Optimization of Medical Device Clinical Evaluation Design Based on Analysis of Clinical Effectiveness and Usability

You Rim Kim

The Graduate School
Yonsei University
Department of Medical Device Engineering and Management

Optimization of Medical Device Clinical Evaluation Design Based on Analysis of Clinical Effectiveness and Usability

A Dissertation Submitted to the
Department of Medical Device Engineering and Management
and the Graduate School of Yonsei University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

You Rim Kim

December 2022

This certifies that the Doctoral Dissertation
of You Rim Kim is approved.

Thesis Supervisor: Sung Uk Kuh

Thesis Committee Member #1: Won Seuk Jang

Thesis Committee Member #2: Byeong Ju Kwon

Thesis Committee Member #3: Hee Kyo Jeong

Thesis Committee Member #4: Sung Bin Park

The Graduate School
Yonsei University
December 2022

ACKNOWLEDGEMENTS

의료기기산업학과에 입학하여 석사학위와 박사학위를 마치며 어느덧 5년의 시간이 흘렀습니다. 지난 시간동안 학위과정을 잘 마칠 수 있도록 도와주신 분들에게 감사의 마음을 전하려고 합니다. 가장 먼저 언제나 저를 신뢰하여 주시고 다양한 연구들을 할 수 있는 기회를 마련해 주신 장원석 교수님께 감사드립니다. 교수님의 지도를 받으며 학위과정 동안 많은 경험을 하고 조금씩 성장해 나갈 수 있었습니다. 또한, 관심과 애정으로 지도해주시고 심사해주신 구성욱 교수님께 진심으로 감사드립니다. 학위논문 지도 외에도 졸업 후에 무엇을 해야 하는지 조언해 주시고 격려해주신 권병주 교수님께도 짐심으로 감사드립니다. 항상 격려의 말씀으로 응원해 주신 정희교 교수님께도 감사드립니다. 교수님께서 건네 주시는 따뜻한 인사와 한마디 말씀이 언제나 큰 힘이 되었습니다. 또한, 석사학위에 이어 관심으로 심사해주시고 향후 연구에 대한 조언도 아낌없이 해 주신 박성빈 교수님께도 진심으로 감사드립니다. 5년동안 함께 했던 연구실의 모든 선배, 동기, 후배들에게도 덕분에 즐거운 대학원 생활을 할 수 있었다는 감사인사를 전합니다. 그리고 항상 옆에서 저를 믿고 응원해준 친구들과 사랑하는 가족들에게도 고맙다는 말을 전합니다. 학위기간 동안 많은 분들의 도움을 받아 학문적으로, 인간적으로 한단계 성장할 수 있었던 시간이었습니다. 앞으로도 계속 배움과 연구에 정진하여 의료기기산업에 기여할 수 있는 사람으로 성장해 나갈 수 있도록 노력하겠습니다. 감사합니다.

December 2022

You Rim Kim

CONTENTS

ABSTRACT ..... **viii**

I. INTRODUCTION **1**

1. Background 1

A. Clinical Evaluation 1

B. Clinical Effectiveness 2

C. Usability 5

D. Patient Monitor 9

2. Purpose 17

3. Method 17

II. MATERIALS AND METHODS **19**

1. Clinical Effectiveness Evaluation of Patient Monitor 19

A. Case of Patient Monitor Clinical Effectiveness Evaluation 19

2. Usability Evaluation of Patient Monitor 29

A. Use Scenarios and Critical Task Analysis 30

B. Usability Evaluation Indicators 37

3. Design of Patient Monitor Clinical Evaluation 43

A. Clinical Evaluation in an Actual Use Environment 44

B. Clinical Evaluation in a Simulated Environment 60

C. Result of Patient Monitor Clinical Evaluation 74

III. RESULT 94

1. Optimization of Clinical Evaluation Design 94

2. Clinical Evaluation Design Optimization Model Application Criteria 96

IV. DISCUSSION 99

V. CONCLUSION 102

REFERENCES 104

ABSTRACT (IN KOREAN) 109

LIST OF FIGURES

Figure 1. Medical Device Clinical Investigation Procedure 3

Figure 2. Medical Device Usability Test Procedure 7

Figure 3. Usability Test Procedure of Actual Medical Device Use 8

Figure 4. Number of Adverse Events Reported to the FDA by Year 12

Figure 5. Device Problems of Adverse Events Reported to FDA 13

Figure 6. Number of Adverse Events Reported to the MFDS by Year 15

Figure 7. Device Problems of Adverse Events reported to MFDS 16

Figure 8. Confusion Matrix 27

Figure 9. NASA Task Load Index 41

Figure 10. Clinical Evaluation Design of Patient Monitor 43

Figure 11. Investigational Devices 46

Figure 12. ECG 5 Lead Placements of Patient Monitor 47

Figure 13. ECG 5 Lead Placements of Investigational Devices 52

Figure 14. Four Areas Where Alarm Messages Can be Displayed on the Screen 59

Figure 15. Configuration of Test Environment 65

Figure 16. Visual Alarm Display of Patient Monitor A 69

Figure 17. Audible Alarm Stop Button for Each Parameter 69

Figure 18. Arrhythmia Alarm Setup Menu 70

Figure 19. Overall Audible Alarm Off Button..... 72

Figure 20. User Preference Survey Results for Display Setting and Alarm 84

Figure 21. User Preference Survey Results on the Need for Specific Functions..... 85

Figure 22. Results of a Preference Survey for Arrhythmia Alarm Activation Settings .. 85

Figure 23. Results of 23 Satisfaction Questionnaires of Participants 90

Figure 24. Satisfaction Result of Participants for Use Scenarios..... 91

Figure 25. Process of Clinical Evaluation Design Optimization Model 94

Figure 26. Clinical Evaluation Optimization Model Application Decision Flow Chart 98

LIST OF TABLES

Table 1. Definitions of The Patient Monitor in MFDS and FDA 9

Table 2. Noise in the Intensive Care Unit 11

Table 3. Reports of Adverse Events Related to Alarm, Use Error, and Human Factor 14

Table 4. Example of Evaluating the Clinical Effectiveness of Patient Monitor Alarms 20

Table 5. Similar Previous Studies on Patient Monitoring Alarm 23

Table 6. Design of Similar Previous Studies on Patient Monitoring Alarm 24

Table 7. Clinical Effectiveness Endpoint of Similar Studies on Patient Monitoring Alarm 26

Table 8. Use Scenario of the Patient Monitor 30

Table 9. Task of Patient Monitor Use Scenario 31

Table 10. Definitions of Severity Levels 33

Table 11. Severity Assessment of Use Scenarios and Tasks 34

Table 12. Results of Important Task Selection 35

Table 13. Task Success Level and Description 37

Table 14. Post-Study System Usability Questionnaire 39

Table 15. Description of NASA-TLX 42

Table 16. Clinical Evaluation Design Synopsis in an Actual Use Environment 44

Table 17. Arrhythmia Alarm of patient Monitor A and B 47

Table 18 Audible Alarm Pattern 48

Table 19. Alarm of Tachycardia and Bradycardia 48

Table 20. Specific Functions in Patient Monitor B 49

Table 21. Distribution of 20 Patients 50

Table 22. Characteristics of 39 Nurses 51

Table 23. Clinical Effectiveness Evaluation Procedure in an Actual Use Environment 53

Table 24. Usability Evaluation Procedure in an Actual Use Environment 54

Table 25. User Preference Survey for Patient Monitoring Devices 57

Table 26. Clinical Evaluation Design Synopsis in a Simulated Environment 60

Table 27. Function of Patient Monitoring Device 61

Table 28. Patient Monitor Display Specifications 62

Table 29. Characteristics of Usability Test Participants 63

Table 30. Test Room Environment 64

Table 31. Usability Evaluation Procedure in Simulated Environment 66

Table 32. Satisfaction Evaluation Questionnaire 67

Table 33. Action of the Alarm Pause Button 71

Table 34. Alarm Occurrence Conditions and Alarm Message Duration for Type of PVCs ... 71

Table 35. User Preference and Use Experience Survey 73

Table 36. Result of False Alarm Occurrence 74

Table 37. Descriptive Statistics of PSSUQ Scores 75

Table 38. Statistical Comparison of PSSUQ Scores 77

Table 39. Descriptive Statistics of TLX Scores 78

Table 40. Statistical Comparison of TLX Scores 79

Table 41. Results of the User Preference Survey 80

Table 42. Success Rate for Each Task and Scenario 86

Table 43. Satisfaction Result of Participants 89

Table 44. Result of User Preference and Use Experience Survey 93

ABSTRACT

Optimization of Medical Device Clinical Evaluation Design Based on Analysis of Clinical Effectiveness and Usability

You Rim Kim

Department of Medical Device Engineering and Management

The Graduate School, Yonsei University

(Directed by Professor Sung Uk Kuh, Won Seuk Jang)

Clinical evaluation is the evaluation of whether the clinical evidence is sufficient to confirm whether a device complies with the requirements for safety and performance. This is evaluated through clinical investigation and usability evaluation. Standards have been created and implemented to properly conduct clinical evaluations and prove the safety, effectiveness, and usability of medical devices. However, adverse events exist in the actual medical environment, and it shows that complete safety cannot be guaranteed even if clinical evaluations produce entirely positive results. Therefore, this study aims to propose a method for evaluating the effectiveness and usability of medical devices through an optimized clinical evaluation design and to present decision criteria for applying it to the clinical evaluation of medical devices.

To derive a clinical evaluation design optimization model, a case study was conducted on the patient monitoring device, one of the important components of the intensive care unit, to design and conduct clinical evaluation, and finally present the clinical evaluation design optimization model and application criteria. The proposed clinical evaluation design optimization model consists of clinical evaluation in the actual use environment and clinical evaluation in the simulated environment. Clinical evaluation in an actual use environment is a model that integrates clinical effectiveness evaluation and usability evaluation and proceeds simultaneously. While conducting the clinical investigation, usability evaluation is also conducted, and both clinical effectiveness endpoint and usability endpoint data are collected. Clinical evaluation in a simulated environment performs qualitative and quantitative evaluation through usability testing. To determine whether to apply the clinical evaluation optimization model when evaluating medical devices, it is necessary to consider whether the medical device is multifunctional and complicated to use, whether urgent decision-making and fast-paced use are required when using the device, whether the medical staff cares for multiple patients, whether there are multiple alarms, lots of noise, and flashing lights. The proposed clinical evaluation model can be used for the clinical evaluation of various medical devices used in environments such as operating rooms, various intensive care units, emergency rooms, and ambulances to consider key aspects of actual patient care or risk-related use. It is possible to identify errors and usability problems that occur during actual use. Therefore, it is expected that the evaluation of clinical effectiveness and usability can be performed more accurately.

Key words: Clinical Evaluation, Effectiveness, Usability, Medical Device, Patient Monitor

Optimization of Medical Device Clinical Evaluation Design Based on Analysis of Clinical Effectiveness and Usability

You Rim Kim

Department of Medical Device Engineering and Management

The Graduate School, Yonsei University

(Directed by Professor Sung Uk Kuh, Won Seuk Jang)

I. INTRODUCTION

1. Background

A. Clinical Evaluation

Clinical evaluation is a structured and ongoing procedure for collecting, evaluating and analyzing clinical data related to medical devices.[1] The purpose is to evaluate whether the available clinical evidence is sufficient to confirm that the device complies with the relevant essential requirements for safety and performance when used according to the manufacturer's instructions for use.[1] Evaluation includes confirmation that the device achieves the performance intended by the manufacturer and includes confirmation of

usability that the design adequately reduces the risk of use error, as far as possible, and is adequate for the intended user.[1]

Clinical performance is evaluated through clinical investigation as the behavior of a medical device or response of the subject to that medical device in relation to its intended use when applied correctly to appropriate subjects.[1] Usability means demonstrating that all risks associated with design and use of the device are minimized, residual risks are acceptable, and that the information material is suitable for use by its intended users.[2]

B. Clinical Effectiveness

(A) Definition of Clinical Effectiveness

Effectiveness refers to the ability of an intervention to have a meaningful effect on the patient under normal clinical conditions.[3] ISO 14155:2020 defines effectiveness as an achievement of a clinically significant intended result in a defined portion of the target population when the investigational medical device is used within its intended uses and according to its instructions for use, the investigator's brochure and the plan, as determined by documented scientific evidence.[4]

(B) Clinical Effectiveness Evaluation

To evaluate the effectiveness of a medical device, clinical trials are conducted that systematically investigate one or more human subjects.[4] A clinical trial is synonymous with clinical investigation.[4] Medical device clinical investigation shall be conducted in

accordance with good clinical practice. ISO14155:2020 is an international standard that addresses good clinical practice for the design, conduct, recording and reporting of clinical investigations conducted on human subjects to evaluate the clinical effectiveness or performance and safety of medical devices.[4] This international standard specifies general requirements for protecting the rights and safety of human subjects and ensuring the scientific conduct of clinical investigations and the reliability of the results of clinical investigations and defines the responsibilities of the sponsor and principal investigator.[4] According to ISO 14155 requirements, before starting a clinical investigation, the sponsor shall obtain an approval document or a favorable opinion document from the ethics board. An ethics committee is an independent body responsible for reviewing clinical investigations to protect the rights, safety, and well-being of human subjects participating in clinical investigations.[4] It is also called a research ethics committee or institutional review committee.[4] Then, Investigator shall obtain informed consent from all subjects before participating in clinical investigations.[4] Accordingly, general clinical trials proceed with the procedure shown in Figure 1.

