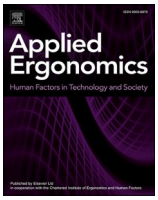
Applied Ergonomics 117 (2024) 104223

Contents lists available at ScienceDirect

Applied Ergonomics

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



Improving maternal safety: Usability and performance assessment of a new medical device for the treatment of postpartum haemorrhage

Sara Candidori [a], Nicolò Dozio [a], Kasra Osouli [b], Serena Graziosi [a,*], Alberto Antonio Zanini [c], Maria Laura Costantino [b], Francesco De Gaetano [b]

[a] Department of Mechanical Engineering, Politecnico di Milano, Via La Masa 1, 20156, Milan, Italy
[b] Department of Chemistry, Materials and Chemical Engineering "G. Natta", Politecnico di Milano, Piazza Leonardo da Vinci 31, 20133, Milan, Italy
[c] Obstetrician Gynecologist, Freelance Professional, Milan, Italy

ARTICLE INFO

Keywords:
User-centered design
Medical device
Uterine balloon tamponade device
Patient safety
Usability
Sustainable development goals
Maternal mortality

ABSTRACT

Postpartum haemorrhage (PPH) is an obstetric emergency causing nearly one-quarter of maternal deaths worldwide, 99% of these in low-resource settings (LRSs). Uterine balloon tamponade (UBT) devices are a non-surgical treatment to stop PPH. In LRSs, low-cost versions of UBT devices are based on the condom balloon tamponade (CBT) technique, but their effectiveness is limited. This paper discusses the experimental study to assess the usability and performance of a medical device, BAMBI, designed as an alternative to current CBT devices. The testing phase involved medical and non-medical personnel and was focused on testing BAMBI's usability and effectiveness compared to a standard CBT solution. We collected measures of the execution time and the procedure outcome. Different training procedures were also compared. Results show a significant preference for the BAMBI device. Besides, medical and non-medical subjects reached comparable outcomes. This aspect is highly relevant in LRSs where the availability of medical personnel could be limited.

1. Introduction

The commercialisation of medical devices is regulated by international standards, such as the EU Medical Device Regulation (MDR) 2017/745 or the Food and Drug Administration (FDA) indications (European Parliament, Council of the European Union, 2017; FDA, 2022). Strict requirements regarding quality and safety must be considered as medical devices interact with the human body. The two standards that have historically been the basis for medical device manufacturers are ISO 13485 and ISO 14971 (ISO, 2016, ISO, 2019), related to quality and risk management. The MDR 2017/745 in Europe represented a fundamental change in this framework, as more stringent conformity assessment procedures are now required. The aim is to increase patient safety compared to previous medical device directives (MDDs) (Ben-Menahem et al., 2020). However, only in recent years, Europe and the FDA have started to focus on safety from a broader perspective, including usability requirements in their evaluations. New standards have emerged for applying usability engineering (UE) or human factors engineering (HFE) to medical devices (IEC, 2015; IEC, 2016; FDA, 2016; AAMI, 2018).

Usability is a qualitative measure of the appropriateness to a purpose of any artifact (Brooke, 1996). It is strictly related to defining the intended use, user, and context. Its main elements are effectiveness, efficiency, and satisfaction field (Brooke, 1996), although engagement, error tolerance, ease of learning, and aesthetic considerations also contribute to it. In medical practice, the growing attention to usability is due to the combined effect of several factors. As healthcare evolves, medical devices are increasingly employed for patient monitoring and treatment. Moreover, they are being used by less skilled users (e.g., the patients) but are also becoming much more complicated (IEC, 2015; Coldewey et al., 2022).

According to the FDA, home care devices are the fastest-growing segment of the medical device industry due to the increased life expectancy and prevalence of chronic conditions (Beer et al., 2014; Tase et al., 2022). As a result, complex devices, such as infusion pumps or ventilators designed for use by trained professionals in clinical settings, are increasingly being used at home (Lyons and Blandford, 2018). Incidents related to the poor design of medical devices, non-intuitive,

* Corresponding author.
E-mail addresses: sara.candidori@polimi.it (S. Candidori), nicolo.dozio@polimi.it (N. Dozio), kasra.osouli@polimi.it (K. Osouli), serena.graziosi@polimi.it (S. Graziosi), albertoantoniozanini@gmail.com (A.A. Zanini), marialaura.costantino@polimi.it (M.L. Costantino), francesco.degaetano@polimi.it (F. De Gaetano).

https://doi.org/10.1016/j.apergo.2023.104223
Received 28 June 2023; Received in revised form 18 December 2023; Accepted 28 December 2023
Available online 14 January 2024

challenging to learn or to use product interfaces have raised the attention to the issue (Lyons and Blandford, 2018). Regarding infusion pumps, regulators (FDA, 2018) and scientific literature (Lyons and Blandford, 2018; Klarich et al., 2022; Tase et al., 2022) highlight how many of the reported adverse events (patients' injuries and deaths) were related to deficiencies in the device development. Inadequate user interfaces include, for example, confusing pump screens, inadequate alarms, unclear warning messages, and confusing and outdated or unavailable user manuals (FDA, 2017; Klarich et al., 2022). Similar usability issues were recently reviewed by Coldewey et al. (2022) for ventilation devices that, as infusion pumps, are increasingly used in multiple contexts as life-saving support. Usability was also evaluated for simpler home care devices, such as blood pressure monitors (Kortum and Peres, 2015; Chaniaud et al., 2020, 2021), pulse oximeters (Gao and Kortum, 2017; Chaniaud et al., 2020, 2021), thermometers (Kortum and Peres, 2015; Gao and Kortum, 2017), and portable ECGs (Bonnette et al., 2017). In addition to home care, usability was also extensively studied for emergency care devices, as in this case, use errors could directly impact patients' survival, especially in time-dependent situations. Numerous studies (Guerlain et al., 2010; Camargo et al., 2013; Robinson et al., 2014; Umasunthar et al., 2015; Edwards et al., 2018; Moss et al., 2018; Kessler et al., 2019) demonstrated, for instance, that the device design is a major determinant for successful adrenaline administration using epinephrine autoinjectors (EAIs, also known as adrenaline autoinjectors, AAIs), the emergency first-line treatment for anaphylactic reactions. Moreover, they demonstrated that EAIs designed using a human-centered approach result in a higher rate of successful and fast treatment in simulated scenarios and are also preferred by several groups of possible users. Similar considerations can be extracted from studies on external defibrillators (Monsieurs et al., 2005; Fairbanks et al., 2007; Hunt et al., 2009; Reeson et al., 2018), both manual (used by healthcare professionals in the hospital) and automated (designed for use also by untrained laypersons). Unintuitive design and lack of usability were indeed identified as barriers to timely intervention, which cannot be sufficiently compensated by additional training. Unlike the general product market, the focus in this context is on the strong relationship between usability and safety (IEC, 2015; IEC, 2016; FDA, 2016).

Medical devices' usability process has strict relationships and interactions with the risk management process. The UE approach for medical devices is based on the use error concept and is conceived to identify and minimise it (IEC, 2015). Since the risk is the combination of the probability of occurrence of harm and the severity of potential harm (ISO, 2019), reducing use errors determines a decrease of the first factor. The term "use error" is used in standards instead of "user error" or "human error" because the main goal is to identify errors that are a direct result of poor user interface design. Indeed, as described by Read and colleagues (Read et al., 2021), using the term "human error" could be damaging since it lacks clarity. Although it could refer to design-induced errors, it might also have a negative or blaming connotation, as it may attribute the cause of the error to the user's negligence. On the contrary, identifying the source of errors in the device interface allows reducing them by increasing the safety of the device. Potential use errors may not emerge in the early stages of the design of a new product; in some cases, they may only become apparent when, for example, the device is used in an emergency or stressful situation (IEC, 2015; Coldewey et al., 2022), in a specific context with spatial and temporal constraints, such as operating rooms (Surma-aho and Katja, 2021) and during simulated-use tests (IEC, 2015; FDA, 2016; Privitera et al., 2017). Therefore, usability testing in the use environment becomes the method for evaluating the user interface of a device.

In the case of medical devices, usability testing should involve the systematic collection of data from test participants to support evidence that the device can be used safely and effectively (FDA, 2016; IEC, 2016). Observational (e.g., performance) and subjective (e.g., comments) data are complementary and extremely useful for assessing the user interface adequacy (IEC, 2016; Valdez et al., 2017). Data can be obtained by observing participants during the test and interviewing them (Valdez et al., 2017); observational data should also include any instances of observed hesitation, apparent confusion, close calls and use difficulties (FDA, 2016; IEC, 2016). As underlined by Carayon and colleagues (Carayon et al., 2015), mixed methods (i.e., methods that combine qualitative and quantitative data) are increasingly used and accepted in UE research in health care, as they allow for a broad and deep assessment of usability issues. Under this approach, the same collection methods, such as interviews or participant observation, could be used to produce both types of data simultaneously, with mixing possible at different stages. Before starting usability testing, it is essential to have a clear, upfront definition of the medical device, its intended use, and clinical indications (Ben-Menahem et al., 2020). Furthermore, the device use process must be transparent. It could be helpful to break it down into a discrete sequence of tasks (FDA, 2016; IEC, 2016). Before testing, user performance that represents success for each task must be specified.

