. However, once safety has been established adequately, then it will be important to involve patients in a stage 3 trial.
Software	The software that was used during the user study was still in early development and not ready for commercial release. However, this provided opportunities for the HFE specialists to provide suggestions to the development team to improve usability for future iterations. Prior to future studies a more robust software version will be developed that is indicative of a commercial release.
Hardware	This was the most mature part of the device and human factors needs had been considered during the design process by the research team. This meant that the physical interaction with the device was a good representation of how it is intended.

that is a novel device that will require a completely new process, this may require exploring similar devices and discussion with clinical experts within the relevant fields.

One of the limitations of this research was the ability to recruit suitable participants, as this study was conducted during the COVID-19. However, regulatory standard IEC:62366-2 in Annex K provides guidance on the sample size, or number of participants, required to improve the probability of detecting use errors. In this study with only four participants, there is potential that some use errors relating to the scenarios evaluated were not identified. In future formative evaluations, we intend to increase the number of participants by working more closely with clinical teams and ensuring planning and budgeting for this from the outset of the project.

6. Conclusion

In this paper, we used the regulatory standard IEC:62366 to evaluate the usability of a novel medical device and highlight lessons learnt for academics and new practitioners. Within academia the translation of clinically viable processes to novel medical devices is becoming more common, requiring closer attention to creating multidisciplinary teams when seeking funding via grant applications. Additionally, the standards provide suggestions but do not provide a clear path to determine the usability of a device. Our analyses found that instructions for use need to be properly assessed as they define 'normal use' for regulatory assessment. Therefore, it is essential that instructions for use need to be considered as part of the overall system that is evaluated, alongside all other accompanying documents for medical devices. Additionally, we address the methods used during this analysis and their strengths and limitations when assessing novel technology and advocate that these analyses be conducted by an interdisciplinary team and not solely by HFE specialists, to create interdisciplinary experts. This will allow for issues outside the domain of individuals to be discovered and improved by the team. Doing so will improve safety and aid in identifying potential use errors during the development stage.

Ethical declaration statement

The research conduct in the article entitled 'An academician's approach to the application of human factors standards: a case study on a liver support system' was done so with ethical clearance as the studies were conducted with human subjects. To comply with the code of ethics, the following sentence is included in the body of this work in Section 3.1 when participants are discussed:

'These analyses were undertaken with ethical clearance and informed consent was obtained from all participants (UCLIC/1819/006/BlandfordProgrammeEthics).'

Funding

This research was funded in whole, or in part, by the Wellcome Trust [203145/Z/16/Z]; EPSRC [Grant number NS/A000050/1]; UCL-HEIF [KEI2020-05-03] award 156780; NIHR [II-LA-0417-20002], and The Liver Group Charity [571348]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jeremy Opie reports financial support was provided by Wellcome Trust. Jeremy Opie reports financial support was provided by Engineering and Physical Sciences Research Council. Ann Blandford reports financial support was provided by Wellcome Trust. Ann Blandford reports financial support was provided by Engineering and Physical Sciences Research Council. Clare Selden reports financial support was provided by National Institute for Health and Care Research.

Acknowledgements

We would like to acknowledge the support of the research team that developed HepatiCan™ and the participants in this study. This work is supported by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) (203145/Z/16/Z).

References

62366-1:2015, I. (2015). *Medical devices part 1: Application of usability engineering to medical devices*. Technical Report, Geneva: International Electrotechnical Commission.

62366-2:2016, I. (2016). *Medical devices part 2: Guidance on the application of usability engineering to medical devices*. Technical Report, Geneva: International Electrotechnical Commission.

9241-11:1998, I. (1998). *Ergonomic requirements for office work with visual display terminals (VDTs) - part 11 guidance on usability*. Technical Report, Switzerland: ISO.

AAMI-FDA (2010). *Infusing patients safely: Priority issues from the AAMI/FDA infusion device summit*. Technical Report.

Blandford, A., Keith, S., Butterworth, R., Fields, B., & Furniss, D. (2007). Disrupting digital library development with scenario informed design. *Interacting with Computers*, *19*, 70–82. https://doi.org/10.1016/j.intcom.2006.07.003.

Borsci, S., Uchegbu, I., Buckle, P., Ni, Z., Walne, S., & Hanna, G. B. (2018). Designing medical technology for resilience: Integrating health economics and human factors approaches. *Expert Review of Medical Devices*, *15*, 15–26. https://doi.org/10.1080/17434440.2018.1418661.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77–101. https://doi.org/10.1191/1478088706qp063oa. Retrieved from http://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa.

Brooke, J. (1996). SUS: A 'quick and dirty' usability scale. In *Usability evaluation in industry* (p. 6). CRC Press.

Burlington, D. B. (1996). Human factors and the FDA's goals: Improved medical device design. *Biomedical Instrumentation & Technology*, *30*, 107–109.

Campbell, G. (2007). Statistics in the world of medical devices: The contrast with pharmaceuticals. *Journal of Biopharmaceutical Statistics*, *18*, 4–19. https://doi.org/10.1080/10543400701668225.

Catchpole, K., & Alfred, M. (2018). Industrial conceptualization of health care versus the naturalistic decision-making paradigm: Work as imagined versus work as done. *Journal of Cognitive Engineering and Decision Making*, *12*, 222–226. https://doi.org/10.1177/1555343418774661.

Escalada-Hernández, P., Soto Ruiz, N., & San Martín-Rodríguez, L. (2019). Design and evaluation of a prototype of augmented reality applied to medical devices. *International Journal of Medical Informatics*, *128*, 87–92. https://doi.org/10.1016/j.ijmedinf.2019.05.004. Retrieved from https://www.sciencedirect.com/science/article/pii/S1386505618312954.