Figure 1. Medical Device Clinical Investigation Procedure.

In order to conduct a clinical investigation, it is necessary to develop a clinical trial protocol first.[5] In the clinical trial protocol, detailed progress of the clinical investigation,

including the purpose of the investigation, medical device information, subject inclusion/exclusion criteria, and investigation period, shall be written.[6]

When a clinical trial protocol is developed, it is reviewed and approved by the Institutional Review Board of each institution(hospital) where the clinical investigation will be conducted.[5] Prior to the start of a clinical investigation, training on clinical trial protocols, procedures, and the use of medical devices is conducted.[6] After that, the subjects participate in the investigation by filling out the consent form, and the investigator applies the medical device to the subject according to the detailed progress of the clinical trial. It is finished by collecting endpoint data to evaluate clinical effectiveness from subjects, analyzing the collected data, and writing a result report.[6] Participants then use medical devices while performing tasks, and observational and subjective data are collected. Participants then use medical devices while performing tasks, and observational and subjective data are collected.

C. Usability

(A) Definition of Usability

Usability refers to the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a particular context of use.[7] ISO 62366-1:2015, a standard for the application of usability engineering to medical devices, defines usability as a characteristic of the user interface that facilitates use and establishes effectiveness, efficiency, and user satisfaction in the intended use environment.[8] Effectiveness is the accuracy and completeness with which a user achieves specified goals, which is a different concept than 'clinical effectiveness' in medical device clinical investigation.[7,8] Efficiency is the resources used in relation to the result achieved, equal to the resources expended in relation to effectiveness. Resources typically include time, human effort, cost, and materials.[7,8] Satisfaction is the extent to which a user's physical, cognitive, and emotional responses to the use of a system, product, or service meet the user's needs and expectations.[7]

(B) Usability Evaluation

Usability can be evaluated through formative and summative evaluation. Formative evaluation is a user interface evaluation conducted with the intention of exploring the strengths, weaknesses and unanticipated use errors of the user interface design.[8] Formative evaluation is performed iteratively throughout the design and development process on a

small scale as an effective way to discover design flaws in user interfaces that can lead to potentially dangerous use errors.[9]

Summative evaluation is a user interface evaluation performed at the end of user interface development to obtain objective evidence that the user interface can be used safely.[8] Summative evaluation is used to confirm the safety of the user interface, and through summative evaluation, it should be determined whether the user interface is acceptable in terms of use-related risks and effects.[9] Usability evaluation can be performed by applying various Usability methods such as advisory panel reviews, heuristic analysis, usability tests, surveys, and Workload assessment. Summative evaluation generally involves usability testing under simulated use conditions.[10]

Usability testing is a method of exploring or evaluating a user interface with intended users within a specified intended use environment.[8] Usability testing involves recruiting users from a specific user group and asking those users to complete a series of tasks, during the implementation, user observation is performed.[10] As required in IEC 623661:2015, Usability testing should be planned to provide the data needed to evaluate the user interface and should be documented in the form of a protocol that describes the purpose of the test and the methods to be used.[10] Usability testing may include training as part of the protocol, as appropriate to simulate realistic use.[10] Accordingly, general usability test proceed with the procedure shown in Figure 2.

Figure 2. Medical Device Usability Test Procedure.

The protocol includes a description of the usability test purpose, test environment and conditions of use, inclusion/exclusion criteria, test scenarios, test equipment, and a training plan.[11] The test environment should be configured in advance, as testing must be conducted in an environment that is reasonably representative of the intended use environment in terms of workspace arrangement, workspace equipment and furnishings, noise levels, and light levels. Mannequins and simulators can be used as simulated patients in a test environment.[12] Participants complete a training session prior to participating in a testing session.[12] The training conducted prior to the test should be identical to the actual training.[11] You may choose not to provide training to all or some of the participants in order to see how well they can use it without training.[12] All participants read and sign an informed consent form before participating in usability testing.[12] Trained participants sign the consent form prior to receiving training, and untrained participants sign it prior to the testing session.[12] Test participants perform use scenarios and use medical devices during the test, and their performance on each task is observed and recorded while performing each use scenario.[10] Observational data and subjective data are collected,[10] and the usability test procedure is completed by analyzing the collected data and writing a report.

In some cases, additional user interface evaluation may be required under actual use

conditions, as simulated use may not be adequate to explore some hazard-related use scenarios in usability testing.[8] figure3 shows the usability test procedure of actual medical device use. IRB approval was added to the procedure as research involving human subjects, such as patients and medical staff, may require approval from an institutional review board or ethics committee.[10]

Figure 3. Usability Test Procedure of Actual Medical Device Use.

In usability tests for actual medical device use, we can evaluate interactions of interest that cannot be fully evaluated through simulated use tests while using medical devices in use environments such as operating rooms, intensive care units, emergency rooms, and ambulances.[10] In addition, objective and subjective data such as descriptions of use errors, performance, opinions, and interview responses interview questions can be collected by conducting interviews at the end of a workday or upon TASK completion.[10]

D. Patient Monitor

A patient monitor is a device that monitors a patient's condition by measuring the patient's biometric data in a variety of patient care settings.[13] This device measures the patient's biological data including invasive blood pressure, electroencephalogram (EEG), electrocardiogram (ECG), carbon dioxide (CO_2) gas measurements and transcutaneous oxygen saturation (SPO_2) and provides visible and audible alarms when hazardous conditions are detected.[14] Table 1 shows the patient monitor medical device classification and definition in the MFDS and FDA.

Table 1. Definitions of The Patient Monitor in MFDS and FDA

Category	Product Code	Product Name	Definition	Class
MFDS	A26090.01	Patient monitor	A device used to monitor vital signs of patients. Visible and audible alarms are sounded when hazardous conditions are observed.	2
MFDS	A26090.02	Patient monitoring system, transportable	A device used to monitor biological data of the patient, which combines data from invasive blood pressure, electroencephalo gram (EEG), and CO_2 gas measurement modules. Visible and audible alarms are sounded when hazardous conditions are observed.	2
FDA	MHX (870.1025)	Monitor, Physiological, Patient (With Arrhythmia Detection or Alarms)	The arrhythmia detector and alarm device monitors an electrocardiogram and is designed to produce a visible or audible signal or alarm when atrial or ventricular arrhythmia, such as premature contraction or ventricular fibrillation, occurs.	2

(A) Patient Monitor in Intensive Care Unit

The patient monitor is commonly used in hospitals, especially in critical care settings such as intensive care units, and other use environments including general hospital rooms and operating rooms. In an intensive care unit environment where patient monitoring devices are used as an important component,[13] medical personnel care for 6 to 12 patients at a time, constantly monitoring vast data streams as patients generate multiple vital sign parameters simultaneously.[15] The physical environment consists of various alarm noises, flashing lights, and machines working without a pause.[15] Table 2 shows the noise sources and intensities in the intensive care unit. The noise level in the intensive care unit is between 50 and 80 dB, and the noise level remains above 60 dB for at least 6 hours a day, and above 70 dB occurs intermittently.[16] Noise in the ICU is caused not only by equipment but also by Nursing care of patients, patient noise, staff talking, and Physician discussion. These environmental factors increase the potential for error and may contribute to burnout for clinicians, negative patient outcomes, and safety issues.[15,17,18]

86% of alarms in the intensive care unit are false alarms, and between 6% and 40% are clinically insignificant alarms that do not require immediate action.[19] Between 2% and 9% of alarms were found to be important to patient care.[20] False alarms False alarms in the intensive care unit can affect both patients and clinical staff through noise disturbances, desensitization to warnings, and delayed response times, resulting in disruption of care or decreased quality of care.[21]

Table 2. Noise in the Intensive Care Unit[16]

Category	Source of Noise	Intensity dB
Equipment	Cardiac monitor	61
	Cardiac monitor alarm	60 - 78
	Ventilator	60 - 65
	Ventilator alarms	71 - 76
	Dialysis alarms	63
	Pulse oximeter	60 - 70
	Intra-aortic balloon pump	60 - 74
	Neonatal incubators	58
Nursing care of patient	O_2 via mask	50 - 60
	Oral/pharyngeal/tracheal suction	55 - 68
	Talking whilst taking vital signs	60 - 68
	Chest percussion/physiotherapy	83
	Patient transfer	60 - 66
	Moving bed	58
	Moving transfusion pole	65
	Moving trolley	65 - 70
Patient noise	Snoring	60
	Coughing	70 - 76
	crying	80
Environmental noise	Staff talking	63 - 70
	Physician discussion	68 - 75
	Telephone ringing	60 - 65
	Computer printer	70 - 72
	Squeaky chair	76
	bedpan in washer/sluice	70 - 80

(B) Adverse Event of Patient Monitor

(1) Patient Monitor Adverse Events reported to FDA

The number of FDA medical device adverse event reports of patient monitoring devices is a total of 10,082 cases from 2017 to 2022, showing an increasing trend except for 2022. Figure 4 shows the number of adverse events reported to FDA by year.

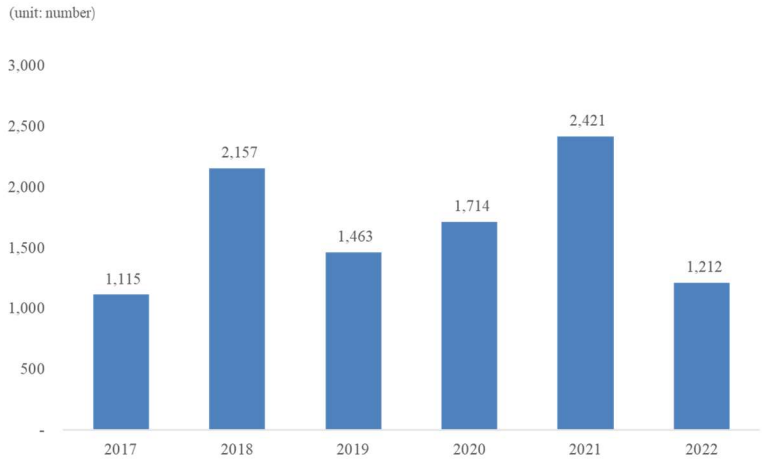


Figure 4. Number of Adverse Events Reported to the FDA by Year.

There are 100 types of device problems with patient monitor adverse events reported to the FDA. Figure 5 shows the percentage of reported adverse events by type of device problem. As for device problems, no audible alarm is the most common with 1,129 cases, followed by communication or transmission problem and device alarm system with 1,044 and 1,040, respectively. In addition, there are device problems such as output problem, device operates differently than expected, display or visual feedback problem, device

displays incorrect message, no display/image, no audible prompt/feedback, and defective alarm. In the event description of the adverse events for each device problem, there are many adverse events related to the alarm problem even if the device problem was different.