User grouping is another essential aspect to be considered. The ability of a user to operate a device depends on several personal characteristics (Medical devices must be carefully validated, 2018), including physical size, dexterity, coordination, sensory and cognitive skills, mental and emotional state, experience with similar devices, ability to learn and adapt to a new device, and willingness and motivation to learn to use it (FDA, 2016). As it is impossible to separate users according to so many unlimited factors, choices must be made, selecting the most relevant aspects and monitoring the others (FDA, 2016; IEC, 2016). The medical background could be pertinent for grouping potential users when evaluating a medical device used in hospital and home settings. Medical professionals, indeed, could have previous experiences with similar devices, existing habits, or clinical "rituals" that inevitably influence device use and could lead to misuse (Ben-Menahem et al., 2020). One of the issues identified for infusion pumps (Klarich et al., 2022) and ventilators (Coldewey et al., 2022) is the lack of a standardised user interface for similar devices from different manufacturers, which brings cross-device interaction-related usability issues (Coldewey et al., 2022).

The environment of use is another relevant factor (Surma-aho and Katja, 2021), as it might include a variety of conditions that could determine optimal user interface design. Medical devices might be used in clinical or non-clinical environments (Tase et al., 2022). Different location-specific conditions, such as lighting, temperature, noise and activity levels, and personal-specific conditions, such as hygienic requirements (e.g., surgical gloves) and social situations (e.g., stress levels and working in teams), contribute to the usability of the device (IEC, 2015; IEC, 2016; Ben-Menahem et al., 2020).

Usability testing can also be used to test the adequacy of the accompanying documentation and training (IEC, 2016). Instructions for use (IFU) and other accompanying documentation are part of the device. Therefore, specific requirements must also be developed for them (FDA, 2016; ISO, 2016). They must be tailored to the user groups, considering the working environment. Different delivery mechanisms can be employed, selecting the one that is deemed the most effective. Training requirements must be addressed early in the development to drive user interface design. In some cases, instructions can be embedded in the device itself, for example, avoiding that one port can connect to two or three different ports (Feinmann, 2019). The most famous example is using distinct colours for arteriosus (red) and venous (blue) fluid lines of extracorporeal circulation systems. This colour coding is adopted today as a convention by most manufacturers, sometimes coupled with a mechanical impossibility to perform the interchanged connection, increasing the safety and usability of such life-support devices (Rajkomar et al., 2014).

This overview demonstrates that UE is gaining attention in medical device development. However, the availability of usability studies in the scientific literature and documents published by manufacturers is still

limited. This consideration can be extended to other aspects of medical devices, and it is nowadays a debate point in Europe and USA (Fraser et al., 2018; Lenzer, 2018). For about a decade, doctors, consumers and even patients have demanded transparency about the evidence of the safety of medical devices, in analogy to what happens for drugs, especially for European approval procedures (Bowers, 2018; Fraser et al., 2018). Even if safety and usability requirements have gained attention, how the new rules will be interpreted and put into practice (Feinmann, 2019) still needs to be clarified. Short summative reports are publicly available for medical devices approved by the FDA but rarely report information about usability testing. Regarding scientific literature, there are already some studies in this field, such as those previously mentioned on infusion pumps, ventilators, haemodialysis machines, and emergency devices and finally also on ventricular assistive devices (Geidl et al., 2009, 2011; Schima et al., 2014). However, further studies are needed to deal with this increasing need for transparency in this field.

Given this intricate framework, the paper reports the usability and performance testing of a new medical device, BAMBI (Balloon Against Maternal BleedIng), in a controlled simulated scenario. BAMBI was designed to stop postpartum haemorrhages (PPHs). The contribution of this study is three-fold. First, even if focused on a specific case, the paper extends the range of examples of how usability testing for invasive medical devices could be designed, carried out and results evaluated. Second, we demonstrate that the usability and performance of the new device are superior to other solutions used daily in low-resource settings (LRSs). Indeed, finding low-cost, effective alternatives to stop PPHs remains an open research challenge (Hu et al., 2020). Third, from a methodological perspective, in this study, we combine qualitative data with quantitative and objective evaluations and support results using a statistics-based approach. We are aware that qualitative assessments are essential for usability; however, in our case, the strong connection with safety and performance issues also requires a quantitative evaluation of the device characteristics as occurs for other design specifications (e.g., biocompatibility). Complementary qualitative data were also acquired by adopting a convergent parallel mixed approach (Carayon et al., 2015), as they are essential for the interpretation of statistical inferential results. We would like to specify that the use of such an approach is not prescribed by the cited standards (IEC, 2015; FDA, 2016; IEC, 2016) and that the UE process required for regulatory purposes is a more comprehensive activity, which comprises a task analysis, with the identifications of hazards and risks related to each task, and a root cause analysis of detected errors. The advised testing is also iterative, with usability evaluations occurring early in the development process, providing "formative" assessments leading to a "summative" evaluation of the final design. However, given the early-stage characteristic of the research, we performed a preliminary assessment, which, even if simplified, allowed us to gather useful insights for improving the BAMBI design instead of preparing the device for a certification process.

The paper is organised as follows: Section 2 presents the BAMBI device and describes the research context; Section 3 provides the methods of the usability testing; Section 4 reports the results; Section 5 discusses them, while Section 6 ends the paper.

2. Uterine balloon tamponade devices

PPH is an obstetric emergency defined as a blood loss above 500 ml within 24 h of birth (International Federation of Gynecology and Obstetrics, 2012; WHO, 2012; WHO, 2021). It affects about 5% of all women giving birth and is associated with nearly one-quarter of maternal deaths worldwide. 99% of these deaths occur in LRSs, where it is the leading cause of maternal deceases (Say et al., 2014; WHO, 2023). Improving access to safe and effective interventions to prevent and treat PPH is critical for achieving the third Sustainable Development Goal (SDG 3), particularly target 3.1 (United Nations), (Alkema et al., 2016). The last published trends show that, despite the successful reduction of the global maternal mortality ratio (MMR, defined as the number of maternal deaths per 100,000 live births) between 2000 and 2015 (from 339 to 227), a strong stagnation happened in the first five years of the SDG era, with an MMR equal to 223 in 2020 (WHO, 2023).

PPH can be managed in developed countries using first-line interventions, such as uterotonics drugs, tranexamic acid, and intravenous fluids (WHO, 2017, WHO, 2018). However, in case of refractory (unresponsive) PPH, uterine balloon tamponade (UBT) devices like the Bakri® Postpartum Balloon (Cook Medical) can be used as a second-line treatment to avoid invasive surgical interventions (hysterectomy) (WHO, 2021). Usually, UBT devices comprise a pre-assembled balloon catheter system that allows blood drainage and tamponade. The tamponade effect is achieved by applying pressure on the bleeding uterine vessels. On the contrary, PPH is life-threatening in LRSs because drugs and commercial UBT solutions are unaffordable and, at the same time, operating theatres are not always available (WHO, 2021). The low-cost improvised version of UBT is the Condom Balloon Tamponade (CBT) solution (Fig. 1a). It consists of a condom or probe cover tied to a catheter, through strings or sutures, inserted into the uterus, and filled with physiological solution or water to stop the bleeding (WHO, 2012) (Fig. 1a). The CBT has several limitations, such as fluid leakages and the need for specific training and manual skills for performing reliable knots. In 2019, a low-cost CBT kit, Every Second Matters for Mothers and Babies™ was approved by the FDA (Massachusetts General Hospital; FDA, 2019). However, it still presents weaknesses and usability issues (manual skills are still required). Also, the publicly available FDA approval report (FDA, 2019) does not provide data on usability evaluations. The device's effectiveness was supported by peer-reviewed publications reporting on device adoption in some sub-Saharan and Indian peninsula countries.

The UBT procedure was introduced in WHO guidelines in 2012 and marked as a "weak recommendation with very-low quality of evidence" (WHO, 2012). However, new evidence on using UBT devices in the following years forced the WHO to update this recommendation. A new guideline was published in 2021 (WHO, 2021). The UBT procedure is now recommended for treating PPH due to uterine atony after vaginal birth in women that do not respond to standard first-line treatment if some conditions are met (WHO, 2021). Context-specific preconditions are set to minimise harm to the woman. They include the non-delay of more invasive treatment if available, the exclusion of other causes of PPH to avoid incorrect patient selection, and the requirement that the procedure is performed by personnel trained and skilled in the management of PPH, including UBT use (WHO, 2021). The guideline also states that where UBT is used, appropriate training of health workers is required. However, the most effective approach to UBT training is not yet known. The WHO recognises this knowledge gap as an urgent priority to address (WHO, 2021).

In such a context, the BAMBI device was developed (Fig. 1b). It is a new low-cost CBT device designed to be used in LRSs. It is equipped with an innovative connector to assemble the condom to the catheter as a more reliable and easier-to-assemble alternative to sutures (Costantino et al., 2021) used for standard CBT devices (Fig. 1a). This connector is included in a package together with all the other elements of the device (Fig. 1b). More details about the BAMBI device can be found in (Candidori et al., 2021; Costantino et al., 2021; Candidori et al., 2023). Hence, a standard CBT device and BAMBI differ in how the condom is assembled to the catheter, i.e., the sutures for the first and the designed connector for the second. All the other components of the kit can be the same.

BAMBI design requirements are discussed in (Candidori et al., 2021). *In vitro* tests concerning the technical performances of the device were performed but not included in this study.

