Health Informatics and Biomedical Engineering Applications, Vol. 142, 2024, 1–11
https://doi.org/10.54941/ahfe1005065

From Concept to Context: Evaluating Medical Device Usability Where It Matters Most

Adrián Morales-Casas[1], Amparo López-Vicente[1], Lorenzo Solano-García[2], and José Laparra-Hernández[1]

[1]Instituto de Biomecánica (IBV), Universitat Politècnica de València, Camino de Vera s/n, edif. 9C, 46022, Valencia, Spain
[2]Institute of Design and Manufacturing, Universitat Politècnica de València, Valencia, 46022, Spain

ABSTRACT

The usability validation process of medical devices outside controlled environments such as test facilities, laboratories, or by expert groups is vital to scrutinising the viability of the developed solution. This work outlines a case study in which the Spanish emergency service 061-Andalucia took part in the validation process of a non-contact vital sign measuring device through image processing, describing the methodology, participant sample, data analysis and conclusions. The measured vital signs were heart rate, respiratory rate, oxygen saturation, temperature, and blood pressure contactless at 2 meters (6.5 ft). In the study, three emergency service teams from three different operation bases in Malaga (Spain) underwent the validation process under semi-real conditions. Each team was provided with one measurement device used during the work shift on patients who were not in a critical stage, conscious and willing to participate in the study after being informed and signing a consent form. The primary goals of the validation were to analyse the ease of the process, reliability, and robustness of the measurements against the standard measurement equipment of the emergency service in different scenarios, as well as detect errors and limitations under semi-real conditions of use. Besides providing evidence of a potential improvement in the service through this new camera system, the satisfaction of the users/ patient and reducing equipment weight. Under these harsh conditions, the measurement device with a technical readiness level 7 reached reliability and robustness between 70% and 100%, depending on the measured vital signs and a high acceptance among the professionals of 66,66%.

Keywords: Project management, Human factors, User experience, Formative evaluation, UCD, Medical device, Non-contact measuring, Vital signs, Field operational tests, Clinical validation

INTRODUCTION

Usability testing is a practice that evaluates the user's performance and acceptance of a system or a product. The literature traces the first tests to the 1980s (Wichansky, 2000). This decade was also defined by fast-developing ground technologies such as the introduction of liquid crystal displays, and the advancements in infrared, radio frequency, and other technology

exponentially expanded the capabilities of cell phones, personal digital assistants (PDA) and home computers (Lewis, 2012). Through these technological enhancements, usability and user-centred design (UCD) rapidly increased in acknowledgement of the necessity to integrate humans in technology's fast-paced development (Nielsen, 1992) and generate a commune format for reporting tests and its results (UNE-EN ISO 9241-11, 2018; The IUSR Project, no date).

Since then, significant progress has been made, and the health industry has extensively adopted UCD for safety reasons (Branaghan, 2018). It is commonly accepted that medical device usability is influenced mainly by the user's behaviour, capabilities and limitations (Hegde, 2013; Knisely et al., 2020). Therefore, US and European regulatory entities have established regulations to identify, understand, and address use-related hazards (UNE-EN 60601-1-6, 2010; FDA, 2016). Nonetheless, the vast extension of medical device applications makes practical implementation of UCD approaches and deployment of tests through formative usability validation in relevant environments demanding (Bitkina, Kim and Park, 2020; Roma and de Vilhena Garcia, 2020). Moreover, access from medical device developers to (I) user groups, (II) acknowledgement by users and stakeholders of their impact on the development, (III) interaction issues and (IV) lack of incentives in the form of compensations are other reported pitfalls on the deployment of UCD approach to overcome (van Berkel et al., 2020).

Despite the considerable academic exploration into user engagement, actionable advice on effectively integrating UCD principles into development processes still needs to be noticed. To address this disparity, management, strategy and establishing a human factors plan are crucial in these domains. While larger companies like Siemens and Phillips may boast exemplary UCD practices, smaller enterprises, particularly Small and Medium Enterprises (SMEs), need help accessing expertise in this domain (Lukiyanto, Pratama and Ningrum, 2023). The present work aims to bridge the gap between the growing acknowledgement of the advantages of UCD in medical device development and the insufficiency of practical guidance accessible to developers for implementing such methods with a focal point on deploying a formative evaluation for a technical readiness level (TRL) 7 required use scenario, democratise the access of technologies by first responders to achieve a humanised tech devices and enhance society's acceptance towards new technology.