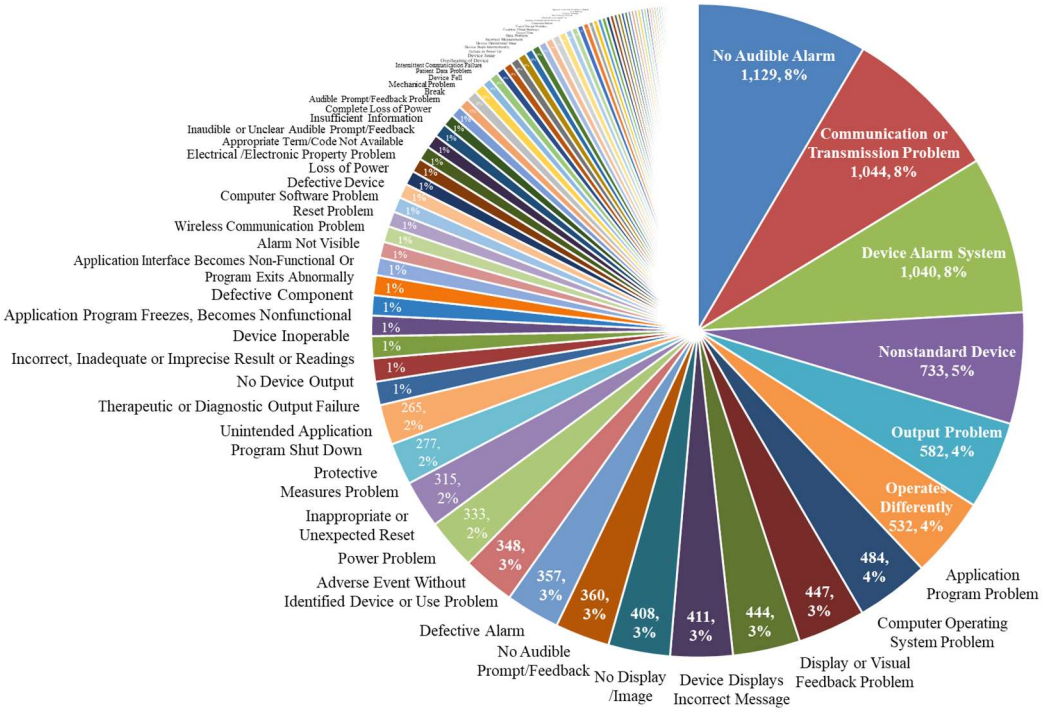


Figure 5. Device Problems of Adverse Events Reported to FDA.

The number of patient monitoring system adverse events that included the word 'alarm' in the device problem was 2,690 in total, which was about 20% of the total adverse events. The number of side effects related to use errors and human factors was 63 and 22, respectively. Table 3 shows the number of adverse event reports related to alarm, use errors and human factors. The reported adverse events include the device triggering a vtach/vfib

alarm when it should have been an afib alarm, and misinterpreting an asystole alarm as a ventricular stroke. In addition, due to frequent false apnea alarms, the nursing team reacted late in the real red alarm situation and the patient died. Other adverse events are that the device did not sound an alarm because the user shutting-off the alarms, or that the alarm pause setting was changed to infinite.

Table 3. Reports of Adverse Events Related to Alarm, Use Error, and Human Factor

Category	Device Problems	Adverse Event
Alarm	No Audible Alarm	1,129
	Device Alarm System	1,040
	Defective Alarm	357
	Alarm Not Visible	105
	Delayed Alarm	22
	Improper Alarm	16
	False Alarm	12
	Low Audible Alarm	9
Use Error	Use of Device Problem	42
	Use of Incorrect Control/Treatment Settings	21
Human Factor	Human Factors Issue	11
	Human-Device Interface Problem	11

(2) Patient Monitor Adverse Events reported to MFDS

As of September 2022, the number of adverse events of patient monitoring devices reported to the Ministry of Food and Drug Safety is 54, showing an increasing trend. Figure 6 shows the number of adverse events reported to MFDS by year.

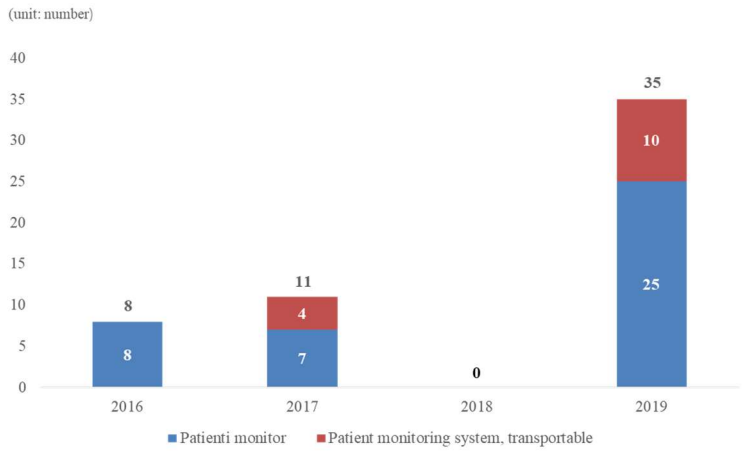


Figure 6. Number of Adverse Events Reported to the MFDS by Year.

There are 30 types of device problems with monitor side effects reported in the MFDS. Figure 7 shows the percentage of reported adverse events by type of device problem. As for device problems, No Device Output and Incorrect Measurement were the most frequent with 4 cases each. Failure to Sense, Fracture, Cut in Material, and Failure to Power-Up were the next most frequent with 3 cases. There were two device problems such as Device Alarm System Issue and Device Operates Differently than Expected. Incorrect Display, No Device Output, and Imperfect Device Output each occurred one case.

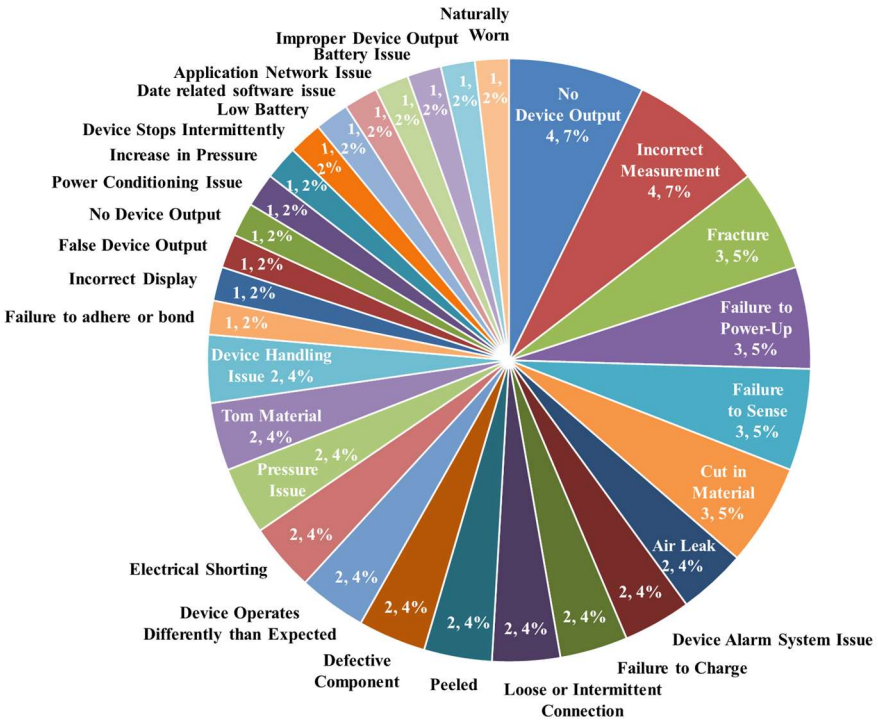


Figure 7. Device Problems of Adverse Events Reported to MFDS.

Reported adverse events include an inaccurate measurement of heart rate value when measuring an electrocardiogram on a patient monitor, or a sudden failure to measure while the monitor SPO2 sensor is in use. Other adverse events include cases in which the alarm does not sound due to a problem with the speaker software, and there are cases in which the mode is set to pediatric rather than adult and the measurement is not performed when an adult cuff is connected and used.

2. Purpose

Standards have been created and implemented to verify the safety, effectiveness, and performance of medical devices and to reduce use errors by improving usability, but adverse events due to problems with medical devices still exist in the actual medical environment. This is because unexpected use errors occur in real medical environments depending on various patients, users, and use environments, and noise is generated for various reasons such as patient movement, nursing treatment.[20] In some cases, even if clinical evaluations generate positive results, manufacturers cannot simulate all hazard-related use scenarios associated with medical devices, and clinical investigations are conducted under controlled conditions, which can lead to assurance gaps.[10]

Therefore, through this study, I would like to propose a method to verify the effectiveness and usability of medical by optimizing the design of clinical evaluation based on clinical effectiveness and usability evaluation. In addition, a flow chart for determining the medical devices to which the clinical evaluation optimization design model can be applied is presented so that the model can be used to evaluate the effectiveness and usability of medical devices.

3. Method

In order to propose a clinical evaluation design optimization model through case studies of patient monitors, one of the important components of an intensive care unit, clinical

trials related to clinical effectiveness evaluation of patient monitors are investigated through Clinicaltrial.gov, and MEDILINE database. previous clinical Investigation cases of the patient monitor are analyzed to derive clinical effectiveness evaluation design and clinical effectiveness endpoints.

For patient monitor case studies to derive a clinical evaluation design optimization model, use scenarios and critical tasks of patient monitors are analyzed and usability evaluation indicators were investigated. Based on the patient monitor's clinical investigation case analysis, clinical effectiveness endpoint investigation, use scenario and task analysis, and usability evaluation index investigation, clinical evaluations design in the actual and simulated environments of the patient monitor are derived. Conduct evaluation according to the derived patient monitor clinical evaluation design to confirm the applicability of the clinical evaluation method. Finally, a clinical evaluation design optimization model is proposed, and application criteria for medical devices to which the model can be applied are presented.

II. MATERIALS AND METHODS

1. Clinical Effectiveness Evaluation of Patient Monitor

As the heart is one of the commonest organs to fail during critical illness, Accurate assessment and monitoring of cardiac function in the intensive care unit is essential,[22] and ECG monitoring is considered the best way to detect abnormalities in the heart.[23] The patient monitor detects various types of arrhythmias through ECG monitoring and provides visual and audible alarms.

Alarms help prevent patient harm by quickly recognizing and responding to critical situations only if they are not false alarms.[24] False alarms in the Intensive Care Unit lead to desensitization of medical staff to alarms with the risk of critical situations potentially being ignored despite correct alarming, with associated slowing in response times and detrimental decreases in the quality of care for the patient.[21,25] Many alarms, as they now exist in most patient monitors, are not usually perceived as helpful by the medical staff due to the high incidence of false alarms.[25] Therefore, it seems possible to prove the effectiveness of the arrhythmia detection of the patient monitoring device through the evaluation of the occurrence of the false alarm of the arrhythmia alarm.

A. Case of Patient Monitor Clinical Effectiveness Evaluation

Studies evaluating the clinical effectiveness of patient monitor alarms was investigated through Clinicaltrial.gov, a database of clinical studies conducted in 50 states and 221

countries, and MEDILINE database developed and maintained by the NLM National Center for Biotechnology Information (NCBI).

As a result of the investigation, it was confirmed on Clinicaltrial.gov that clinical studies comparing the efficacy and safety in the Monitoring of ECG are being conducted. Table 4 shows an example design of a study evaluating the clinical effects of patient monitor alarms. The secondary outcome of the study was monitoring for 24 hours per subject and comparing the number of correct yellow and red alarms between the investigational device and the control device.

Table 4. Example of Evaluating the Clinical Effectiveness of Patient Monitor Alarms

Item	Description
Brief Title	CardioSenseSystem Compared Study Regarding Efficacy and Safety in the Monitoring of ECG
NCT Number	NCT03610529
Design Details	• Allocation: Non-Randomized • Intervention Model: Single Group Assignment • Intervention Model Description: The study participants will be using both the experimental device and the control device simultaneously throughout the study participation. The participant will thereby act as its own control. • Masking: None (Open Label) • Primary Purpose: Prevention
Intervention	Experimental device: CardioSenseSystem (a novel wireless ECG monitoring system) Control device: Philips Intellivue (an established ECG monitoring system)
Enrollment	60
Inclusion Criteria	• Male or female at least 18 years old. • Patient hospitalized at the investigational site and in need of ECG monitoring. • Patient with expected alarms during the 24 hours ECG monitoring. • Patient who has been informed of the clinical trials purpose, limitations, and relevance, and who has voluntarily agreed to participation in the clinical trial by signing the informed consent form.