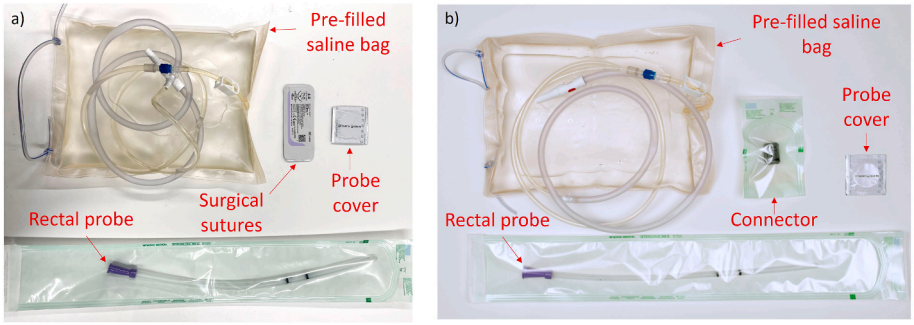


Fig. 1. Components of the two CBT kits analysed in this study. a) An example of the components of a standard CBT device: rectal probe, pre-filled saline bag, surgical sutures, and probe cover. Ethicon Vicryl™ 4/0, 45 cm, was used for the testing, even if other surgical sutures can be used. b) The BAMBI device with its rectal probe, pre-filled saline bag, connector, and probe cover. Apart from the connector, the other components are the same as the standard CBT device (a). In both cases, a catheter can substitute the rectal probe.

3. Methods

3.1. Study design and participants

The study aims to assess the usability and performance of the BAMBI device for the treatment of PPH in a controlled *in vitro* environment when used by subjects with and without a gynaecological/obstetric background (from now on, "medical" and "non-medical" users, respectively). Non-medical users were included because, in LRSs, it is not uncommon for unqualified subjects (e.g., relatives) to help with the birth. The study was approved by Politecnico di Milano's Ethical Committee. Recruited subjects agreed to be filmed in the consent form.

The proposed method of investigation consists of an *in vitro* simulation of the tamponade procedure, followed by a questionnaire.

The study involved a total of 62 participants who voluntarily took part in the experiment. In particular:

- medical users: 17 participants (12 females, 5 males), aged between 25 and 54 years (M = 30.82; SD = 7.59);
- non-medical users: 45 participants (27 females, 18 males), aged between 23 and 58 years (M = 26.58; SD = 7.60).

Medical users were healthcare professionals represented by gynaecology/obstetrics specialists and residents with a medicine and surgery degree, and midwives with a nursing and midwifery sciences degree. They work in the hospitals of Carate Brianza and Monza (Fondazione IRCCS San Gerardo dei Tintori) and were recruited on a voluntary basis through a flyer posted on the hospital notice board. Non-medical users were recruited mainly from students and professors of Politecnico di Milano who frequent the campus daily, simply asking them if they would like to participate in our experimental campaign. Each participant provided written informed consent for participation in the study. Both groups of participants were asked to perform a CBT procedure with the BAMBI device. Besides, further experimental conditions were also studied.

To compare the BAMBI device (Fig. 1b) with a standard CBT solution (Fig. 1a), a within-subjects variable was included in the group of medical users, who, therefore, were asked to use both devices after a live session training. This comparison was limited to medical users to collect reliable data from qualified personnel to be used as a baseline.

Besides, providing clear instructions is essential, especially for non-medical users. Considering that BAMBI is used in emergencies, where the chances that the user has received in-depth formations with qualified personnel could be meagre, we decided to introduce a between-subjects variable represented by the training modality in the non-medical participants' group. The training was provided through:

1. Live session (i.e., participants received live training from qualified personnel on how to perform the procedure);
2. Paper IFU;
3. Video training (i.e., a short video showing how to perform the procedure).

Frames of the video used for the training are provided in the Supplementary Material (Fig A1). This figure clearly explains the main steps to assemble and insert the device into the uterus. Non-medical participants were thus divided into three further subgroups, depending on the training modality. The study's objectives are reported in Table 1.

3.2. Procedure

The testing was performed in our laboratory at Politecnico di Milano (with non-medical users) and in the hospitals of Carate Brianza and Monza (with medical users) using the same portable experimental setup, easy-to-assemble on a standard table. The tests involving medical users were performed in the hospital for practical and logistical reasons and in an empty meeting room. Medical users were left free to join the room for the tests when they were free from their work commitments. Fig. 2 shows the two types of testing scenarios.

The experiment consisted of a simulation of PPH and the related CBT

Table 1
Overview of the study's objectives.

# Analysis	User (device)	Objective To evaluate the usability and performance of:
1	Medical users (BAMBI and CBT)	the BAMBI device vs the CBT solution when used by medical users
2	Non-medical users (BAMBI)	the BAMBI device when used by non-medical users undergoing different training (i.e., live training session, paper IFU reading, video watching)
3 4	Medical and non-medical users (BAMBI)	the BAMBI device when used by medical vs non-medical users (same training, i.e., only live session, is considered in analysis 3, while all non-medical users were pooled in analysis 4)
5 6	Medical (CBT) and non-medical users (BAMBI)	medical users using the emergency CBT solution vs non-medical users using the BAMBI device (same training, i.e., only live session, is considered in analysis 5, while all non-medical users were pooled in analysis 6)

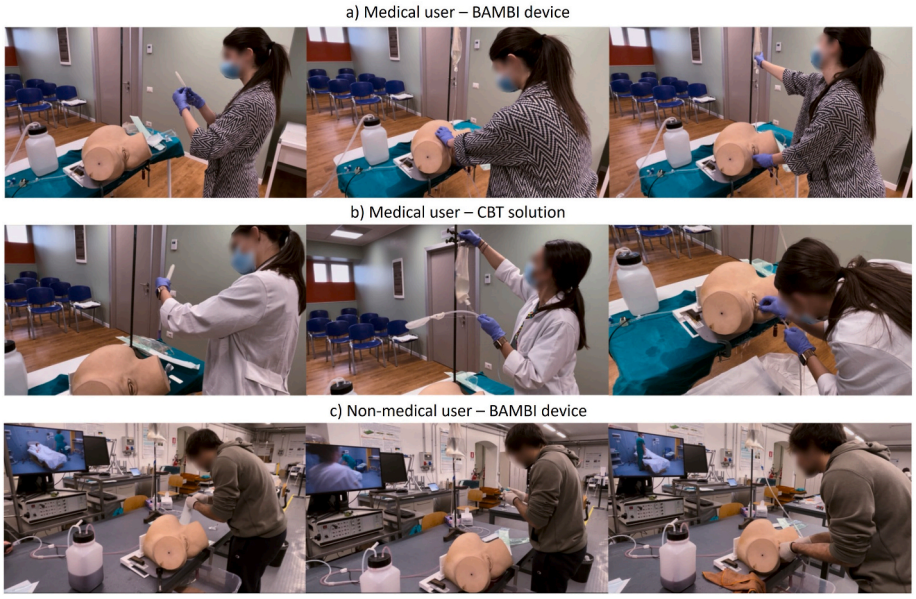


Fig. 2. Usability experimental tests. Snapshots taken from the videos: a) medical user performing the tamponade procedure with the BAMBI device; b) medical user performing the tamponade procedure with the CBT solution; c) non-medical user performing the tamponade procedure with the BAMBI device. a) and b) show the testing scenario inside the hospital. c) shows the testing scenario inside the laboratory.

procedure using BAMBI (for both groups) and a standard CBT solution (only for medical users). At least two experimenters were present. Before the test, all recruited participants were requested to read and sign a consent form and to perform a pre-test questionnaire, in which data on age, gender, and profession were collected.

Medical users were trained individually by the same experimenter for both devices. They performed the first CBT procedure with one of the two devices. After a break, the second CBT procedure with the other device was fulfilled. The order of the tested devices was counterbalanced between participants. Since no standard procedure for the CBT solution (Fig. 1a) is available, medical users were left free to secure the probe cover to the rectal probe using sutures according to their medical experience (an example is shown in Fig A2, Supplementary Material). After the test, participants were asked to fill in a questionnaire, through which data about the usability of the two devices, their preferred device and any free comments were collected.

Non-medical users were divided into three subgroups according to three training modalities. After the training, each participant performed the CBT procedure with the BAMBI device. Finally, they were asked to fill in a questionnaire, through which data about the device's usability were gathered. Further details about each training modality are provided in the Supplementary Material (A and B).

During each test, a video recording was performed by one of the two operators; the *a posteriori* analysis of the videos allowed us to obtain data about procedure correctness and times. Each video lasts approximately 3:30 min, with video recording starting a few seconds before the official start of the test (the operator declared that the subject could start and begin the bleeding simulation), from which times were measured.

When the positive tamponade effect was achieved, or the test was stopped in advance, e.g., in case of incorrect assembly, the two experimenters checked the procedure correctness, effectiveness, and the balloon's anatomical position. Fig. 2 shows some moments of the tests. Each session lasted about 20 min for medical and 10 min for non-medical users. All data were analysed and reported anonymously.

3.3. Experimental setup

The experimental set-up (Fig. 3) consisted of a manikin for obstetric simulations and training (Model-med). It is the external portion of the female reproductive system. The manikin was supplemented with a 3D-printed phantom of a uterus. The phantom is an updated version of the one we presented in (Candidori et al., 2023). It was obtained by stereolithography using a flexible and transparent material. The geometry was shaped to fit the back cavity of the simulator, using a 3D scanning technology to acquire the cavity's geometry. This phantom was used to mimic the presence of the uterus and thus to allow a standard manikin to work appropriately as a testing set-up for simulating the bleeding. A support (Fig. 3, left) was designed and then 3D-printed using the Fused Filament Fabrication technology to let the uterus maintain its position during the procedure. The support was only visible from the experimenter's point of view. It did not alter the test setup.