BACKGROUND

The formative evaluation was conducted within the framework and context of the EQUILIN project focused on creating a TRL 7 prototype (Casas et al., 2023). This prototype aimed to measure vital signs, including heart rate, respiratory rate, temperature, oxygen saturation, and blood pressure, using real-time image analysis based on non-contact photoplethysmography (PPG), see Figure 1.

This project assembled a diverse team of eight specialists, each fulfilling roles aligned with the scrum team structure in Agile project development.

Within this Agile framework, the team worked for 16 months to incorporate Human Factor Engineering (HFE) and Design for Usability (DfU) methodologies into the development process aligned with the Medical Device Regulation (MDR). HFE and DfU focus on ensuring that products are designed with the end-user in mind, optimising usability, safety, and efficiency. The team embraced a dynamic approach to continuous improvement throughout the project, implementing strategic checkpoints at three sprints. These strategic checkpoints allowed for regular assessment and adjustment of the prototype, contributing to its iterative enhancement.

Figure 1: EQUILIN TRL 7 prototype: remote PPG device and dashboard user interface.

The proactive utilisation of DfU was particularly crucial in this context. The team identified and addressed potential design flaws and concerns before reaching the formative validation phase by integrating usability considerations early in the design process. This proactive approach likely saved time and design and development costs by mitigating issues early on and ensuring the prototype was better aligned with user needs and expectations.

However, precise planning had to be implemented to get Malaga's ethics provincial investigation committee (H.R.U.–Málaga, no date) (Spain) approval to deploy the testing. The acquisition processes of the ethical approval needed careful planning of the user roles, definition of the test environment and scenarios to recreate, participant training, patient involvement level and exclusion criteria. Besides, the software and hardware involved in the study and the data analysis protocols were outlined appropriately, ensuring sufficient care on safety and data protection policies was given.

METHODOLOGY

Once formal approval was given, the deployment of the formative usability testing in practice followed the scheme illustrated in Figure 2, which, starting from the bottom, follows a back-to-front end deconstruction of the management blocks put in place. This structure was chosen to overcome the three main potential drawbacks that can occur during the deployment of a formative evaluation in a semi-real environment. The drawbacks are classified into three categories from the back-to-front end: technical, experimental design protocol/procedure-related and human, see Figure 2.

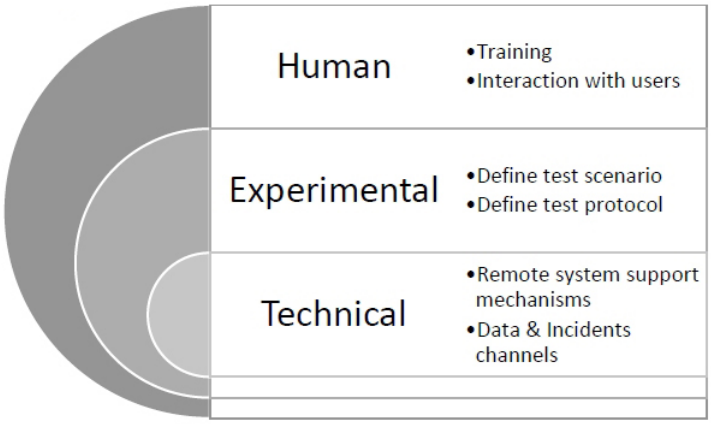


Figure 2: Methodology scheme classified by drawbacks categories.

In the first category, technical issues, it was necessary to ensure the capability of collecting data and reporting incidents. This was done by implementing online communication channels accessible by the users on their phones and endowing the prototype with a separate remote monitoring and access modem. Enabling remote real-time technical support was essential to maintain the continuity of the formative validation deployment, which spanned nine and a half weeks. Additionally, to provide backup options, a total of four camera devices were distributed. One device was allocated to each of the three emergency service teams stationed at separate bases in Malaga, while the fourth device was retained in the coordination office. This arrangement ensured contingency measures in case the three deployed devices encountered technical issues that could not be resolved remotely.