Fairbanks, R. J., & Caplan, S. (2004). Poor interface design and lack of usability testing facilitate medical error. *The Joint Commission Journal on Quality and Safety*, *30*, 579–584. https://doi.org/10.1016/S1549-3741(04)30068-7. Retrieved from https://www.sciencedirect.com/science/article/pii/S1549374104300687.

Food and Drug Administration (2016). *Applying human factors and usability engineering to medical devices*. Technical Report FDA-2011-D-0469, Rockville: U.S. Department of Health and Human Services Food and Drug Administration. Retrieved from https://www.fda.gov/media/80481/download.

Gelijns, A. C., Medicine, I.o.M.U.C.o.T.I.i. (1989). 3. The development of medical devices. In *Technological innovation: Comparing development of drugs, devices, and procedures in medicine*. US: National Academies Press. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK222708/.

MHRA (2014). *Medical devices - adverse incidents reported to MHRA 2011 to 2013*. Technical Report.

Pelayo, S., Marcilly, R., & Bellandi, T. (2021). Human factors engineering for medical devices: European regulation and current issues. *International Journal for Quality in Health Care*, *33*, 31–36. https://doi.org/10.1093/intqhc/mzaa103.

Petrie, H., & Bevan, N. (2009). The evaluation of accessibility, usability and user experience. In *The universal access handbook 1* (pp. 1–16).

Selden, C., Bundy, J., Erro, E., Puschmann, E., Miller, M., Kahn, D., Hodgson, H., Fuller, B., Gonzalez-Molina, J., Le Lay, A., Gibbons, S., Chalmers, S., Modi, S., Thomas, A., Kilbride, P., Isaacs, A., Ginsburg, R., Ilsley, H., Thomson, D., … Spearman, C. W. (2017). A clinical-scale BioArtificial liver, developed for GMP, improved clinical parameters of liver function in porcine liver failure. *Scientific Reports*, *7*(1), Article 14518. https://doi.org/10.1038/s41598-017-15021-4. Retrieved from https://www.nature.com/articles/s41598-017-15021-4.

Selden, C., & Fuller, B. (2018). Role of bioreactor technology in tissue engineering for clinical use and therapeutic target design. *Bioengineering*, *5*(2), 32. https://doi.org/10.3390/bioengineering5020032. Retrieved from https://www.mdpi.com/2306-5354/5/2/32.

Selden, C., Spearman, C. W., Kahn, D., Miller, M., Figaji, A., Erro, E., Bundy, J., Massie, I., Chalmers, S. A., Arendse, H., Gautier, A., Sharratt, P., Fuller, B., & Hodgson, H. (2013). Evaluation of encapsulated liver cell spheroids in a fluidised-bed bioartificial liver for treatment of ischaemic acute liver failure in pigs in a translational setting. *PLoS ONE*, *8*, Article e82312. https://doi.org/10.1371/journal.pone.0082312. Retrieved from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0082312.

Spencer, R. (2000). The streamlined cognitive walkthrough method, working around social constraints encountered in a software development company. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 353–359). The Hague The Netherlands: ACM. Retrieved from https://dl.acm.org/doi/10.1145/332040.332456.

Tase, A., Ni, M. Z., Buckle, P. W., & Hanna, G. B. (2022). Current status of medical device malfunction reporting: Using end user experience to identify current problems. *BMJ Open Quality*, *11*, Article e001849. https://doi.org/10.1136/bmjoq-2022-001849. Retrieved from https://qir.bmj.com/lookup/doi/10.1136/bmjoq-2022-001849.

Thimbleby, H., Lewis, A., & Williams, J. (2015). Making healthcare safer by understanding, designing and buying better IT. *Clinical Medicine*, *15*, 258–262. https://doi.org/10.7861/clinmedicine.15-3-258. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4953110/.

van Berkel, N., Clarkson, M. J., Xiao, G., Dursun, E., Allam, M., Davidson, B. R., & Blandford, A. (2020). Dimensions of ecological validity for usability evaluations in clinical settings. *Journal of Biomedical Informatics*, *110*, Article 103553, Article 10.1016/j.jbi.2020.103553. Retrieved from http://www.sciencedirect.com/science/article/pii/S1532046420301817.

Vincent, C. J., & Blandford, A. (2015). Usability standards meet scenario-based design: Challenges and opportunities. *Journal of Biomedical Informatics*, *53*, 243–250. https://doi.org/10.1016/j.jbi.2014.11.008. Retrieved from https://www.sciencedirect.com/science/article/pii/S153204641400238X.

Wiklund, M. E. (2022). Reflecting on four decades of progress in applying human factors engineering to medical devices. *Human Factors in Healthcare*, *2*, Article 100024. https://doi.org/10.1016/j.hfh.2022.100024. Retrieved from https://www.sciencedirect.com/science/article/pii/S2772501422000215.

Jeremy Opie is a Research Fellow at the UCL Interaction Centre and a Wellcome/ESPRC Centre for Interventional and Surgical Sciences (WEISS) member. His research explores interactions between people and technology using qualitative methods, with a focus on Digital Health.

Clare Selden is a Professor of Experimental Hepatology, at University College London on the Royal Free Hospital campus. The main thrust of her research has been to utilise liver cells for therapeutic purposes. Her current aim is to deliver a novel cell therapy and medical device to patients with liver failure, in an environment where there are insufficient donor organs available for patients who need a liver transplant.

Ann Blandford is a Professor of Human-Computer Interaction in the Department of Computer Science at UCL, and a member of UCL Interaction Centre. Her funded work is on evaluating complex systems "in the wild", with a focus on Digital Health. She takes a pragmatic approach to developing and applying theory in practice, recognising and working with the inherent "messiness" of the real world. She has been leading projects on HCI/digital health funded by EPSRC and NIHR.