A hydraulic circuit connected to the phantom allowed the bleeding

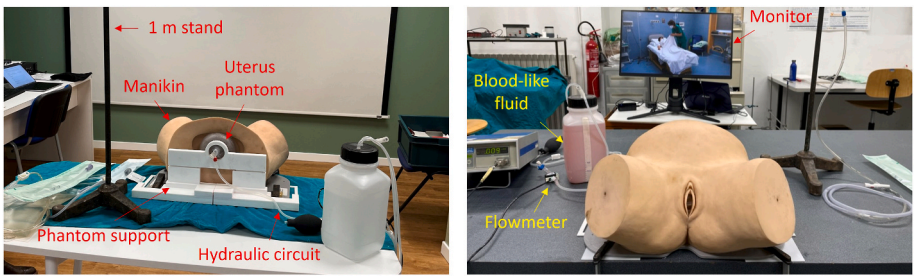


Fig. 3. Experimental setup used for the simulation of the PPH tamponade procedure: back view (left, experimenter's point of view) and front (right, subject's point of view). The set-up is portable and easy-to-assemble.

simulation using red-coloured water in a pressurised container. The circuit was equipped with a flowmeter to monitor the volumetric flow rate of the bleeding, set at 500 ml/min. This value was chosen by slightly reducing the standard blood flow rate to the uterus at full term, approximately 600 ml/min (Bienstock et al., 2021). For each test, the participants were provided with all the components of the standard CBT (Fig. 1a) or BAMBI (Fig. 1b). A 1 m stand hung the saline bag during the test. To simulate an emergency and stressful situation, a monitor was used to project a video in which a critical care condition in a hospital setting was replicated for the tests involving non-medical users (Fig. 3, right). A laptop or a tablet was used to record questionnaire answers using the Qualtrics website platform (Qualtrics International Inc., Seattle, Washington, USA).

3.4. Measures

The data obtained from the experimental campaign are related to *demographics*, *usability score* and *procedure-related variables*.

Demographic data concern the subject's age, gender, and profession. They were collected through pre-test questionnaires. While the profession divided users into two main groups (i.e., medical or non-medical), age and gender were acquired but not considered.

The *usability score* of the two devices was evaluated using the Italian-translated and validated version of the System Usability Score (SUS), a reliable usability scale typically used for global assessments of usability (Brooke, 1996). This score can quantify usability intended as appropriateness to a purpose, combining several aspects of effectiveness, efficiency and satisfaction (Brooke, 1996). The SUS scale was selected as it is relatively quick and easy to use, it is technology-agnostic (which allows its application on a wide range of products, comprising healthcare devices), non-proprietary, reliable and easy to understand also by non-experts as it finally provides a single score on a percentage scale (Bangor et al., 2008). This single number is very useful for relative judgements (e.g., comparing competitive alternatives or iterative versions of a product), but has the limitation of not being fully and easily interpretable as a single absolute value (Bangor et al., 2008). The standard ten-item Likert (1–5 points) scale was modified according to the specific application. The first question, "I think that I would like to use this system frequently", was removed because it was not deemed appropriate for our study. Previous studies showed that it is possible to leave out any SUS items without a practically significant effect on the resulting scores (Lewis and Sauro, 2017). We converted the nine-item contributions into an overall SUS score, expressed in percentage (Brooke, 1996; Lewis and Sauro, 2017). A SUS score was calculated for each test.

For each test, the following *procedure-related data* were extracted. Three dichotomous (yes/no) variables were evaluated from the videos: *procedure correctness*, *tamponade effectiveness* and *uterine positioning*. *Tamponade effectiveness* was positively scored (i.e., yes) if a positive tamponade effect (defined in the literature as the stopping of the bleeding (Georgiou, 2010)) was achieved, regardless of the balloon position. If the balloon was positioned inside the uterine cavity at the uterine fundus, which is the optimal location (Bienstock et al., 2021), *uterine positioning* was also scored positively. This variable was introduced because there are pieces of evidence of a positive tamponade effect also when the balloon inflates into the vaginal canal or in the lower uterus segment (Cho et al., 2008). The procedure was considered correct if all the tasks explained during the training were performed correctly. Difficulties and "close calls" (IEC, 2016) were annotated but did not result in an incorrect evaluation because they did not ultimately compromise the overall correct use. Lastly, assembly and manoeuvre times were measured in seconds. Assembly time was calculated from the beginning of the procedure to the end of the assembly. This step concerns the connection of the rectal probe to the bag (step 7 in Fig A1, Supplementary Material). Manoeuvre time was measured from the end of the assembly to the end of the procedure (i.e., the positive tamponade was achieved, or the experimenter stopped the test). The total procedure time is the sum of the assembly and manoeuvre time.

3.5. Data analysis

When normally distributed, we present continuous data as mean (M) and standard deviation (SD). The median value (Q2) and the interquartile range (IQR) were used when data were not normally distributed. Contingency tables were used to describe categorical procedure-related data.

For the medical user analysis (*#1* in Table 1, Section 3.1), a within-subjects approach (paired tests) was followed. A between-subjects analysis was performed for all the other comparisons (*#2–6*).

Two normality tests (i.e., Anderson-Darling and Ryan-Joiner) were carried out to determine if data related to the SUS score, assembly and manoeuvre times were normally distributed. When data were not normally distributed, outlier detection tests (boxplot analysis and Grubb's test) were used to detect these outliers; normality was then rechecked after removal.

As regards the continuous variables, paired *t*-test (normal data) and Wilcoxon signed-rank test (not normal data) were used for the comparisons in the medical user analysis (*#1*). T-test and one-way ANOVA were used to compare two or three groups when data were normally distributed for all the other analyses (*#2–6*). Mann-Whitney U and Kruskal-Wallis H tests were used to compare groups for skewed distributions. If a statistically significant difference was detected among more than two groups, the Bonferroni correction was applied to lower the critical p-value when performing multiple comparisons. For the non-expert user analysis (*#2* in Table 1), for example, a critical value of $\alpha/[N(N-1)/2]$, with N = 3, was considered when comparing the three subgroups. Dichotomous variables were compared using the χ^2 test for proportions.

Two-sided p-values of less than 0.05 (α) were considered statistically significant. All the analyses were performed using Minitab® (version 21.3.1).

4. Results

The results are presented following the scheme in Table 1. A summary of the outcomes is provided in Table 2. The comparisons between continuous variables are shown in Fig. 4, while those between dichotomous ones are summarised in Fig. 5.

4.1. Analysis #1: medical users (BAMBI and CBT)

SUS score. The usability score of the BAMBI device (M = 93.92, SD = 4.78) is significantly higher (t = 5.84, p < 0.001) than that of the CBT solution (M = 62.15, SD = 24.23) (Fig. 4a). The distribution of the data is narrower for the BAMBI score than the CBT one, as underlined by the lower SD and standard error (SE) of the mean ($SE_{BAMBI} = 1.20$, $SE_{CBT} = 6.06$), showing more agreement between the SUS score of the different expert users.

Categorical procedure-related variables. There is a significant difference between procedure correctness (χ^2 (1) = 14.17, p < 0.001), tamponade effectiveness (χ^2 (1) = 14.17, p < 0.001) and uterine positioning (χ^2 (1) = 12.88, p < 0.001) proportions of the two devices. In all cases, the differences favour BAMBI (Fig. 5a).

Assembly and manoeuvre time. The assembly time of BAMBI (Q2 = 40.00, IQR = 10.00) is significantly lower (U = 0.00, p = 0.003) than that of the CBT solution (Q2 = 60.00, IQR = 22.50). No statically significant difference (t = −1.60, p = 0.130) was found between the manoeuvre times of the two devices ($M_{BAMBI} = 74.06$, $SD_{BAMBI} = 8.61$, $M_{CBT} = 85.00$, $SD_{CBT} = 24.43$) (Fig. 4b).

Table 2
Summary of the results of the study. The symbol "-" is used when p-value> 0.05 (no statistical significance detected), while "*" when 0.01 < p-value ≤0.05 (statistical significance detected). The symbol "**" is used when 0.001 < p-value ≤0.01 (high statistical significance detected) and "***" when p-value ≤0.001 (very high statistical significance detected).

# Analysis	Users	Device	Training	SUS score	Procedure correctness	Tamponade effectiveness	Uterine positioning	Assembly time	Manoeuvre time
1	Medical	BAMBI CBT	live	***	***	***	***	**	–
2	Non-medical	BAMBI	live paper IFU video	–	–	–	–	***	*
3	Medical Non-medical	BAMBI	live	–	–	–	*	***	***
4	Medical Non-medical	BAMBI	live all	–	–	–	**	***	***
5	Medical Non-medical	CBT BAMBI	live	**	***	***	–	–	–
6	Medical Non-medical	CBT BAMBI	live all	***	***	***	–	***	–

Notes.
#1. The BAMBI device was better in all cases in which a statistically significant difference was shown.
#2. Regarding the assembly time: live < (paper IFU = video); regarding the manoeuvre time: paper IFU < video.
#3. and 4. The medical users performed the task better in all cases in which a significant difference was obtained.
#5. In all cases where a significant difference was shown, the non-medical users using BAMBI performed the task better than medical users using CBT.
#6. SUS score, procedure correctness and tamponade effectiveness are better in the non-medical group using BAMBI, while assembly time is shorter in the medical one using CBT.