Second, the experimental design, the use case and scenarios, the participants (including patients and clinicians, more on this in the sample description) and the hypothesis were thoroughly defined and agreed on through the experimental protocol. This middle layer is a connection point between the technical and human factors during testing. Last, the human elements, this block compressed the training of the emergency service staff and strategy to ensure a friendly and natural interaction.

Three primary topics were covered to optimise the training session for emergency service staff. Firstly, written and audio-visual instructions were provided for the maintenance and initial utilisation of the camera system. Second, agreeing with the emergency team members on the best distribution of roles for recruiting patients, measurements, and data registration as they imagine better integrating into the intervention process. Last, the available incidence report tools were introduced to the teams, and awareness of the importance of their feedback was given. Due to the work dynamics of the emergency service staff, some participants were unavailable at the first scheduled training session. Therefore, a second online training session was planned five weeks later to give the participants feedback on the results.

Sample Description

Of the 15 users who participated in the study, 12 fully completed the survey. All 12 users tested the camera device in scenario one, emergency service intervention outside the hospital. Four used the device in hospital triage scenarios, and there was a massive event, as shown in the Sample description in Figure 3. The participants in the test belonged to three emergency service teams on three different bases in Malaga. The patients who participated in the study were also listed and classified according to the scenario, adding to 126 patients.

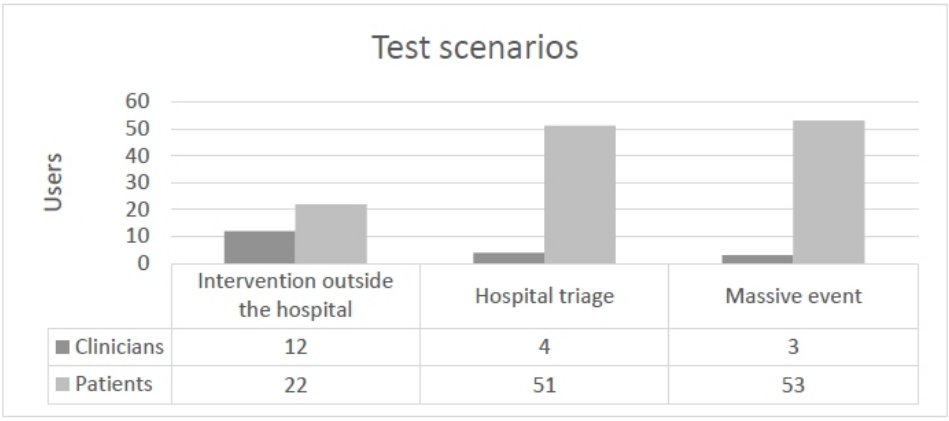
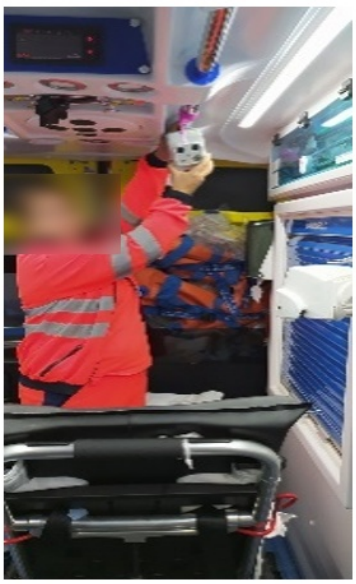


Figure 3: Sample description of users involved in each use case.

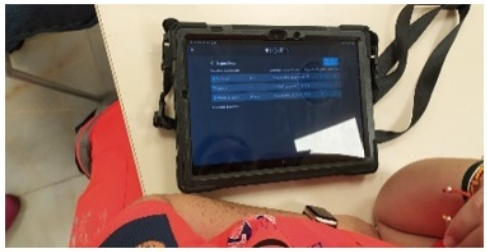
Experimental Plan and Data Analysis

The goals of the testing were, first, to prove the possibility of improving the emergency services by reducing the intervention time. Second, musculoskeletal injuries can be reduced by improving the ergonomics of the measurement equipment. Last, prove the feasibility of getting reliable and robust results in the three test scenarios with a contactless vitalising measurement technology based on the image.