Item	Description
Exclusion Criteria	• Patient with burns. • Patient with known allergy or sensitivity to any of the compositions in CardioPatch. • Patient with infection in the area where the electrodes are to be placed. • Patient with fragile skin (e.g., after prolonged cortisone treatment). • Patient with open sternum / sternum (e.g., severe heart failure postoperatively) or treatment for infection of the sternum. • Patient with mechanical auxiliary heart or ECMO. • Patient with implantable defibrillator. • Severely ill patient during end of life. • Patient participating in any other clinical trial. • Patient where the investigator judges that participation may be risky for the patient or obstruct or interfere the implementation of the trial as approved.
Primary Outcome	Loss of monitoring data: Compare time of interruptions in the monitoring system between the investigational device and the control device.
Secondary Outcome	• Management time: Compare management time between the investigational device and the control device. This is done by measuring the time required for the sterilization of cables, battery replacements, application of electrodes and cables, and extra management time for applying electrodes and cables if unconnected. • Number of correct yellow and red alarm: Compare the number of correct yellow and red alarm between the investigational device and the control device. • Number of false yellow and red alarm: compare the number of false yellow and red alarm for the investigational device and the control device. • Incidence and severity of Adverse Events: The incidence and severity of adverse events associated with the investigational device and the control device.

Table 5 shows similar prior studies on intensive care unit patient monitoring alarms. As a result of searching the MEDILINE database through PubMed, 7 similar previous studies on patient monitoring alarms in intensive care units were identified. Lawless suggested that 94% of all alarms in the pediatric intensive care unit (PICU) were not clinically relevant,[19] and in another study, Tsien and Fackler also found that 92% of alarms observed in the PICU were false alarms.[20]

These results are not specific to the PICU. O'Carroll reported that only 8 of 1,455 alarms were caused by potentially life-threatening situations and Siebig reported that only 17% of alarms were related and 44% were technically false.[26,27]

In one of the studies conducted in the intensive care unit, 26% of the alarms had marginal consequences, for example leading to re-positioning of sensors.[24,28] 24% were caused by manipulation, 17% of alarms were the result of technical problems, and only 6% did the alarm led to a call for a doctor.[24,28] There were studies on false alarm reduction as well. Chambrin reported that an average of 42.7% of ECG arrhythmia alarms were false, and that the suppression algorithm reduced the incidence of false ECG arrhythmia alarms to 22.7%. Muroi reclassified the alarms and found that the sensitivity of the red alarm was 87.0% and the specificity was 29.6%, while the sensitivity of the green alarm was 30.2% and the specificity was 87.2%.

Table 5. Similar Previous Studies on Patient Monitoring Alarm

Case Number	Study Title	Result
Case 1	Lawless ST. Crying wolf: false alarms in a pediatric intensive care unit. Crit Care Med. 1994 Jun;22(6):981-5. PMID: 8205831.	94% of all alarms in the pediatric intensive care unit (PICU) were clinically irrelevant.[19]
Case 2	Tsien CL, Fackler JC. Poor prognosis for existing monitors in the intensive care unit. Crit Care Med. 1997 Apr;25(4):614-9. doi: 10.1097/00003246-199704000-00010. PMID: 9142025.	92% of alarms are false alarms in their observation in a PICU.[20]
Case 3	O'Carroll TM. Survey of alarms in an intensive therapy unit. Anesthesia. 1986 Jul;41(7):742-4. doi: 10.1111/j.1365-2044.1986.tb12844.x. PMID: 3463228.	Only 8 of 1,455 alerts are caused by potentially life-threatening situation.[26]
Case 4	Siebig S, Kuhls S, Imhoff M, Langgartner J, Reng M, Schölmerich J, Gather U, Wrede CE. Collection of annotated data in a clinical validation study for alarm algorithms in intensive care--a methodologic framework. J Crit Care. 2010 Mar;25(1):128-35. doi: 10.1016/j.jcr c.2008.09.001. Epub 2009 Jan 17.	Only 17 % of the alarms were relevant, with 44 % being technically false.[27]
Case 5	Chambrin MC, Ravaux P, Calvelo-Aros D, Jaborska A, Chopin C, Boniface B. Multicentric study of monitoring alarms in the adult intensive care unit (ICU): a descriptive analysis. Intensive Care Med. 1999 Dec;25(12):1360-6. doi: 10.1007/s0013400510 82. PMID: 10660842.	26% of alarms had negligible consequences leading to sensor relocation, only 6% alarms called doctors, 17% were the result of technical problems and 24% were caused by staff manipulation.[28]
Case 6	Aboukhalil A, Nielsen L, Saeed M, Mark RG, Clifford GD. Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform. J Biomed Inform. 2008 Jun;41(3):442-51. doi: 10.1016/j.jbi.2008.03.003. Epub 2008 Mar 21. PMID: 18440873; PMCID: PMC2504518.	An average of 42.7% of ECG arrhythmia alarms are false, A suppression algorithm reduced the incidence of false ECG arrhythmia alarms from 42.7% to 22.7%.[21]
Case 7	Muroi C, Meier S, De Luca V, Mack DJ, Strässle C, Schwab P, Karlen W, Keller E. Automated False Alarm Reduction in a Real-Life Intensive Care Setting Using Motion Detection. Neurocrit Care. 2020 Apr;32(2):419-426. doi: 10.1007/s12 028-019-00711-w. PMID: 31290067.	A total of 2349 alarms from 45 patients were reclassified. The sensitivity of the RED alarm is 87.0% and the specificity is 29.6%, and the sensitivity of the GREEN alarm is 30.2% and the specificity is 87.2%.[29]

(A) Design of Patient Monitor Clinical Effectiveness Evaluation

Table 6 shows the clinical trial design of prior studies. In these studies, an observational study was conducted targeting intensive care unit patients and pediatric intensive care unit patients. The duration of the study ranged from 7 days to 10 weeks. Each study was conducted with 26 to 447 subjects and collected up to 3188 data. The data collected through the study included the type and number of alerts, frequency, cause and origin, appropriateness, and reliability. False alarms were determined based on the data collected through the study, and the results were analyzed according to the clinical endpoint of each study.

Table 6. Design of Similar Previous Studies on Patient Monitoring Alarm

Case Number	Subject	Study Type	Period	Number of Data and Subjects	Collection Data	Clinical Endpoint
Case 1	PICU patient	Observational	7 days	2,176 data	Type and number of alarm soundings	Number of False Alarms
Case 2	PICU patient	Observational	10 weeks	2,942 data	Occurrence rate, cause, and appropriateness of all alarms	True Positive, clinically relevant (TP-R) True Positive, clinically irrelevant (TP-I) False Positive (FP)
Case 3	ICU patient	Observational	3 weeks	1,455 data 26 patients	Origin and frequency of alarm soundings	Number of False Alarms

Case Number	Subject	Study Type	Period	Number of Data and Subjects	Collection Data	Clinical Endpoint
Case 4	ICU patient	Observational	-	3,682 data 38 patients	Alarm frequency and reliability	Number of False Alarms
Case 5	ICU patient	Observational	-	3,188 data 131 patients	Origin and frequency of alarm soundings	Positive Predictive Value (PPV), Negative Predictive Value (NPV), Sensitivity, Specificity
Case 6	ICU patient	Observational	-	447 patients	ECG arrhythmia alarms	False Alarm Rate
Case 7	ICU patient	Observational	-	2,349 data 45 patients	Reclassified alarms (true, possibly false, false)	Positive Predictive Value (PPV), Negative Predictive Value (NPV), Sensitivity, Specificity

(B) Endpoint of Patient Monitor Clinical Effectiveness

The clinical effectiveness endpoint is the outcome by which the effectiveness of medical devices in a clinical trial is evaluated.[30] Table 7 shows clinical effectiveness endpoints used in previous studies on patient monitoring alarm. Clinical effectiveness endpoints used in previous studies included the number of false alarms, False alarm rate, Positive Predictive Value (PPV), Negative Predictive Value (NPV), Sensitivity, and Specificity.

Table 7. Clinical Effectiveness Endpoint of Similar Studies on Patient Monitoring Alarm

Case Number	Clinical Effectiveness Endpoint
Case 1	Number of false alarms
Case 2	True Positive, clinically relevant (TP-R) True Positive, clinically irrelevant (TP-I) False Positive (FP)
Case 3	Number of false alarms
Case 4	Number of false alarms
Case 5	Positive Predictive Value (PPV), Negative Predictive Value (NPV) Sensitivity, Specificity
Case 6	False alarm rate
Case 7	Positive Predictive Value (PPV), Negative Predictive Value (NPV) Sensitivity, Specificity

Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) can be calculated using classical confusion matrices for the entire data set and specific subsets.[29] Figure 8 shows the confusion matrix and calculation formula. The confusion matrix, also known as the error matrix, is a table that is used to evaluate the

performance of a diagnostic device, classification model, or algorithm for the diagnosis, classification, or prediction.[31-33]

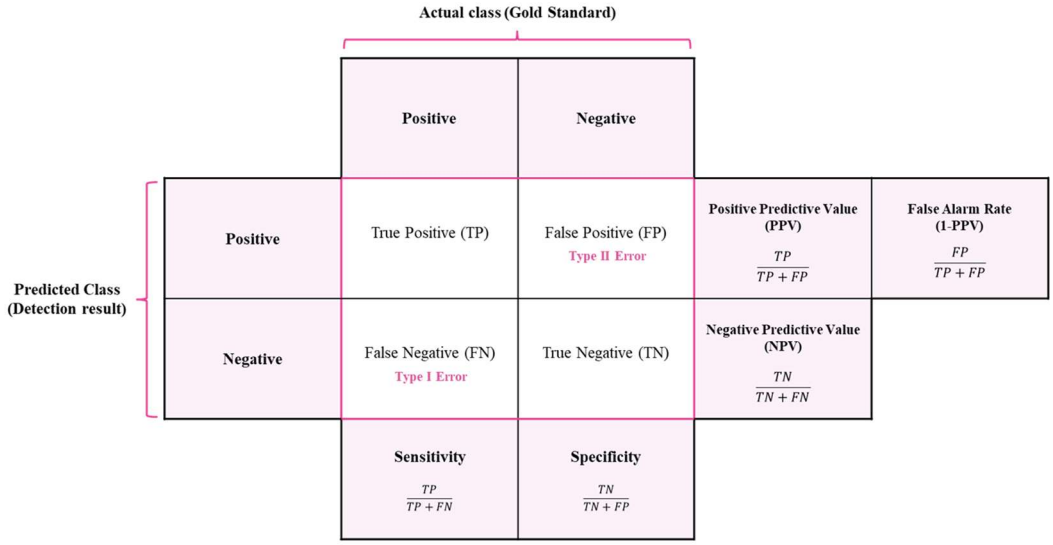


Figure 8. Confusion Matrix.

Sensitivity is the probability of identifying a person with the condition of interest among people who actually have the condition, and specificity is the probability of identifying a person without the condition of interest among people who do not actually have the condition.[33,34] PPV is the probability that people with positive test results actually have the condition of interest, and NPV is the probability that people with negative test results do not actually have the condition of interest.[33,34] True Positive (TP) represents the number of patients who actually have the condition of interest and have been properly classified as positive.[32,33,35] In previous studies, it was classified into true positive clinically relevant

(TP-R) and irrelevant (TP-I). TP-R indicates the sounded alarm was appropriate given the actual data value, and the patient's condition required prompt attention.[28] TP-I indicates the sounded alarm was appropriate given the input data value, but the patient's condition had not changed in a way that required additional medical attention.[28] False Positive (FP) represents the number of patients who actually do not have the condition of interest and have been misclassified as positive.[32,33,35]

False alarms indicate that the monitor sounded alarms, but the alarm was inappropriate given the input data value.[28] The number of false alarms equals the number of false positives, which indicates alarms that actually do not have the condition of interest and have been misclassified as positive. The false alarm rate is the probability that sounded alarms do not actually have the condition of interest.