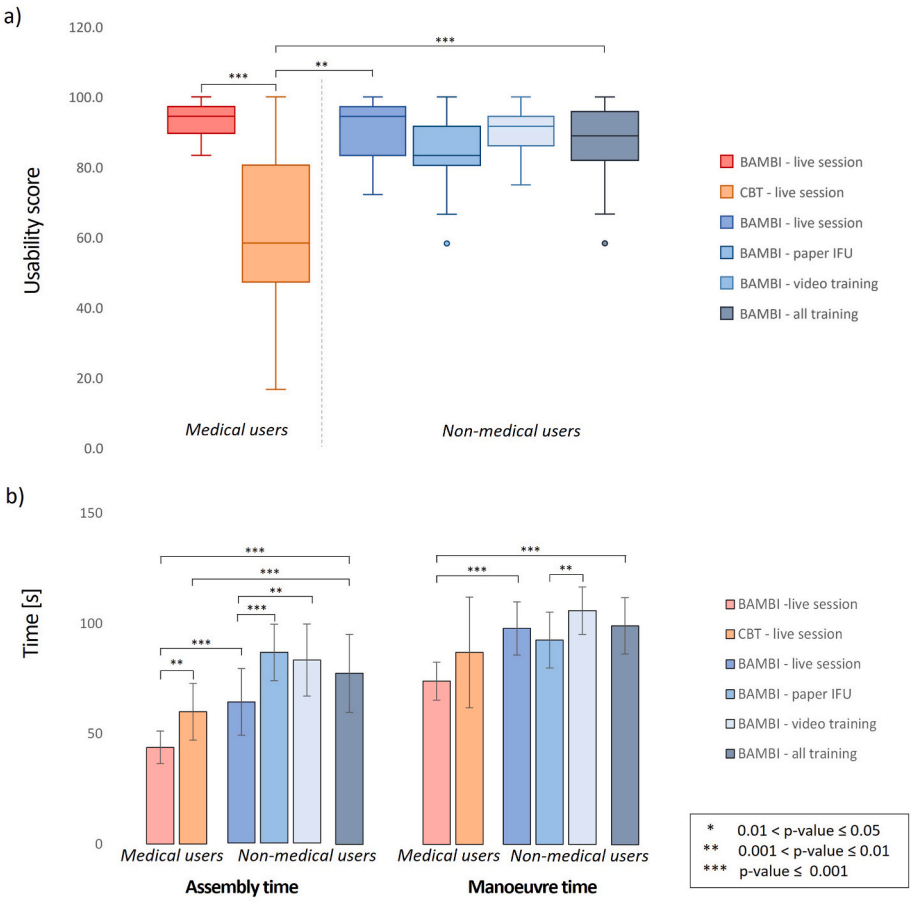


Fig. 4. Comparison of the user groups' usability score (boxplot in a), assembly and manoeuvre time (histograms in b). Statistically significant differences are reported (as in Table 2). For non-medical users, both the training-based subgroups and pooled data are highlighted.

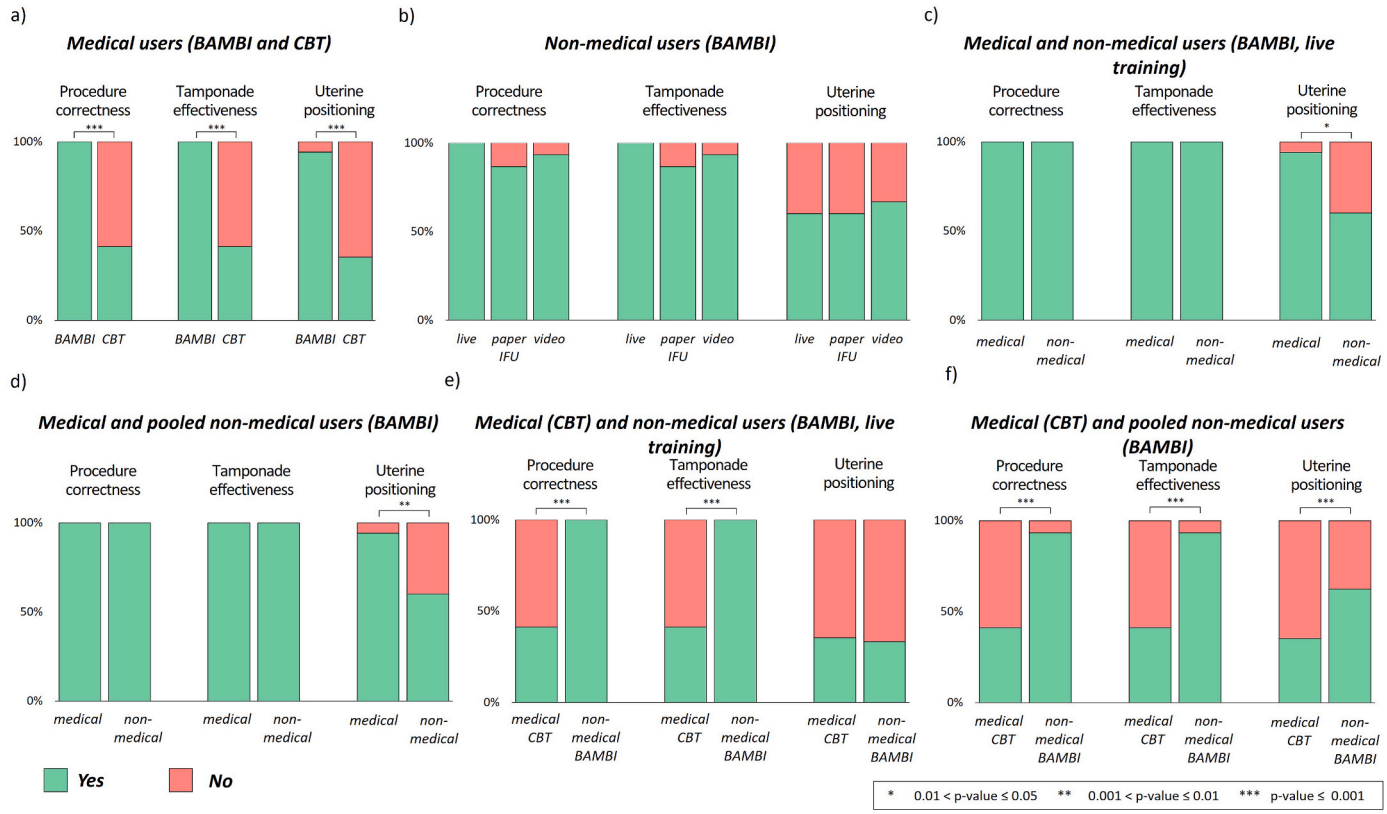


Fig. 5. Stacked histograms of the procedure correctness, tamponade effectiveness, and uterine positioning proportions for the different analyses (a-f, corresponding to analyses #1–6). "Yes" results are shown in green, while "no" in red. Statistically significant differences are reported (as in Table 2).

4.2. Analysis #2: non-medical users (BAMBI)

SUS score. There is no statically significant difference (H = 3.35, p = 0.187) between the usability score of the BAMBI device after different training modalities ($Q2_{live}$ = 94.44, IQR_{live} = 13.89, $Q2_{IFU}$ = 83.33, IQR_{IFU} = 11.11, $Q2_{video}$ = 91.67, IQR_{video} = 8.33) (Fig. 4a).

Categorical procedure-related variables. There is no statistically significant difference between procedure correctness (χ^2 (2) = 2.14, p = 0.343), tamponade effectiveness (χ^2 (2) = 2.14, p = 0.343) and uterine positioning (χ^2 (2) = 0.19, p = 0.910) proportions of BAMBI following the three training modalities (Fig. 5b).

Assembly and manoeuvre time. The assembly time after IFU reading (M = 86.43, SD = 12.77) is significantly higher (t = −4.32, p < 0.001, with α = 0.0167 after Bonferroni correction) than after the live training (M = 64.00, SD = 15.14). The same result (t = −3.30, p = 0.003) was obtained by comparing live session training with video watching (M = 83.00, SD = 16.34). No statistically significant difference (t = 0.63, p = 0.533) was detected between IFU reading and video watching. Considering manoeuvre times, no statistically significant difference (W = 274.50, p = 0.083) was found between the live session (Q2 = 100.00, IQR = 15.00) and IFU reading (Q2 = 90.00, IQR = 15.00). A similar result (W = 185.50, p = 0.049) was obtained comparing the live session and video watching (Q2 = 110.00, IQR = 10.00). Eventually, the manoeuvre time after the IFU reading is statistically significantly lower (W = 158.5, p = 0.002) than after video watching (Fig. 4b).

4.3. Analysis #3: medical and non-medical users (BAMBI, live training)

SUS score. There is not a statistically significant difference (W = 285.00, p = 0.878) between the usability score of BAMBI from medical (Q2 = 94.44, IQR = 8.33) vs. non-medical users (Q2 = 94.44, IQR = 13.89) after live training (Fig. 4a).

Categorical procedure-related variables. There is no statistically significant difference between procedure correctness (identical proportions, p = 1.000) and tamponade effectiveness (identical proportions, p = 1.000) proportions between medical and non-medical users when using BAMBI after live training. A statistically significant difference (expert better than non-expert) (χ^2 (1) = 5.82, p = 0.016) was found between the uterine positioning proportions with BAMBI of the two groups (Fig. 5c).

Assembly and manoeuvre time. The assembly time of medical users (Q2 = 40.00, IQR = 10.00) was significantly lower (W = 181.50, p < 0.001) than that of non-medical users (Q2 = 70.00, IQR = 25.00) when using the BAMBI device after live training. A similar result (t = −6.32, p = 0.001) was obtained comparing the manoeuvre times of the two groups, with that of medical users (M = 74.06, SD = 8.61) lower than that of non-medical ones (M = 98.00, SD = 12.07) (Fig. 4b).

4.4. Analysis #4: medical and pooled non-medical users (BAMBI)

SUS score. There is not a statistically significant difference (W = 640.50, p = 0.097) between the usability score of BAMBI assigned by medical (Q2 = 94.44, IQR = 8.33) and non-medical users (Q2 = 88.89, IQR = 13.89). This means that the results of analysis #3 about the usability score can be extended independently of the training modality (Fig. 4a).