In the experimental plan, three scenarios were described and tested, see Figure 4. The first scenario was emergency service interventions outside the hospital with an advanced coordination team formed by one paramedic, an emergency medical technician, and an emergency nurse. The aim was to measure the vital signs of 150 patients. Second, a triage scenario in a hospital to measure the vital signs of 50 patients. The third scenario was a massive event with a medical outpost for triage purposes and 50 patients. The plan also detailed the communication channels for incidents and the measurement protocol. Under this condition, 126 patients were eligible to participate in the study from the initial goal of 200 patients.



Scenario 1: Out-of-hospital



Scenario 2: Triage in a hospital

Scenario 3: Massive event

Figure 4: Three test scenarios.

The data analysis and validation of the camera system were done through a survey adapted from the technological acceptance model (TAM) (Davis, 1989) and system usability scale (SUS) (Martins et al., 2015) provided at the end of the testing. After the testing, two surveys were passed, one following the TAM model and another following the SUS model. On the other hand, data collection was done by reading the stored measurement values on

the camera device and the measurements of the reference current vital signs measurement equipment used by the emergency service collected in a data log. The traceability of the stored and manually collected measurements was completed by generating a numerical ID and registering the time stamp.

The measurements were classified into four categories throughout the testing. The first category was No ID; these were those measurements due to transcription problems that could not be traced back to the reference measurement equipment. The second category was Quality 1 (Q1); these measurements were correctly traced but with recordings that lasted less than 30 seconds, with much movement and not well-centred. Third category Q2: These measurements needed to be better-centred but still well-traced and of sufficient length. The last category, Q3, included measurements with optimal quality. It is acknowledged that many measurements were inaccurately recorded because of the system's orientation and distance from the patient.

DISCUSSION

The case study illustrated three complex test scenarios in which a novel device (using real-time image analysis based on non-contact PPG) for measuring the five main vital signs without contact was tested together with an APP for nine and a half weeks. The APP function was to visualise the real-time pre-processed data of the camera device, create the patient file and establish the communication between the device and the dummy server (emulating the health clear patient case server).

Although sufficient care was given to the strategic planning of such deployment to address technical issues, detail out the experimental part and promote and prepare the users' engagement, real-world scenarios can be unpredictable. A popular explanation of the acceptable results achieved in deploying a formative evaluation is that prior development, testing iteration, and thorough planning and preparation will help ensure consistent results. Nonetheless, the unpredictable nature of these conditions is necessary to identify unforeseen risks, difficult to reproduce in labs or semi-real conditions.

Participation and inclusiveness of patients in such studies are also complex. In the described case study, three emergency service bases in Malaga ensured the involvement of 15 users. As mentioned earlier, three users must complete the final survey fully. This detail is highlighted as all three interrupted the survey, abruptly raising the suspicion of a possible call of duty during the fulfilment of the study. Also, by having such a small sample, losing 20% has a high impact on the show results and a loss of more profound insights. On the other hand, the reduced participation of patients, particularly in scenario 1, was mainly due to exclusion criteria. Therefore, with a related nature.

However, we acknowledge considerable discussions among researchers about other uncommunicated aspects, such as skills to foresee potential risks in complex problem scenarios or the ability to adapt to changing events. Here, we highlight two values that perhaps help minimise the impact of unforeseen eventualities: the collaboration level with the client and the response-ability of the work team assigned to the project. In these projects, the team worked

as a scrum team; therefore, its values and principles were aligned with the scrum method, which showed a positive outcome.

RESULTS

Technical Acceptance Model

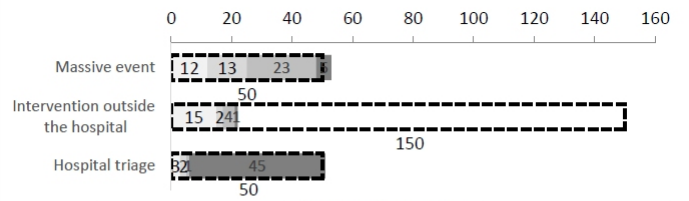
The TAM model survey revealed that the system provided the following perceived advantages related to the improvement of the service: reducing intervention times, assessing risk without the need for contact, reducing the risk of infectious diseases, monitoring several patients alternately without having to connect/disconnect devices, reduction of the expense associated with consumables and the corresponding environmental impact. In the same survey, the elements that are most highly rated about the acceptance of the prototype were: "easy-to-interpret visual information (55,6%)", "convenient to carry (66,7%)", and "necessary for the healthcare of the future (55,6%)".