2. Usability Evaluation of Patient Monitor

In intensive care units, doctors and nurses don't treat just one patient at a time,[17,36] they commonly use complex and error-prone equipment, and there's too much patient data for one person to handle effectively.[15] This high workload required of medical personnel can be the root cause of medical errors.[18] In addition, The ICU environment consists of a variety of alarm noises, flashing lights, and continuous operation of medical devices, which can increase the potential for errors and may contribute to burnout for clinicians, negative patient outcomes, and safety issues.[15,17,18]

Adverse events are often the result of poorly designed user interfaces without considering the capabilities and limitations of the end user, and user interface defects can be the root cause of adverse events.[37-39] Use errors occur in equipment with complex user interfaces, and in fact, injuries or deaths may occur due to mistakes such as pressing the wrong button or misperceiving a number while using the medical device.[40] Usability should be considered when designing user interfaces for medical devices and systems to facilitate use and prevent use errors, and the evaluation of a medical device’s user interface can be achieved through usability test.[41,42] For patient monitoring devices used in intensive care units, usability tests can be used to assess device effectiveness, satisfaction, and user design preferences. Furthermore, additional user interface evaluation may be required under actual use conditions as simulation use may not be sufficient for usability evaluation.

A. Use Scenarios and Critical Task Analysis

A use scenario is a description of a user interacting with the medical device to achieve a certain result under a specific use environment and it can be written in many different forms, ranging from narratives to simple lists of user tasks or steps in a task.[8] The purpose of the use scenario is to describe how the functions of a medical device are used while users are trying to achieve a result.[8] Use scenarios of the patient monitor consist of eight scenarios covering a wide range of situations including patient management, waveform, parameter setting, basic setting, alarm setting, arrhythmia alarm setting and alarm occurrence, display mode setting, and patient discharge (Table 8).

Table 8. Use Scenario of the Patient Monitor

No.	Use Scenario
1	Patient Setup
2	Waveform/Parameter Setting
3	Basic Setting
4	Alarm Setting
5	Arrhythmia Alarm Setting
6	Arrhythmia Alarm Occurs
7	Display Mode Setting
8	Discharge

Task is a distinct action or step within an overall workflow to achieve a desired result.[8,12] The tasks performed in each use scenario can be divided into 43 tasks (Table 9).

Table 9. Task of Patient Monitor Use Scenario

Use Scenario	Task Number	Task Description
Patient Setup	Task1	Admit new patient. • Patient Mode: Adult • Patient ID: ICU01 • Name: Jason • Birthdate: 1970.07.20 • Gender: Male
	Task2	Change the birthdate from 1970.07.29 to 1970.07.20
Waveform/Parameter Setting	Task3	Change the number of waveforms on the main screen
	Task4	Change the label of P2 to PAP
	Task5	Change the waveform from Respiration to PAP.
	Task6	Select ECG sweep speed to proper speed.
	Task7	Select ECG waveform size menu to proper size.
	Task8	Set NIBP interval to 1 hours.
	Task9	Change Respiration source randomly.
Basic Setting	Task10	By default, the settings window closes after a certain amount of time. Change the settings so that the settings window does not close.
	Task11	Set not to make a sound when the screen is touched
	Task12	Set the sound to be muted according to your heart rate.
Alarm Setting	Task13	Set the alarm limit display ON to display established alarm value on the monitor. Check the alarm limit location on the screen and point out it.
	Task14	Check all areas where the current visual alarm (SpO_2) occurs.
	Task15	Pause the alarm when the alarm occurs.
	Task16	Change the lower alarm limit of SpO_2 to 85%.

Use Scenario	Task Number	Task Description
Alarm Setting	Task17	Check all areas where the current visual alarm (P1) occurs.
	Task18	Change the systolic alarm upper limit value of P1 to 170mmHg through the ‘Alarm Setting Button’.
	Task19	Check the alarm message in Message List and check the alarm message on the screen.
	Task20	Set so that only audible alarms for 'diastolic pressure' and 'average pressure' of P1 are not generated.
Arrhythmia Alarm Setting	Task21	Set the arrhythmia alarm ON to display.
	Task22	Change the alarm condition for ventricular tachycardia to 135 bpm.
	Task23	Change the Run PVCs alarm condition to 7 beats.
Arrhythmia Alarm Occurs	Task24	Check the alarm of high-risk V-FIB, and point to the area where the visual and alarm messages occur.
	Task25	Pause the alarm when the alarm occurs.
	Task26	Check the visual and audible alarm for a medium-risk Tachycardia alarm.
	Task27	Check the visual and audible alarm for the low-risk Bigeminy alarms.
	Task28	Check that the current ECG waveform is normal and click the message list to check the arrhythmia alarms that have occurred so far. (V-FIB alarm record does not remain)
	Task29	Delete the arrhythmia alarm history on the alarm message list.
	Task30	Check a visual alarm to the current Pair PVCs alarm.
	Task31	Check ECG waveform and PVCs alarm message list.
	Task32	Check the visual alarm messages for new Run of PVCs alarms and compare them to ECG waveforms.
	Task33	Check that the ECG waveform is Multiform PVCs and that the alarm message is properly occurred. If the visual alarm message is not appropriate, wait for Multiform PVCs to appear.
	Task34	The ECG waveform returned to normal. However, visual and audible alarms for Multiform PVCs are still occurring. Turn off the current alarm by pressing the button indicating that you have recognized Multiform PVCs.
	Task35	Turn off all audible alarms.

Use Scenario	Task Number	Task Description
Display Mode Setting	Task36	Display the Big Number mode.
	Task37	Display the Tabular Trend mode.
	Task38	Tabular trend list is ordered by descending now. Change the display order from ascending to descending.
	Task39	Display the Graphical Trend mode.
	Task40	Change the display interval randomly.
	Task41	Display the Event review mode.
	Task42	Check all the data for 10 seconds before and after the event occurs.
Discharge	Task43	Discharge the patient.

A critical task is a task which, if performed incorrectly or not performed at all, would or could cause serious harm to the patient or user, where harm is defined to include compromised medical care.[8] The severity is divided into 6 levels as shown in (Table 10), and it is selected as a critical task if task has a severity higher than that causing injury or impairment requiring professional medical treatment.

Table 10. Definitions of Severity Levels

Rating	Severity	Description
1	Inconsequential	Inconvenience only
2	Negligible	Inconvenience or temporary discomfort
3	Minor	Results in recoverable injury or impairment not requiring professional medical intervention
4	Serious	Results in recoverable injury or impairment requiring professional medical intervention
5	Critical	Results in permanent impairment or life-threatening injury
6	Catastrophic	Results in death of patient or user

Table 11 shows the severity assessment of the hazardous situations related use scenarios and tasks. The severity was rated as 2 because it gives only inconvenience to the user in the situation where the user failed to register the patient. The severity of the failure of the user to set waveforms and parameters on the screen, failure to use the basic settings, and failure to use the screen mode was evaluated as 3, because temporary injury or impairment not requiring professional medical intervention may occur due to improper treatment and delay in treatment. In addition, failure to use the alarm settings and failure to use the arrhythmic function was evaluated as 4 severity because it which could cause deterioration of the patient's condition.

Table 11. Severity Assessment of Use Scenarios and Tasks

Hazardous Situation	Severity	Related Use Scenario	Related Task
Failure to use when a user registers a patient	2	US1	T1, T2
Failure to use when the user sets waveforms and parameters on the screen	3	US2	T3, T5, T6, T7, T8, T9
Failure when user uses basic settings	3	US3	T10, T11, T12
Failure to use user's alarm settings	4	US4	T13, T14, T15, T16, T17, T18, T19, T20
Failure of the user to use the arrhythmic function	4	US5, US6, US7	T21, T22, T23, T24, T25, T26, T27, T28, T30, T31, T34, T34,
Failure of user to use screen mode	3	US8	T36, T37, T38, T39, T40, T41, T42,

Depending on the severity assessed for each hazardous situation-related use scenario, it was determined whether the related task was a critical task (Table12). Task 4 was selected as a critical task because the severity of it was rate high because blood pressure

measurement was not performed well when use failure occurred when the label of P2 was changed to PAP. In the case of deleting the arrhythmia alarm history displayed on the alarm message list of Task 29, it was excluded from the critical task because it was of negligible severity, such as inconvenience or temporary discomfort if use failure occurred because it was simply deleting the record. Task 35 was not selected as a critical task because the severity of use failure was low because it is supposed to indicate that the audible alarm is turned off at the top of the screen when the setting is changed so that the audible alarm does not occur.

Table 12. Results of Important Task Selection

Use Scenario	Task Number	Severity	Critical Task
Patient Setup	Task1	2	N
	Task2	2	N
Waveform/Parameter Setting	Task3	3	N
	Task4	4	Y
	Task5	3	N
	Task6	3	N
	Task7	3	N
	Task8	3	N
	Task9	3	N
Basic Setting	Task10	3	N
	Task11	3	N
	Task12	3	N
Alarm Setting	Task13	4	Y
	Task14	4	Y
	Task15	4	Y

Use Scenario	Task Number	Severity	Critical Task
Alarm Setting	Task16	4	Y
	Task17	4	Y
	Task18	4	Y
	Task19	4	Y
	Task20	4	Y
Arrhythmia Alarm Setting	Task21	4	Y
	Task22	4	Y
	Task23	4	Y
Arrhythmia Alarm Occurs	Task24	4	Y
	Task25	4	Y
	Task26	4	Y
	Task27	4	Y
	Task28	4	Y
	Task29	3	N
	Task30	4	Y
	Task31	4	Y
	Task32	4	Y
	Task33	4	Y
	Task34	4	Y
	Task35	3	N
Display Mode Setting	Task36	3	N
	Task37	3	N
	Task38	3	N
	Task39	3	N
	Task40	3	N
	Task41	3	N
	Task42	3	N
Discharge	Task43	3	N

B. Usability Evaluation Indicators

(A) Task Success Rate

Task success rate is one of the quantitative methods of determining usability and is an indicator of how well a user achieves a specific goal.[43] It is useful to identify success levels when there are reasonable shades of gray associated with task success, such as completing a task partially.[44] In usability testing, there are cases where tasks are completed successfully, and cases where tasks are not completed or completed incorrectly and required assistance from the Moderator.

In addition to this, it is observed that cases of almost causing use errors, performing the task multiple times or taking a long time to perform, and mentioning difficulties are observed. Accordingly, the level of success can be divided into three levels: Completed (C), Completed with Issues (CI), and Not Complete (NC).[11] Table 13 shows success levels and their definitions.

Table 13. Task Success Level and Description

Level of Task Success	Description
Completed (C)	Completion of task successfully with no observed or reported close calls or use errors.[11]
Completed with Issues (CI)	Close call: Completion of a task in which a user almost commits an error but corrects the mistake himself or herself in time before making the error and completes the task.[11,12]
	Difficulty: Completion of a task with observed or expressed difficulties during the process, which may be revealed by multiple attempts to perform the task, comments of difficulties, facial expressions, and taking longer than expected. [11,12]

Level of Task Success	Description
Did Not Complete (NC)	Assistance: Assistance in which the participant has difficulty performing the task and asks for help to complete the task, or the participant is unable to complete the task for more than a predetermined amount of time.[11,12]
	Use error: User action or lack of action during the use of the medical device causes a result different from the intention of the manufacturer or the user's expectations.[8,11]

(B) Satisfaction

Satisfaction is one of the attributes included for usability according to ISO 9241-11:2018, meaning the positive association and absence of discontent that the user experiences during the performance.[43] Satisfaction refers to everything a user says or thinks about interacting with a product, and the user may have an opinion that the product is easy to use, confusing, or unreliable. User satisfaction is important in products where users have choices in using them.[44]

In the usability test, while performing tasks according to use scenarios, it is possible to evaluate the level of satisfaction with whether functions related to performing each task were easy to use or convenient. The 7-point Likert item is considered the best solution for questionnaires such as those used in usability evaluations.[45] This is because the 7-point Likert items provide a more accurate, easier-to-use, and better reflection of a respondent's actual evaluation than the 5-point Likert items.[45] Whether usability practitioners are developing a new summary scale, satisfaction survey, or a simple one-item post-test evaluation item, it is more appropriate to use a 7-point rather than a 5-point scale.[45]

Applying a 7-point Likert scale to rating satisfaction, a score of 1 means very dissatisfied and a score of 7 means very satisfied.