Categorical procedure-related variables. There is no statistically significant difference between procedure correctness (χ^2 (1) = 1.980, p = 0.555) and tamponade effectiveness (χ^2 (1) = 1.980, p = 0.555) proportions between medical and non-medical users when using BAMBI. A statistically significant difference (expert better than non-expert) (χ^2 (1) = 7.43, p = 0.006) was found between the uterine positioning proportions with BAMBI of the two groups (Fig. 5d).

Assembly and manoeuvre time. The assembly time of medical users

(Q2 = 40.00, IQR = 10.00) was significantly lower (W = 186.00, p < 0.001) than that of non-medical ones (Q2 = 80.00, IQR = 20.00) when using the BAMBI, independently on the training modality. A similar result (t = −8.66, p = 0.001) was obtained comparing the manoeuvre times of the two groups, with that of medical (M = 74.06, SD = 8.61) lower than that of non-medical users (M = 99.19, SD = 12.77) (Fig. 4b).

4.5. Analysis #5: medical (CBT) and non-medical users (BAMBI, live training)

SUS score. The usability score of BAMBI for non-medical users (Q2 = 94.44, IQR = 13.89) is significantly higher (W = 192.00, p = 0.001) than that of the CBT solution for the medical ones (Q2 = 58.33, IQR = 33.33) (Fig. 4a).

Categorical procedure-related variables. There is a significant difference between procedure correctness (χ^2 (1) = 16.72, p < 0.001) and tamponade effectiveness (χ^2 (1) = 16.72, p < 0.001) proportions of the two groups. No statistically significant difference (χ^2 (1) = 1.97, p = 0.160) was found between the uterine positioning proportions of the two groups (Fig. 5e).

Assembly and manoeuvre time. There is not a statistically significant difference (t = −0.81, p = 0.435) between the assembly time of medical users with the CBT solution (M = 60.00, SD = 12.87) and that of non-medical ones when performing the procedure with BAMBI (M = 64.00, SD = 15.14). Moreover, there is not a statistically significant difference (t = −1.60, p = 0.124) between the manoeuvre times of the two groups ($M_{med,CBT}$ = 87.06, $SD_{med,CBT}$ = 25.13, $M_{non\text{-}med,BAMBI}$ = 98.00, $SD_{non\text{-}med,BAMBI}$ = 12.07) (Fig. 4b).

4.6. Analysis #6: medical (CBT) and pooled non-medical users (BAMBI)

SUS score. The usability score of BAMBI for all non-medical users (Q2 = 88.89, IQR = 13.89) is significantly higher (W = 288.50, p < 0.001) than that of the CBT solution for medical ones (Q2 = 58.33, IQR = 33.33) (Fig. 4a).

Categorical procedure-related variables. There is a significant difference between procedure correctness (χ^2 (1) = 18.60, p < 0.001) and tamponade effectiveness (χ^2 (1) = 18.60, p < 0.001) proportions of the two groups. No statistically significant difference (χ^2 (1) = 3.63, p = 0.057) was found between the uterine positioning proportions of the two groups (Fig. 5f).

Assembly and manoeuvre time. The assembly time of medical users with the CBT solution (M = 60.00, SD = 12.87) is significantly lower (t = −4.44, p < 0.001) than that of all non-medical users when performing the procedure with BAMBI (M = 78.89, SD = 19.42). On the other hand, no statistically significant difference (t = −12.13, p = 0.073) was found between the assembly time of the two groups ($M_{med,CBT}$ = 87.06, $SD_{med,CBT}$ = 25.13, $M_{all\ non\text{-}med,BAMBI}$ = 99.19, $SD_{all\ non\text{-}med,BAMBI}$ = 12.77) (Fig. 4b).

Two additional summary tables can be found in the Supplementary Material. Table C1 reports the descriptive statistics of the three continuous variables (SUS score, assembly, and manoeuvre time), divided for the different analyses; Table C2 shows the contingency tables of the three dichotomous variables (procedure correctness, tamponade effectiveness, and uterine positioning), divided for the different analyses.

5. Discussion

The first analysis (Table 2) showed that BAMBI (Fig. 1b) could be used more effectively and quickly for PPH management than the improvised CBT device (Fig. 1a). Besides, all healthcare workers prefer BAMBI over the CBT device. Several positive comments were made about BAMBI: six users said it was quicker to use than the CBT, five found it easier to use, four stated it was quicker and safer in terms of hydraulic sealing, and three pointed out that it was easier to handle. Eventually, one physician stated that it was more comfortable to use, and another that it was easier to handle with medical gloves. These aspects are reflected in the higher SUS score assigned to BAMBI than CBT (Fig. 4a) and the higher percentages of correct and effective tamponade procedures (Fig. 5a). The shorter assembly time confirms the perception of a faster procedure; however, the manoeuvre time is independent of the device, as the second part of the procedure is the same for both devices (steps 7–12 Fig A1, Supplementary Material). A difference of approximately 30 s was measured between the total time of the two procedures (120.6 ± 16.7 s and 147.1 ± 25.7 s with BAMBI and the CBT solution, respectively), which is highly relevant in a life-threatening emergency such as PPH. The 100% procedure correctness is relevant for BAMBI, whereas, in all 10 cases of incorrect use of the CBT solution, the problem was the failure of the knot connection. The lower percentages of correct and tamponade effective procedures (i.e., 41% (7/17) for both variables) of the CBT solution than the BAMBI (i.e., 100% for both variables) were indeed due to the difficulty in performing reliable knots that guarantee hydraulic sealing. A leaky connection leads to a wrong procedure, and if the fluid leakage is abundant, it can result in the inability of the balloon to tamponade the bleeding. In our tests, the fluid leakage due to unreliable knots was always so copious to prevent the tamponade. The lower correct uterine positioning percentage of the CBT solution (6/17, corresponding to 35%) compared to BAMBI (16/17, corresponding to 94%) was due to several aspects. First, the presence of the connector in the BAMBI kit makes it easier to verify proper positioning and avoids inserting a longer portion of the rectal probe than necessary into the uterus, which is less prone to bending than the CBT kit. Second, when using the CBT, the handling of the device in the correct position (i.e., the user has to avoid the slipping of the balloon toward the vaginal canal during inflation) has to be done for a time longer than BAMBI because the fluid leakage slows down the filling. This extra effort can cause more frequent incorrect uterine positioning.

No statistically significant differences were found between the three training modalities for usability score, correct procedure, and positive tamponade, indicating they are all effective. Minor differences were observed in the assembly and manoeuvre times, the most significant being the assembly time for the live training group, which was approximately 20 s less than the other two. The live session, which is more interactive, is thus more effective in conveying the importance of rapid treatment for the mother's survival. Looking at the whole pool of non-medical users, only 3 out of 45 did not perform the procedure correctly. Two users did not open the clamp of the saline bag (step 10 Fig A1, Supplementary Material); one user did not perform the three rotations of the upper part of the connector, which is essential to secure the probe cover to the rectal probe (step 6 Fig A1, Supplementary Material). These errors suggest possible improvements in the design of BAMBI. Following the same approach used to indicate the direction of insertion and rotation of the connector, the words "3 × 360°" could be engraved or printed on top of it to help users remember the number of full rotations to be performed. Similarly, the word "open" and an arrow could be engraved on the clamp.

The third and fourth analyses showed that although the two groups' usability scores and accuracy rates were equivalent, healthcare workers performed the entire procedure more quickly. As the timing of the tasks was clinically relevant to the procedure, as suggested by the standards (FDA, 2016), subjects were informed in advance of the importance of performing the tamponade procedure correctly and quickly. The mean total time for medical and non-medical users was 120.6 ± 16.7 s and 168.9 ± 41.7 s, respectively. A difference of approximately 50 s translates into an additional blood loss of almost 420 ml of blood, given an average blood flow rate of 500 ml/min. These are added to the 1000 ml lost during the first 120 s. Hence, the importance of time may have needed to be further emphasised for non-medical users, who were more dedicated/anxious to complete each task without error and to follow the instructions precisely.

On the other hand, healthcare professionals were fully aware of

every second's importance for the patient's survival. This was demonstrated by their better time performance and their behaviour. Three of them tried to massage the manikin's abdomen to stimulate uterine contraction (Fig. 2a, centre), which in real cases helps to stop bleeding, stimulating uterine contractility (Bienstock et al., 2021). Many also asked if they could squeeze the bag (Fig. 2a, right) to speed up balloon inflation, even though this was not indicated during the training. Eventually, once a positive tamponade was achieved, they immediately asked how much blood had been lost, recognising this as a relevant clinical indication of the procedure's success. Some non-medical users found it challenging to open the sterilisation pouches (steps 2, 3, Fig A1, Supplementary Material) (Fig. 2c, left), contributing to slower performance. Medical users who daily deal with such pouches opened them without problems, and in some cases in clever ways (different from that shown during the training) to further accelerate the intervention. It is worth explaining that these pouches are fundamental to guarantee sterilisation. However, usability and correct and effective use frequencies are equivalent between medical and non-medical users, demonstrating that the provided training was also successful for non-medical subjects.

The fifth and sixth analyses showed that non-medical users' performance with BAMBI is better than that of medical ones with the CBT solution. Theoretically, childbirth should be assisted by skilled professionals in clinical settings. However, hospital accessibility is uneven around the world. Even in a developed country such as the USA (which, according to the latest trends, is the only rich country where the MMR increased from 12 to 20 between 2000 and 2020 (WHO, 2023)), in the last decades, a substantial increase of "out-of-hospital births" and "births in a hospital without an obstetric unit" was observed in rural areas (Kozhimannil et al., 2018). This inevitably results in lower quality maternal and infant care, which is reflected in higher rates of pregnancy-related hospitalisation and infant mortality (Liu et al., 2015; Alkema et al., 2016; Lespérance et al., 2016). BAMBI could thus be spread at the community level as an effective and easy-to-use solution to save mothers' lives worldwide.