System Usability Scale

The lack of real-time feedback on the quality or aiming orientation of the camera device in specific usage contexts posed significant challenges, see **Figure 5.** This situation may have influenced the SUS assessment of speed (22,2% positive), simplicity (66,7% positive), and convenience (44,4% positive) since the system stopped recording data by not correctly detecting the patient and did not generate any measurements, or the values were abnormal.

Robustness

The Q3 section measurements were used, as seen in **Figure 5**, to compare with the standard contact vital sign measurement equipment used in the ambulances of the Malaga emergency service and the context of hospital triage. A dashed line represents the target sample of each scenario, as shown in the table in **Figure 5**.



	Hospital triage	Intervention outside the hospital	Massive event
No ID	3	15	12
Q1	2	2	13
Q2	1	4	23
Q3	45	1	5
Target	50	150	50

Figure 5: Reliability and robustness results.

The measurements could be traced to the patient through the automatically generated IDs and the registered protocol for the standard measurement equipment. The results are collected in Figure 6, and the used reliability ranges were ±2 bpm for heart rate, ±2 RR for respiratory rate, ±2% for oxygen saturation, ±10 mmHg for blood pressure and ±0.5 °C for temperature.

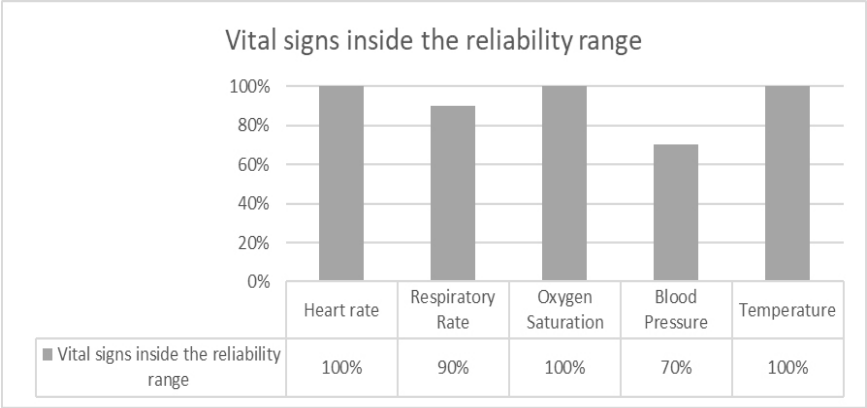


Figure 6: The reliability range of our non-contact device's critical vital signals to the current EPES equipment is used as the gold standard.

CONCLUSION

The testing phase involved the active involvement of management and developers in implementing the infrastructure and protocols necessary to prevent and enable the possibilities of real-time technical assistance during the test.

During the testing, the camera software was updated two times and the APP four times; on average, the incident resolution response time was four days, and none of the incidents required the pull out of the hardware devices during the ongoing testing.

Therefore, the incidents and flaws detected during the deployment could be classified as minor. Besides, the feasibility of getting reliable and robust results in the three test scenarios could be enhanced with further work. However, the tested device has proven to maximise service and ergonomic comfort by reducing weight and improving simplicity.

To sum up, the developed remote PPG and the ease to use interface can considerably improve current emergency services, reducing time and costs, and minimizing clinicians' physical risks.

As future work, this technology is being explored in other social and health care services from triage at hospitals to home care monitoring.

ACKNOWLEDGMENT

The authors would like to acknowledge the European Regional Development Fund (ERDF) through the Pluri-regional Operational Program of Spain

(POPE) 2014–2020, aid granted by the Ministry of Science and Innovation through the FID program "Promotion of Innovation from Demand".

The coordination work of the 061 Health Emergencies Centre of the Andalusian Health Service (in Spain) provided access to their facilities and inside knowledge from the professionals of the emergency service workers.