(C) System Usability

A standardized system usability assessment questionnaire used to assess perceived usability for a product or system is the Post-Study System Usability Questionnaire (PSSUQ). The PSSUQ is a questionnaire consisting of 16 items and has three subscales of system usefulness, information quality, and interface quality.[46,47] PSSUQ has been widely used in healthcare for the evaluation of anesthesia,[48] telerehabilitation systems,[49] radiation therapy systems,[50] and clinical monitoring.[51]

16 items were evaluated on a 7-point Likert scale, and the score range was 1 to 7, and the lower the score, the better the usability of the system and the higher the satisfaction. [46,50,52] The score for each subscale is calculated as an average score by dividing the total score by the number of questions answered. [46]

Table 14. Post-Study System Usability Questionnaire

No.	Post-Study System Usability Questionnaire	Subscale
1	Overall, I am satisfied with how easy it is to use this system.	System Usefulness
2	It was simple to use this system.	
3	I was able to complete the tasks and scenarios quickly using this system.	
4	I felt comfortable using this system.	

No.	Post-Study System Usability Questionnaire	Subscale
5	It was easy to learn to use this system.	System Usefulness
6	I believe I could become productive quickly using this system.	
7	The system gave error messages that clearly told me how to fix problems.	Information Quality
8	Whenever I made a mistake using the system, I could recover easily and quickly.	
9	The information (such as online help, on-screen messages, and other documentation) provided with this system was clear.	
10	It was easy to find the information I needed.	
11	The information was effective in helping me complete the tasks and scenarios.	
12	The organization of information on the system screens was clear.	
13	The interface of this system was pleasant.	Interface Quality
14	I liked using the interface of this system.	
15	This system has all the functions and capabilities I expect it to have.	
16	Overall, I am satisfied with this system.	-

(D) Workload

An evaluation index for evaluating workload is the National Aeronautics and Space Administration Task Load Index (NASA-TLX).[53,54] NASA-TLX is an evaluation index composed of six subscales: mental demand, temporal demand, physical demand, performance, effort, and frustration.[54] Although originally designed for aviation, it is a validated and commonly used tool in human factors engineering that has become the gold

standard for measuring subjective workload in a variety of applications.[55-57] In the healthcare field, it has been used to evaluate medical devices such as ventilators,[52] radiation therapy systems,[50] infusion pumps and physiological monitoring displays.[58,59]

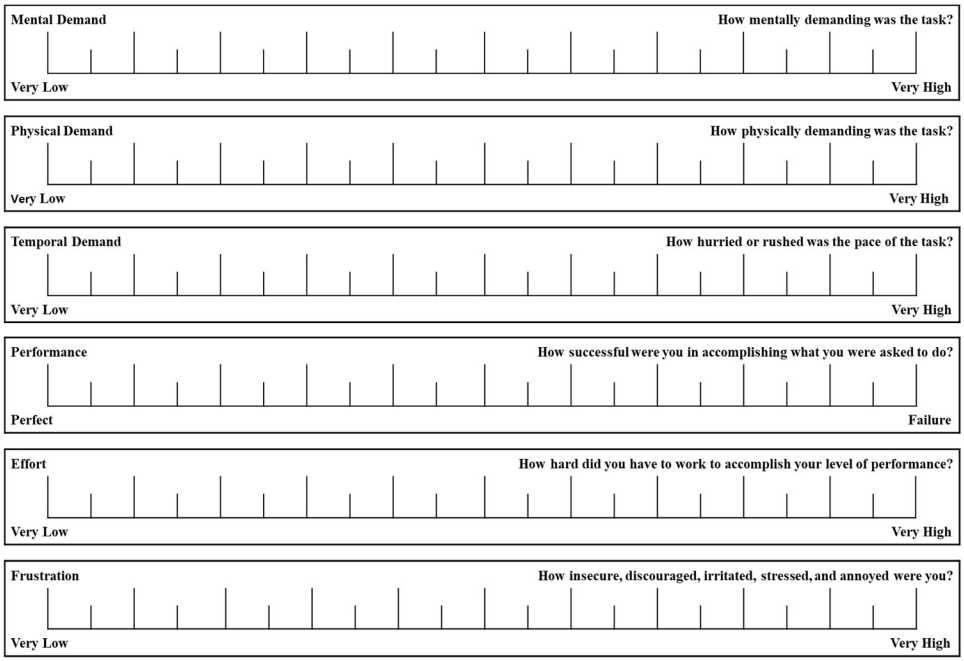


Figure 9. NASA Task Load Index.

Figure 9 is a graphical scale of the NASA Task load index. The increments of high, medium, and low on the graphical scale of the NASA Task load index are divided on a 21-point rating scale and are rated on a 20-point scale from 0 (very low) to 20 (very high).[57,60] The NASA-TLX score ranges from 0 to 100 as raw responses are transformed and mapped onto a 100-point scale.[61] The higher scores indicate a higher perceived workload and a more difficult user interface.[61]

Table 15. Description of NASA-TLX

Title	Description
Mental Demand	How much mental or perceptual activity such as thinking, deciding, searching, calculating, and remembering was required to perform a given task? Was the task easy or difficult, simple, or complex?[54,62]
Physical Demand	How much mental or perceptual activity such as thinking, deciding, searching, calculating, and remembering was required to perform a given task? Was the work slow or brisk, loose or strenuous?[54,62]
Temporal Demand	How much time pressure have you felt during your work due to the pace of work? Was the pace leisurely or frantic?[54] (e.g., if you need to perform many actions to complete a given task, you feel high time pressure)[62]
Performance	How successful and accurate do you think you were in accomplishing the goals of the task? How satisfied were you with completing these goals?[54,62]
Effort	How hard did you have to work mentally and physically to reach your performance level?[53] How much effort did you put in to accomplish a given task?[62] (e.g., if a great deal of concentration is required, it is a high-effort task)[62]
Frustration	How much did you feel discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent while you performed given a task?[47] (e.g., you can feel high frustration when you can't figure out how to do your job or when it's judged to be unrealistic)[62]

(E) User Preference

User preference survey is conducted to identify user requirements, and user preference survey data can be used as formative evaluation data.[10] The user preference survey questionnaire is configured to be used in decision-making for design reflection, such as whether a specific function is required, preference for basic settings, and what needs to be improved compared to other devices and what is more preferred.

3. Design of Patient Monitor Clinical Evaluation

Based on the clinical investigation case review of the patient monitor, clinical effectiveness endpoint investigation, use scenario and task analysis, and usability evaluation indicators investigation, clinical evaluation methods in the actual use environment and simulated environment of the patient monitor were derived (Figure 10).

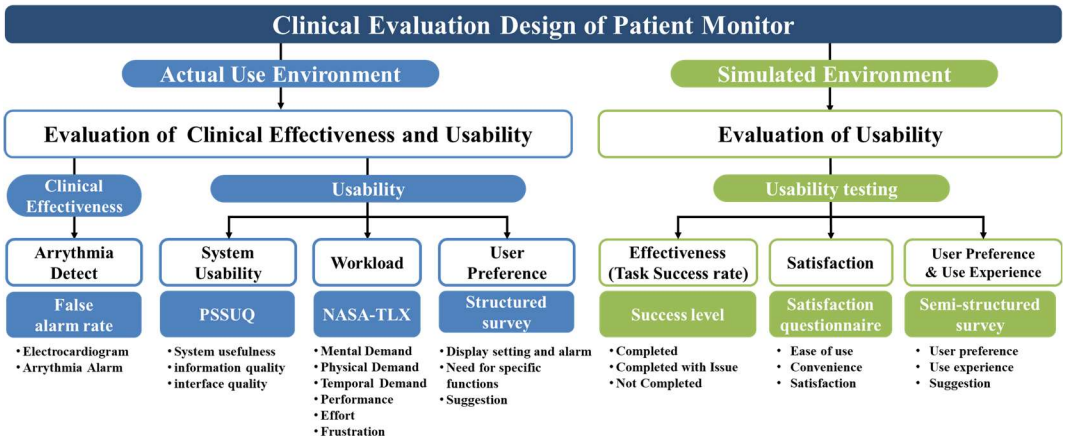


Figure 10. Clinical Evaluation Design of Patient Monitor.

As for the clinical evaluation in an actual use environment, a study was designed to evaluate the effectiveness of arrhythmia detection, which is a major monitoring parameter of patient monitors in the cardiology intensive care unit. For patient monitors used in intensive care units, usability testing in a simulated environment is not sufficient to evaluate the various risk controls, including alarm systems that can attract the user's attention.[10] so usability evaluation indicators were also added to the study design to evaluate the usability of actual patient monitor use. Clinical evaluation in a simulated environment was designed

to conduct a usability test based on usability evaluation indicators that cannot be evaluated in an actual use environment.

A. Clinical Evaluation in an Actual Use Environment

(A) Design

In a clinical evaluation in an actual use environment, a study was designed to evaluate the clinical effectiveness and usability of patient monitor A. Table 16 presents the design synopsis for clinical evaluation in an actual use environment. The study was conducted through clinical investigation as evaluates the clinical effectiveness of arrhythmia detection and alarm through electrocardiogram monitoring and evaluates the usability of the patient monitoring device. The experimental device was patient monitor A, and a comparative study was conducted with patient monitor B. The study was conducted in the cardiology intensive care unit at Severance Hospital, and participants were patients admitted to the intensive care unit and monitored by ECG, and nurses who use both experimental and control devices while nursing the patient. False alarm rate, system usability (PSSUQ), workload (NASA-TLX), and user preference data are collected.

Table 16. Clinical Evaluation Design Synopsis in an Actual Use Environment

Item	Description
Study Purpose	This study aims to compare and evaluate the effectiveness of the arrhythmia detection and alarm through the ECG monitoring of the two devices and evaluate the usability for intensive care unit nurses who have experienced using the two devices.

Item	Description
Design Details	• Allocation: Non-Randomized • Intervention Model: Single Group Assignment • Intervention Model Description: The study participants will be using both the experimental device and the control device simultaneously throughout the study participation. The participant will thereby act as its own control. • Masking: single-blinded (assessor) • Primary Purpose: Prevention
Institution	Severance Hospital
Investigational Device	Experimental device: patient monitor A Control device: patient monitor B
Participants	Participant in ECG monitoring: 20 patients
	Participant in the usability survey: 39 nurses
Participation Criteria for ECG Monitoring	Inclusion Criteria: • Male or female at least 20 years old. • Patient hospitalized at the intensive care unit and in need of ECG monitoring. • Patient who has been informed of the purpose and method of the clinical trial and has voluntarily agreed to participate in the clinical trial by signing an informed consent form or obtaining consent from a legal representative.
	Exclusion Criteria: • Patient with burns. • Patient with known allergy or sensitivity to any of the compositions in patch. • Patient with fragile skin (e.g., after prolonged cortisone treatment). • Patient with open sternum / sternum (e.g., severe heart failure postoperatively) or treatment for infection of the sternum. • Patient with mechanical auxiliary heart or ECMO. • Patient with implantable defibrillator. • Severely ill patients who refused life-sustaining treatment • Patient participating in any other clinical trial. • Patients who are judged inappropriate to participate in this clinical trial by the principal investigator or other investigators.
Participation Criteria for Usability Survey	Inclusion Criteria: • Nurses working in an intensive care unit who have used both the experimental device and the control device for clinical trial • Those who voluntarily signed consent after hearing the explanation of the purpose and method of this clinical trial
	Exclusion Criteria: • Those who are judged inappropriate to participate in this clinical trial by the principal investigator or other investigators
Primary Endpoint	Clinical effectiveness: False Alarm Rate
Secondary Endpoint	Usability: PSSUQ, NASA-TLX, User Preference

(B) Device

The Investigational devices for clinical effectiveness and usability evaluation in an actual use environment are patient monitor A and patient monitor B (Figure 11). Patient monitor A is a medical device developed by a Korean manufacturer and used for a case study in this study to design an optimized clinical evaluation. Patient monitor B is a medical device developed by a foreign manufacturer and has a high market share due to its good performance, high reliability, and preference, so it was selected as a comparative device in this study.