There are some limitations to this study. While three different training modalities were evaluated for non-medical users, only live training was considered for medical professionals. After a short discussion with the medical participants, we decided to comply with the common procedure to learn how to use new devices in clinical contexts. Typically, a representative of the device manufacturer explains the procedure. This aspect is also highlighted in the FDA report, which states that nurses prefer to learn how to use a medical device through a hands-on demonstration by a colleague or a manufacturer's representative and that only a limited proportion of them would spend time reading IFUs (FDA, 2016). This illustrates how the protocol can be adapted based on end-user feedback, leading to ongoing changes in choosing the most appropriate training mechanism.

Another limitation is that the simulation of the tamponade procedure was performed immediately after the training. This choice was made to shorten the time necessary for the experimental campaign, which involved many participants. However, the retention of training decays over time. The procedure should not occur immediately after it, as they should approximate real conditions. Possible future research could be aimed at analysing the effects of training timing.

Another limitation could be the small number of participants in each user group. From a statistical perspective, estimating the number of participants is based on assumptions regarding, for example, a fixed, known probability of encountering a problem and the independence of the problems (Virzi, 1992; Nielsen, 1994). However, making these assumptions is highly impractical and misleading in the real world for ordinary usability testing (Faulkner, 2003); moreover, individuals' probabilities of encountering a problem vary significantly depending on the user's characteristics (Faulkner, 2003). In usability testing, the initial historical suggestion of a minimum of 5 participants (Virzi, 1992; Lewis, 1994; Nielsen, 1994) has been gradually increased to 15 as a practical guide (Faulkner, 2003; FDA, 2016), which is sufficient to find at least 90% and, on average, 97% of all problems (Faulkner, 2003). In this study, 15 was set as the minimum number for each user group. Instead of performing a power analysis to estimate sample size, confidence intervals were used to check the power of the statistical results. However, it is essential to emphasise that human factors validation testing for medical devices, as required by regulatory authorities (IEC, 2015; FDA, 2016; IEC, 2016), is primarily qualitative rather than quantitative. Use errors are recorded, but the purpose is not to quantify the frequency of a particular use error or to establish acceptability with respect to the manufacturer's predefined "acceptable level". The goal is to identify the part of the user interface involved in a usability problem or error and to investigate the causes so that the design can be optimised for a safe and effective use (FDA, 2016). In this study, observed use errors suggested useful improvements to the user interface of the BAMBI device.

A general comment on the design of our device is the lack of feedback on the secure connection between the rectal probe and the saline bag, which is a simple press-fit conical type (step 7 Fig A1, Supplementary Material). This is a standard connection, already used in other medical devices. Whilst this did not result in any problem during the usability tests, some users, particularly non-medical ones, needed help understanding whether it was reliable and performed correctly. This could lead to delayed or missing treatment, as the user could damage the device trying to over-tighten it. Future studies could further explore novel solutions based on tactile and visual feedback. None of the participants directly raised this insight, but a few commented aloud that they were unsure about the quality of the connection.

Hence, overall, the performed testing campaign has allowed us to collect multiple feedback from what concerns both the device's design and the training modality. Besides, the simulated scenario has successfully supported us in reaching this target. To this aim, 3D printing technologies have been revealed to be strategically important in supporting the prototyping of the simulated testing set-up.

6. Conclusions

This study collects the measures related to the usability and performance of a new low-cost medical device, BAMBI, conceived to stop PPHs. This device was compared with a standard CBT solution through tests performed in a simulated environment by medical and non-medical subjects. Results demonstrate that BAMBI is more successful than a standard CBT concerning the procedure correctness frequency, tamponade effectiveness and assembly time. Different training modalities were considered. They provided similar results, further supporting the usability of the BAMBI device and its versatility for different users and contexts of use. These results offer a promising initial assessment regarding the usability and effectiveness of BAMBI. Future studies must focus on extending the data collection and improving device validation, as well as on performing clinical evaluations. Together with these experimental results, the study also provides methodological indications on how usability tests of medical devices, particularly invasive medical devices, could be performed. This kind of insight is scarce in scientific and technical literature. Besides, the study also strengthens the relevance of this kind of test, even performed in a simulated environment, to collect feedback to improve the medical device's design and training modalities.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Sara Candidori, Francesco De Gaetano, Kasra Osouli, Serena Graziosi, Alberto Antonio Zanini and Maria Laura Costantino have patent "Uterine device for treating postpartum haemorrhage" (WO2021/220151) pending to Politecnico di Milano.

Acknowledgements

This study is part of the BAMBI (Balloon Against Maternal BleedIng) project, funded by Polisocial Award (2020) – Politecnico di Milano grant. The authors acknowledge Paola Russo for the support in the 3D modelling and fabrication of the uterus phantom and BAMBI device.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apergo.2023.104223.