Also, we thank our colleagues from Plux Pedro Duque and Goçalo Telo. Lastly, I thank our colleagues from IBV, Úrsula Martinez, Juanma Belda, Daniel Gómez, Cristina Díaz, David Rubio, David Moro, Carlos Chirivella, Amparo López and José Laparra, who provided insight and expertise that greatly assisted the research and development.

REFERENCES

van Berkel, N. *et al.* (2020) 'Dimensions of ecological validity for usability evaluations in clinical settings', *Journal of Biomedical Informatics*, 110, p. 103553. Available at: https://doi.org/10.1016/J. JBI.2020.103553.

Bitkina, O. V., Kim, H. K. and Park, J. (2020) 'Usability and user experience of medical devices: An overview of the current state, analysis methodologies, and future challenges', *International Journal of Industrial Ergonomics*, 76, p. 102932. Available at: https://doi.org/10.1016/J. ERGON.2020.102932.

Branaghan, R. J. (2018) 'Human Factors in Medical Device Design: Methods, Principles, and Guidelines', *Critical Care Nursing Clinics of North America*, 30(2), pp. 225–236. Available at: https://doi.org/10.1016/J. CNC.2018.02.005.

Casas, A. M. *et al.* (2023) '3 sprints from zero to innovative medical device in 16 months: Benefits of combining Human Factors and Agile', *Production Management and Process Control*, 104(104). Available at: https://doi.org/10.54941/AH FE1003047.

Davis, F. D. (1989) 'Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology', *MIS Quarterly*, 13(3), p. 319. Available at: https: //doi.org/10.2307/249008.

FDA (2016) 'Applying Human Factors and Usability Engineering to Medical Devices', *FDA Guidance*, (301), pp. 1601–1605.

Hegde, V. (2013) 'Role of human factors / usability engineering in medical device design', *Proceedings - Annual Reliability and Maintainability Symposium* [Preprint]. Available at: https://doi.org/10.1109/RAMS.2013.6517650.

H. R. U. - Málaga (no date). Available at: http://www.hospitalregionaldemalaga.es/ Profesionales/Comit%C3%A9deEticaAsistencialdeM%C3%A1laga.aspx (Accessed: 26 March 2024).

Knisely, B. M. *et al.* (2020) 'An Analysis of FDA Adverse Event Reporting Data for Trends in Medical Device Use Error', *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, 9(1), pp. 130–134. Available at: https://doi.org/10.1177/2327857920091024.

Lewis, J. R. (2012) 'Usability Testing', *Handbook of Human Factors and Ergonomics: Fourth Edition*, pp. 1267–1312. Available at: https://doi.org/10.1002/9781118131350. CH46.

Lukiyanto, K., Pratama, A. R. F. and Ningrum, I. K. (2023) 'The Challenges of Applying Ergonomics to Small Medium Enterprises', *E3S Web of Conferences*, 388. Available at: https://doi.org/10.1051/E3SCONF/202338801012.

Martins, A. I. *et al.* (2015) 'European Portuguese Validation of the System Usability Scale (SUS)', *Procedia Computer Science*, 67, pp. 293–300. Available at: https://doi.org/10.1016/J. PROCS.2015.09.273.

Nielsen, J. (1992) 'The Usability Engineering Life Cycle', *Computer*, 25(3), pp. 12–22. Available at: https://doi.org/10.1109/2.121503.

Roma, M. S. G. and de Vilhena Garcia, E. (2020) 'Medical device usability: literature review, current status, and challenges', *Research on Biomedical Engineering*, 36(2), pp. 163–170. Available at: https://doi.org/10.1007/S42600-019-00037-8/FIGURES/4.

The IUSR Project (no date). Available at: https://www.nist.gov/itl/iad/iusr-project-industry-usability-report (Accessed: 14 March 2024).

UNE-EN 60601-1-6 (2010). Available at: https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0046353 (Accessed: 14 March 2024).

UNE-EN ISO 9241-11 (2018). Available at: https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma?c=N0060329 (Accessed: 14 March 2024).

Wichansky, A. M. (2000) 'Usability testing in 2000 and beyond', *Ergonomics*, 43(7), pp. 998–1006. Available at: https://doi.org/10.1080/001401300409170.