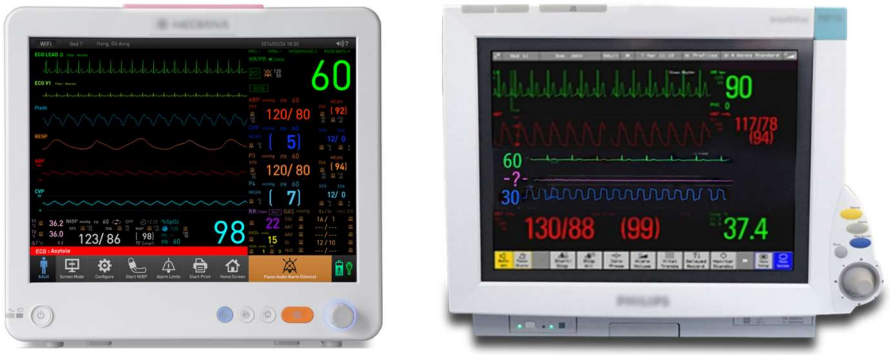


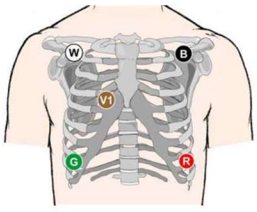
Figure 11. Investigational Devices (left: patent monitor A, right: patient monitor B).

There are several differences between the two devices. Table 17 shows the types of arrhythmia alarms generated through electrocardiogram monitoring of each patient monitor. Patient monitor A has fewer arrhythmia alarm types than patient monitor B, and the arrhythmia alarms of patient monitor B are more subdivided.

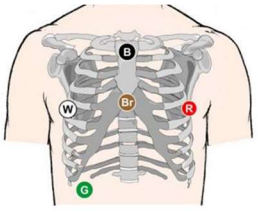
Table 17. Arrhythmia Alarm of patient Monitor A and B

Device	Arrhythmia Alarms
Patient Monitor A	Asystole, V-FIB, VTACH, HR High, HR Low, Ventricular rhythm, Tachy, Brady, Bigeminy, Trigeminy, Missed/Pause, R-on-T PVCs, Frequent PVCs, Pair PVCs (Couplet), Multiform PVCs,
Patient Monitor B	Asystole, Extreme Tachy, Vent Bigeminy, R-on-T PVCs, Run PVCs High, ST Multi, Vent FIB/TACH, Extreme Brady, Vent Trigeminy, PVCs/min High, Non-sustan VT, SVT, VTACH, HR HIGH, Missed Beat, Pair PVCs, ST High, Irregular, HR, Vent Rhythm, HR Low, Pause, Multiform PVCs, ST LOW

Patient monitor A and B have a difference in the placement of the ECG 5 leads. Patient monitor A uses the standard 5-lead placement and patient monitor B uses the EASI electrode placement. Figure 12 shows standard 5-lead placement and EASI electrode placement on patient monitors.



W(RA) Directly below the clavicle and near the right shoulder
B(LA) Directly below the clavicle and near the left shoulder
G(RL) On the right lower abdomen
R(LL) On the left lower abdomen
V1(V) On the chest



W(RA) On the right midaxillary line at the same level as the Br electrode
B(LA) On the upper sternum
G(RL) Below the sixth rib on the Right hip
R(LL) On the left midaxillary line at the same level as the Br electrode
Br(V) On the lower sternum at the level of the fifth intercostal space

Figure 12. ECG 5 Lead Placements of Patient Monitor. (left: Standard 5-lead placement, right: EASI electrode placement.)

The alarm patterns of patient monitor A and B have difference as shown in Table 18. In particular, there is a significant difference in the patterns of alarms that detect danger.

Table 18 Audible Alarm Pattern

Alarm Level	Patient Monitor A	Patient Monitor B
Danger (High)	○○○-○○ 10 times in 5 second (540Hz)	○○○ Repeat every second (High pitch.)
Warning (Medium)	○-○-○ 3 times per second (480Hz)	○○ Repeat every 2 seconds (Low pitch)
Caution (Low)	○ Once per second (400Hz)	○ Repeat every 2 seconds (Low pitch)

There are also differences in alarm trigger conditions and alarm levels. Table 19 shows the difference in tachycardia and bradycardia alarms between the two patient monitors. Patient monitor A occurs an alarm as soon as the alarm limit is exceeded and a low-level alarm occurs, but Patient monitor B generates a high-level alarm when the alarm limit is exceeded by more than 20.

Table 19. Alarm of Tachycardia and Bradycardia

Feature	Patient Monitor A	Patient Monitor B
Occurrence Condition	Alarm occurs immediately when getting out the alarm limit range.	Alarm triggers in the event of alarm limit +/- 20 bpm
Alarm Level	Low alarm	High alarm

Patient monitor B has specific functions that patient monitor A does not have. Its functions are wave freeze, drug calculation, standby mode, and alarm acknowledgment. Table 20 shows a description of each specific function.

Table 20. Specific Functions in Patient Monitor B

Function	Description
Wave Freeze	Freeze waves with a history of 20 seconds on the screen and measure parts of the wave using cursors.
Drug Calculation	Performing calculations for a non-specific drug and a specific drug. It calculates the dosage of drugs according to patient categories such as adults, children, and neonatal, types of drugs, and weight and height.
Standby Mode	Temporarily suspend monitoring. The patient data information and all settings are retained, and all waveforms and figures disappear from the display. Monitoring resumes when you select anything on the screen.
Alarm Acknowledging	Silence permanent key switches off the audible alarms and alarm lamps. A check mark appears next to the alarm message. After acknowledging the alarm, the alarm message will still be displayed if the condition that triggered the alarm still exists.

(C) Participant

(1) Patients in Cardiology Intensive Care Unit

Subjects to be applied and monitored with Investigational devices were recruited for patients who entered the intensive care unit and required electrocardiogram monitoring for more than 3 days in order to evaluate clinical effectiveness. Twenty patients in the cardiology intensive care unit were recruited and the patient distribution is shown in Table 21. The participating patients had paroxysmal atrial fibrillation, aortic aneurysm, myocardial infarction coronary artery occlusive disease, heart failure, aortic dissection, and ischemic cardiomyopathy.

Table 21. Distribution of 20 Patients

Category	Range	Number of People
Gender	Male	14
	Female	6
Age (In years)	Twenties (In 20's)	0
	Thirties (In 30's)	1
	Forties (In 40's)	2
	Fifties (In 50's)	4
	Sixties (In 60's)	7
	Seventies (In 70's)	3
	Eighties (In 80's)	3

(2) Nurses in Cardiology Intensive Care Unit

In order to evaluate the usability of the investigational device, subjects for the usability survey were recruited targeting nurses belonging to the intensive care unit who had used both the experimental device and the control device in the patient ECG monitoring process for clinical effectiveness evaluation. Thirty-nine nurses in the cardiology intensive care unit were recruited.

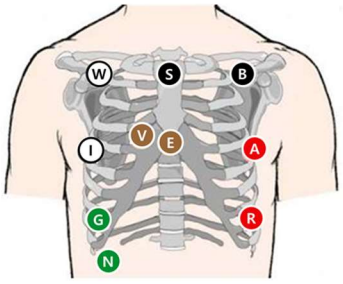
Table 22 shows the characteristics of 39 nurses. The intensive care unit nurses who participated in this study have a four-year college nursing degree, worked in an intensive care unit for an average of 7 years or more, know how to use and operate a patient monitoring system, and have clinical expertise.

Table 22. Characteristics of 39 Nurses

Category	Range	Number of People
Age	Twenties (In 20's)	26
	Thirties (In 30's)	13
	Forties (In 40's)	0
Career Period	More than 2 years and less than 5 years	10
	More than 5 years and less than 8 years	12
	More than 8 years and less than 11 years	14
	More than 11 years and less than 14 years	3

(D) Setting

For Clinical Evaluation in an actual use environment, patient monitor A and patient monitor B were installed in the cardiology intensive care unit of the cardiology department of Severance Hospital. The study participants simultaneously applied the experimental device and the control device throughout the study participation. In order to collect ECG and arrhythmia alarm occurrence data, ECG leads of the experimental device and control device are applied to the subject at the same time, and the ECG is measured from one patient through the ECG leads. Figure 13 presents the ECG 5 lead Placements of the test device (patient monitor A) and reference device (patient monitor B). Patient monitor A used standard 5-lead placement to connect leads and patient monitor B used EASI electrode placement to connect leads.



Patient Monitor A	W(RA)	Directly below the clavicle and near the right shoulder
	B(LA)	Directly below the clavicle and near the left shoulder
	G(RL)	On the right lower abdomen
	R(LL)	On the left lower abdomen
	V1(V)	On the chest
Patient Monitor B	E(V)	On the lower sternum at the level of the fifth intercostal space
	A(LL)	On the left midaxillary line at the same level as the Br electrode
	S(LA)	On the upper sternum
	I(RA)	On the right midaxillary line at the same level as the Br electrode
	N(RL)	Below the sixth rib on the Right hip

Figure 13. ECG 5 Lead Placements of Investigational Devices.

(E) Procedure

After the patient admitted to the intensive care unit (and/or legal representative) signs the informed consent form for participation in clinical evaluation, patients who met the inclusion criteria according to participation criteria for ECG monitoring and who did not meet the exclusion criteria participated in the clinical evaluation. To monitor the electrocardiogram and arrhythmia alarm, ECG leads of patient monitors A and B were applied to the subject at the same time. The electrocardiogram of the subject was monitored using the patient monitoring device, and the arrhythmia alarm and electrocardiogram waveforms that occur are recorded. Two independent cardiologists analyzed the arrhythmia alarm record and the electrocardiogram waveform where the alarm occurred to determine whether the arrhythmia alarm is a false alarm or an accurate alarm. Table 23 shows the clinical effectiveness evaluation procedure in an actual use environment.

Table 23. Clinical Effectiveness Evaluation Procedure in an Actual Use Environment

Procedure	Description
Obtain Written Informed Consent	Obtain informed consent from the patient or legal representative.
Review Participant Criteria	Review of participants according to inclusion/exclusion criteria
Application of Investigational Devices	Attach two patient monitor ECG leads to the patient
Electrocardiogram Measurement	Electrocardiogram monitoring for 3 days
Record Arrhythmia Alarm	Record the type of arrhythmia alarm, occurrence date, and time
Determine False Alarm	Determine false alarm by comparing arrhythmia alarm and electrocardiogram

After obtaining informed consent forms from nurses who wish to voluntarily participate in this clinical evaluation through a notice posted in the cardiology intensive care unit, nurses who met the inclusion criteria according to the participation criteria for the usability survey and did not meet the exclusion criteria were selected. The selected nurse used patient monitors A and B while nursing the patient who was monitoring ECG and arrhythmia alarms. Nurses conducted usability (PSSUQ), workload (NASA-TLX), and preference surveys at the end of shifts based on their experience of using investigational devices during working hours. Table 24 shows usability evaluation procedure in an actual use environment.

Table 24. Usability Evaluation Procedure in an Actual Use Environment

Procedure	Description
Obtain Written Informed Consent	Obtain informed consent from the patient or legal representative.
Review Participant Criteria	Review of participants according to inclusion/exclusion criteria
Use of Investigational Devices	Use patient monitors A and B while nursing patients who applied them.
System Usability Evaluation	System usability is evaluated using the PSSUQ
Workload Assessment	Workload is evaluated using NASA-TLX
User Preference Survey	Fill out the user preference questionnaires

(F) Data source & Analysis

(1) False Alarm

Arrhythmia alarms and electrocardiogram waveforms generated during the period of application of the investigational devices were collected, and false alarms were determined by comparing the arrhythmia alarms and ECG waveforms. Descriptive statistics (frequency and percentage) of false alarm rates for 3 days of application of each medical device were presented. The false alarm was evaluated by two independent evaluators, and if there was a difference of opinion, the additional opinions of one independent evaluator were collected to finally determine whether arrhythmia occurred. The Chi-square test or Fisher's exact test were performed to compare the false alarm rates of experimental and control devices.