References

AAMI, 2018. ANSI/AAMI HE75:2009/(R)2018, Human Factors Engineering — Design of Medical Devices.
Alkema, L., Chou, D., Hogan, D., Zhang, S., Moller, A.B., Gemmill, A., Fat, D.M., Boerma, T., Temmerman, M., Mathers, C., Say, L., 2016. Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the un Maternal Mortality Estimation Inter-Agency Group. Lancet 387, 462–474.
Bangor, A., Kortum, P.T., Miller, J.T., 2008. An empirical evaluation of the system usability scale. Int. J. Hum. Comput. Interact. 24, 574–594.
Beer, J.M., McBride, S.E., Mitzner, T.L., Rogers, W.A., 2014. Understanding challenges in the front lines of home health care: a human-systems approach. Appl. Ergon. 45, 1687–1699.
Ben-Menahem, S.M., Nistor-Gallo, R., Macia, G., von Krogh, G., Goldhahn, J., 2020. How the new European regulation on medical devices will affect innovation. Nat. Biomed. Eng. 4, 585–590.
Bienstock, J.L., Eke, A.C., Hueppchen, N.A., 2021. Postpartum hemorrhage. N. Engl. J. Med. 384, 1635–1645.
Bonnette, B., Suggs, J., Tremoulet, P.D., 2017. How useful are handheld ECG devices? Proc. Int. Symp. Hum. Factors Ergon. Heal. Care 6, 154–158.
Bowers, S., 2018. Doctors demand to see evidence on safety of medical devices approved in Europe. BMJ 363, k5105.
Brooke, J., 1996. SUS - a quick and dirty usability scale. In: Usability Evaluation in Industr, vol. 189, pp. 4–7.
Camargo, C.A., Guana, A., Wang, S., Simons, F.E.R., 2013. Auvi-Q versus EpiPen: preferences of adults, caregivers, and children. J. Allergy Clin. Immunol. Pract. 1.
Candidori, S., De Gaetano, F., Osouli, K., Re, A., Volonté, P., Zanini, A.A., Graziosi, S., Costantino, M.L., 2021. Fighting maternal bleeding in low-resource settings: an analysis of design and measurement issues. 2021 IEEE Int. Work. Metrol. Ind. 4.0 IoT, MetroInd 4.0 IoT 2021 - Proc 324–329. doi:10.1109/MetroInd4.0IoT51437.2021.9488502.
Candidori, S., Graziosi, S., Russo, P., Osouli, K., De Gaetano, F., Zanini, A.A., Costantino, M.L., 2023. Design and 3D printing of a modular phantom of a uterus for medical device validation. Rapid Prototyp. J. 29, 7–20.
Carayon, P., Kianfar, S., Li, Y., Xie, A., Alyousef, B., Wooldridge, A., 2015. A systematic review of mixed methods research on human factors and ergonomics in health care. Appl. Ergon. 51, 291–321.
Chaniaud, N., Métayer, N., Megalakaki, O., Loup-Escande, E., 2020. Effect of prior health knowledge on the usability of two home medical devices: usability study. JMIR mHealth uHealth 8.
Chaniaud, N., Megalakaki, O., Capo, S., Loup-Escande, E., 2021. Effects of user characteristics on the usability of a home-connected medical device (Smart Angel) for ambulatory monitoring: usability study. JMIR Hum. Factors 8, 1–16.
Cho, Y., Rizvi, C., Uppal, T., Condous, G., 2008. Ultrasonographic visualization of balloon placement for uterine tamponade in massive primary postpartum hemorrhage. Ultrasound Obstet. Gynecol. 32, 711–713.
Coldewey, B., Diruf, A., Röhrig, R., Lipprandt, M., 2022. Causes of use errors in ventilation devices - systematic review. Appl. Ergon. 98.
Costantino, M.L., De Gaetano, F., Osouli, K., Zanini, A., Possenti, L., Graziosi, S., Candidori, S., 2021. Uterine Device for Treating Postpartum Hemorrhage - WO 2021/220151.
Cook Medical. Bakri® postpartum balloon with rapid instillation components. https://www.cookmedical.com/products/wh_sosr_webds/.
Edwards, E., Kessler, C., Cherne, N., Dissinger, E., Shames, A., 2018. Human factors engineering validation study for a novel 0.1-mg epinephrine auto-injector. Allergy Asthma Proc. 39, 461–465.
European Parliament, Council of the European Union, 2017. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on Medical Devices.
Fairbanks, R.J., Caplan, S.H., Bishop, P.A., Marks, A.M., Shah, M.N., 2007. Usability study of two common defibrillators reveals hazards. Ann. Emerg. Med. 50, 424–432.
Faulkner, L., 2003. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav. Res. Methods, Instruments, Comput. 35, 379–383.
FDA, 2016. Applying Human Factors and Usability Engineering to Medical Devices. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/applying-human-factors-and-usability-engineering-medical-devices.
FDA, 2017. Examples of Reported Infusion Pump Problems. https://www.fda.gov/medical-devices/infusion-pumps/examples-reported-infusion-pump-problems#3.
FDA, 2018. Infusion Pumps. https://www.fda.gov/medical-devices/general-hospital-devices-and-supplies/infusion-pumps.
FDA, 2019. 510 (K) Clearance for ESM-UBT. https://www.accessdata.fda.gov/cdrh_docs/pdf19/K191264.pdf.
FDA, 2022. How to Determine if Your Product Is a Medical Device. https://www.fda.gov/medical-devices/classify-your-medical-device/how-determine-if-your-product-medical-device.
Feinmann, J., 2019. Medical devices: new EU regulations won't guarantee safe design, doctors warn. BMJ 365, l4446.
Fraser, A.G., et al., 2018. The need for transparency of clinical evidence for medical devices in Europe. Lancet 392, 521–530.
Gao, M., Kortum, P., 2017. Measuring the usability of home healthcare devices using retrospective measures. Proc. Hum. Factors Ergon. Soc. 2017-Octob 1281–1285.
Geidl, L., et al., 2009. Usability and safety of ventricular assist devices: human factors and design aspects. Artif. Organs 33, 691–695.
Geidl, L., et al., 2011. Intuitive use and usability of ventricular assist device peripheral components in simulated emergency conditions. Artif. Organs 35, 773–780.
Georgiou, C., 2010. Intraluminal pressure readings during the establishment of a positive 'tamponade test' in the management of postpartum haemorrhage. BJOG An Int. J. Obstet. Gynaecol. 117, 295–303.
Guerlain, S., Hugine, A., Wang, L., 2010. A comparison of 4 epinephrine autoinjector delivery systems: usability and patient preference. Ann. Allergy Asthma Immunol. 104, 172–177.
Hu, K., et al., 2020. Improved treatment of postpartum hemorrhage: design, development, and bench-top validation of a reusable intrauterine tamponade device for low-resource settings. J. Med. Devices, Trans. ASME 14, 16–18.
Hunt, E.A., et al., 2009. Delays and errors in cardiopulmonary resuscitation and defibrillation by pediatric residents during simulated cardiopulmonary arrests. Resuscitation 80, 819–825.
IEC, 2015. IEC 62366-1:2015 Medical Devices — Part 1: Application of Usability Engineering to Medical Devices.
IEC, 2016. IEC/TR 62366-2: 2016 Medical Devices — Part 2: Guidance on the Application of Usability Engineering to Medical Devices.
International Federation of Gynecology and Obstetrics, 2012. FIGO Guidelines: prevention and treatment of postpartum hemorrhage in low-resource settings. Int. J. Gynecol. Obstet. 117, 108–118.
ISO, 2016. ISO 13485:2016 Medical Devices — Quality Management Systems — Requirements for Regulatory Purposes.
ISO, 2019. ISO 14971:2019 Medical Devices. Application of Risk Management to Medical Devices.
Kessler, C., Edwards, E., Dissinger, E., Sye, S., Visich, T., Grant, E., 2019. Usability and preference of epinephrine auto-injectors: auvi-Q and EpiPen jr. Ann. Allergy Asthma Immunol. 123, 256–262.
Klarich, A., Noonan, T.Z., Reichlen, C., Marie, S., Barbara, J., Cullen, L., Pennathur, P.R., 2022. Usability of smart infusion pumps : a heuristic evaluation. Appl. Ergon. 98, 103584.
Kortum, P., Peres, S.C., 2015. Evaluation of home health care devices: remote usability assessment. JMIR Hum. Factors 2.
Kozhimannil, K.B., Hung, P., Henning-Smith, C., Casey, M.M., Prasad, S., 2018. Association between loss of hospital-based obstetric services and birth outcomes in rural counties in the United States. JAMA, J. Am. Med. Assoc. 319, 1239–1247.
Lenzer, J., 2018. Medical device industry: international investigation exposes lax regulation. BMJ 363, k4997.
Lespérance, S., Miller, K., Dworkin, R., Smith-Windsor, T., 2016. Maternal morbidity and perinatal outcomes in rural versus urban areas (2). CMAJ (Can. Med. Assoc. J.) 188, 1261–1262.
Lewis, J.R., 1994. Sample sizes for usability studies: additional considerations. Hum. Factors 36, 368–378.
Lewis, J.R., Sauro, J., 2017. Can I leave this one out ? The effect of dropping an item from the SUS. J. Usability Stud. 13, 38–46.
Liu, L., Oza, S., Hogan, D., Perin, J., Rudan, I., Lawn, J.E., Cousens, S., Mathers, C., Black, R.E., 2015. Global, regional, and national causes of child mortality in 2000-13, with projections to inform post-2015 priorities: an updated systematic analysis. Lancet 385, 430–440.
Lyons, I., Blandford, A., 2018. Safer healthcare at home: detecting, correcting and learning from incidents involving infusion devices. Appl. Ergon. 67, 104–114.
Massachusetts General Hospital. Every second Matters for mothers and babies: uterine balloon tamponade for postpartum hemorrhage. https://www.massgeneral.org/emergency-medicine/global-health/initiatives-and-programs/every-second-matters-for-mothers-and-babies-uterine-balloon-tamponade-for-postpartum-hemorrhage.
Medical devices must be carefully validated. Nat. Biomed. Eng. 2, 2018, 625–626.
Model-med. Lucy and lucy's mum instrumental delivery birth simulator. https://modelmed.com.au/products/lucy-instrumental-delivery-birth-simulator/.
Monsieurs, K.G., Vogels, C., Bossaert, L.L., Meert, P., Calle, P.A., 2005. A study comparing the usability of fully automatic versus semi-automatic defibrillation by untrained nursing students. Resuscitation 64, 41–47.
Moss, R.B., Daniels, K., Moll, T., Carlo, D.J., 2018. Human factors study in untrained adolescents comparing a recently approved single-dose epinephrine prefilled syringe with an approved autoinjector. Ann. Allergy Asthma Immunol. 120, 540–541.
Nielsen, J., 1994. Usability Engineering. Morgan Kaufmann.
Privitera, M.B., Evans, M., Southee, D., 2017. Human factors in the design of medical devices – approaches to meeting international standards in the European Union and USA. Appl. Ergon. 59, 251–263.
Rajkomar, A., Farrington, K., Mayer, A., Walker, D., Blandford, A., 2014. Patients' and carers' experiences of interacting with home haemodialysis technology: implications for quality and safety. BMC Nephrol. 15, 1–12.
Read, G.J.M., Shorrock, S., Walker, G.H., Salmon, P.M., 2021. State of science: evolving perspectives on 'human error'. Ergonomics 64, 1091–1114.

Reeson, M., Kyeremanteng, K., D'Egidio, G., 2018. Defibrillator design and usability may Be impeding timely defibrillation. Joint Comm. J. Qual. Patient Saf. 44, 536–544.
Robinson, M.N., Dharmage, S.C., Tang, M.L.K., 2014. Comparison of adrenaline auto-injector devices: ease of use and ability to recall use. Pediatr. Allergy Immunol. 25, 462–467.
Say, L., Chou, D., Gemmill, A., Tunçalp, Ö., Moller, A.-B., Daniels, J., Gülmezoglu, A.M., Temmerman, M., Alkema, L., 2014. Global causes of maternal death: a WHO systematic analysis. Lancet Global Health 2, e323–e333.
Schima, H., Schlöglhofer, T., zu Dohna, R., Drews, T., Morshuis, M., Roefe, D., Schmitto, J.D., Strüber, M., Zimpfer, D., 2014. Usability of ventricular assist devices in daily experience: a multicenter study. Artif. Organs 38, 751–760.
Surma-aho, A., Katja, H., 2021. Usability Issues in the Operating Room – towards Contextual Design Guidelines for Medical Device Design, vol. 90.
Tase, A., Vadhwana, B., Buckle, P., Hanna, G.B., 2022. Usability challenges in the use of medical devices in the home environment: a systematic review of literature. Appl. Ergon. 103.
Umasunthar, T., Procktor, A., Hodes, M., Smith, J.G., Gore, C., Cox, H.E., Marrs, T., Hanna, H., Phillips, K., Pinto, C., Turner, P.J., Warner, J.O., Boyle, R.J., 2015. Patients' ability to treat anaphylaxis using adrenaline autoinjectors: a randomized controlled trial. Allergy Eur. J. Allergy Clin. Immunol. 70, 855–863.
United Nations. UN sustainable development goals. https://www.un.org/sustainabledevelopment/.
Valdez, R.S., McGuire, K.M., Rivera, A.J., 2017. Qualitative ergonomics/human factors research in health care : current state and future directions. Appl. Ergon. 62, 43–71.
Virzi, R.A., 1992. Refining the test phase of usability evaluation: how many subjects is enough? Hum. Factors 34, 457–468.
WHO, 2012. WHO Recommendations for the Prevention and Treatment of Postpartum Haemorrhage.
WHO, 2017. WHO Recommendation on Tranexamic Acid for the Treatment of Postpartum Haemorrhage.
WHO, 2018. WHO Recommendations: Uterotonics for the Prevention of Postpartum Haemorrhage.
WHO, 2021. WHO Recommendation: Uterine Balloon Tamponade for the Treatment of Postpartum Haemorrhage.
WHO, 2023. Trends in Maternal Mortality 2000 to 2020 Estimates by WHO, UNICEF, UNFPA. World Bank Group and UNDESA/Population Division.