(2) System Usability

System usability of patient monitor A and patient monitor B was quantitatively collected through PSSUQ targeting nurses in cardiology intensive care units who nursed patients applying investigational devices. A total of 16 Questionnaires consisting of system usefulness, information quality, and interface quality are scored on a 7-point Likert scale, and the higher the satisfaction, the lower the score. Descriptive statistics (, standard deviation, median, first quartile, and third quartile) of the PSSUQ score for evaluating the system usefulness of the survey device were presented. PSSUQ scores were presented for all 16 items, system usefulness, information quality, and interface quality. The independent two-sample t-test or Mann-Whitney's U test was conducted to compare the usability

(PSSUQ) scores between experimental devices and control devices.

(3) Workload

The workload of patient monitor A and patient monitor B was evaluated using NASA-TLX for cardiology intensive care unit nurses who care for patients with investigational devices applied. The degree of subjective workload felt by the user in terms of mental demand, physical demand, time demand, performance, effort, and dissatisfaction were scored on a graphical scale on the NASA Workload Index in the range of 0 to 21Descriptive statistics (mean, standard deviation, median, first quartile, and third quartile) of NASA-TLX scores that evaluated the workload of the survey device were presented. NASA-TLX scores were presented for all 6 subscales and overall scores, and an independent two-sample t-test or Wilcoxon rank-sum test was conducted to compare the workload (NASA-TLX) scores between test and control devices.

(4) User Preference

The User preference survey was conducted to evaluate subjective medical device design preferences felt by users after using patient monitor A and B equipment targeting cardiology intensive care unit nurses who nursed patients applying investigational devices.

The user preference survey consisted of 13 questions related to patient monitor display settings, the necessity for six specific functions, alarm patterns and alarm occurrence criteria, and arrhythmia alarm activation settings (Table25). For each option, the higher the selected ratio, the higher the user preference. Survey results for user preferences were

expressed as the ratio of options selected in each question. For each option, the higher the selected ratio, the higher the user preference.

Table 25. User Preference Survey for Patient Monitoring Devices

No.	User Preference Questionnaire			
1	When using a patient monitoring device, which language do you prefer, English or Korean?	1-1	High use frequency	English
				Korean
		1-2	Preference	English
				Korean
2	Do you think that the wave freeze function that keeps the previously measured parameter waveform on the screen is necessary when module is detached?			Yes
				No
3	What is your preference for screen composition change when module is detached?			Automatic change
				Holding the previous state
				User's setting
				Others
4	Do you prefer to have a not-measured/unused menu displayed on the screen?			Yes
				No
5	Do you think you need standby mode?			Yes
				No
6	Do you think you need an Early Warning Scoring (EW) function that provides a signal for a patient's deterioration in health?			Yes
				No
7	Do you think the Drug Calculation function is necessary for the patient monitoring system?			Yes
				No
8	Where would you like the alarm message to be displayed on the screen? (Refer to the figure. Duplicate selection is possible.)			Parameter waveform area
				Parameter numerical area
				Area at the top of the screen
				Area at the bottom of the screen
9	Which audible alarm pattern do you prefer?			Patient monitor A
				Patient monitor B

No.	User Preference Questionnaire			
10	Which one do you prefer when changing the pulse tone pitch according to the SpO_2 value change?			Change in tone pitch according to SpO_2 figure change during ECG and SpO_2 measurement
				No change in tone pitch when measuring ECG, but change in tone pitch according to change in SpO_2 figure when measuring only SpO_2
11	When an alarm occurs, do you think it is necessary to have an acknowledging alarm function that can permanently turn off the audible alarm and alarm lamp for the alarm that has occurred even if the corresponding alarm is still occurring?			Yes
				No
12	What is your alarm preference for tachycardia and bradycardia?	12-1	Occurrence condition	Alarm occurs immediately when getting out the alarm limit range.
				Alarm triggers in the event of alarm limit +/- 20 bpm
		12-2	Alarm level	High alarm
				Low alarm
13	What is your preferred setting for arrhythmia alarm activation (on/off)?			
13-1	ASYSTOLE	13-13	R-ON-T PVCs	
13-2	VENT FIB/TACH	13-14	PVCs/min HIGH	
13-3	VTACH	13-15	PAIR PVCs	
13-4	VENT RHYTHM	13-16	MULTIFORM PVCs	
13-5	EXTREME TACHY	13-17	RUN PVCs HIGH	
13-6	EXTREME BRADY	13-18	NON-SUSTAIN VT	
13-7	HR HIGH	13-19	ST HIGH	
13-8	HR LOW	13-20	ST LOW	
13-9	VENT BIGEMINY	13-21	ST MULTI	
13-10	VENT TRIGEMINY	13-22	SVT	
13-11	MISSED BEAT	13-23	IRREGULAR HR	
13-12	PAUSE			

In Question 8 of the questionnaire, the four areas in which an alarm message can be displayed on the screen of the patient monitoring device are the parameter waveform area, the numerical area, the top area of the screen, and the bottom area of the screen, as shown in Figure 14.

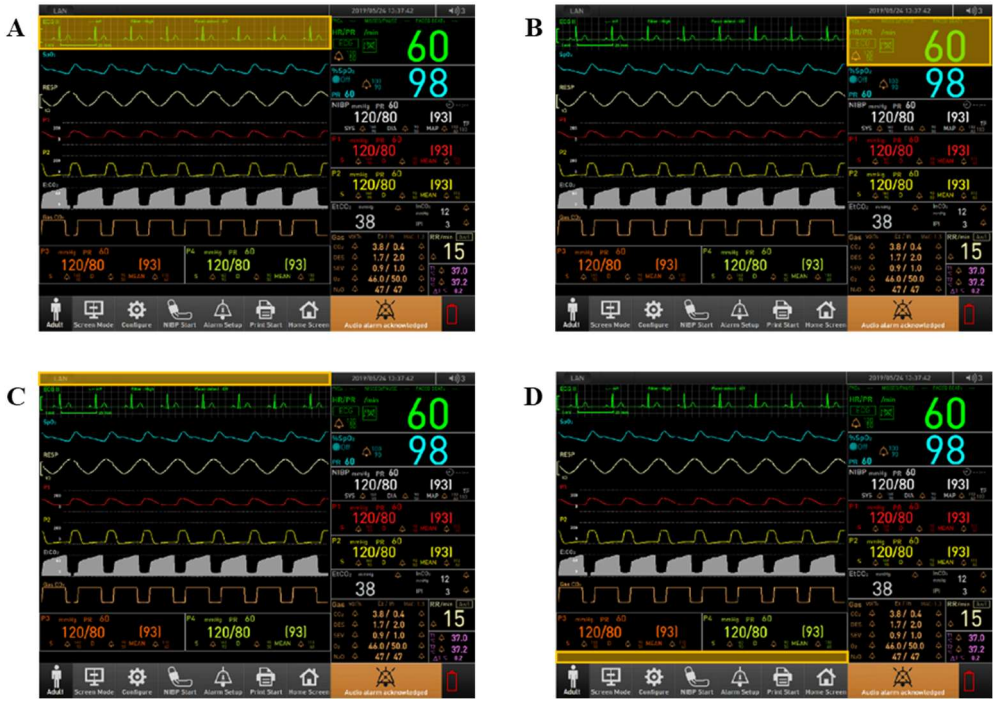


Figure 14. Four Areas Where Alarm Messages Can be Displayed on The Screen. (A: Parameter waveform area, B: Parameter numerical area, C: Top area of the screen, D: Bottom area of the screen)

B. Clinical Evaluation in a Simulated Environment

(A) Design

In a clinical evaluation in a simulated environment, A study was designed to evaluate the usability of the user interface of patient monitor. Table 26 presents the design synopsis for clinical evaluation in a simulated environment.

Table 26. Clinical Evaluation Design Synopsis in a Simulated Environment

Item	Description
Study Purpose	This study aims to evaluate the usability of the user interface for patient monitoring devices, including Alarm Occurs, Arrhythmia Alarm, Waveform and Parameter Setting, Display Mode Setting, Patient Setup, and specific functions of patient monitoring devices.
Design Details	• Evaluation type: Formative Evaluation • Evaluation metho: Usability test • Use scenarios: 8 scenarios • Tasks: 43 tasks
Institution	Medical Device Usability Research Center, Gangnam Severance Hospital
Test Device	Patient monitor A
Participants	Nurse, Physician, Biomedical Engineer: 6
Participation Criteria	Inclusion Criteria: • Career period More than 5 years
Observational data	Task Success Rate (Success level: task completed, task completed with Issue, Not Completed.)
Subjective data	Satisfaction, User Preference

The study was conducted through formative evaluation as a usability test, and 44 tasks were performed according to 8 use scenarios of the patient monitoring device. The test device was patient monitor A, and the test was conducted at Medical Device Usability

Research Center, Gangnam Severance Hospital. Participants were recruited from Nurses, Physicians, and Biomedical Engineers with more than 5 years of experience, and task success rate, satisfaction, and user preference data were collected.

(B) Device

The usability evaluation device was patient monitoring device A. Patient monitor A monitors electrocardiogram (ECG), heart rate/pulse rate (HR./PR), noninvasive blood pressure (NIBP), a saturation of percutaneous oxygen(SpO2), invasive blood pressure (IBP), respiration, partial pressure of carbon dioxide (EtCO2, InCO2), body temperature (Temperature), and multi-gas are measured to monitor the patient's condition for adult, pediatric, and neonatal patients in a hospital or environment equipped with hospital facilities. Parameters that can be monitored by the patient monitor in the simulated environment are ECG, Respiration, SpO2, NIBP, IBP, and EtCO2, and the functions that can be performed task are shown in the following table 27.

Table 27. Function of Patient Monitoring Device

Function	Description
Patient Setup	Patient Admit, Patient Discharge, Changing information
Monitoring	ECG, Respiration, SpO2, NIBP, IBP, EtCO2, Waveform and parameter change
Alarm	Visual alarm. Audible alarm, Alarm limit, Alarm Pause, Alarm OFF, Alarm message,
Display	Main screen, Big Number screen, Event Review screen, Graphic Trend screen

This medical device is a 15-inch monitor with screen touch function. The display specifications are shown in Table 28.

Table 28. Patient Monitor Display Specifications

Feature	Description
Size	15 in (Diagonal)
Type	Liquid Crystal Display (LCD) Color
Resolution	1080p, 720p, 480p
Waveform	8 waveforms
External Screen	Screen display (via HDMI port)

(C) Participant

To conduct the usability test, we recruited participants for the usability test targeting nurses, clinicians, and biomedical engineers who had experience in using patient monitoring devices. According to IEC 62366-2, formative evaluation involves 5 to 8 people, so a total of 6 participants were recruited, including 5 medical staff in the intensive care unit (4 nurses and 1 clinician) and 1 biomedical engineer. Intensive care medical staff were selected as the target group for this study because they are the major daily users of patient monitors at Severance Hospital. Since it is a medical device that is widely used in intensive care units, medical staff with more than 5 years of experience in using patient monitoring devices among medical staff in the intensive care unit of the Department of Cardiology

were recruited. Table 29 shows the characteristics of the usability test participants, such as age, area of expertise, experience, experience using similar devices, and period of patient monitoring device use.

Table 29. Characteristics of Usability Test Participants

Category	ID*	Age	Specializations	Career period	Similar Device Use Experience	Patient Monitor Use Period
Nurse	N1	40's	critical care nursing in CCU	More than 10 years	Philips	More than 10 years
	N2	30's	critical care nursing in CCU	More than 5 years and less than 10 years	Philips, Mediana	More than 3 years and less than 5 years
	N3	40's	critical care nursing in CCU	More than 10 years	Philips	More than 10 years
	N4	40's	critical care nursing in CCU	More than 10 years	Philips, Mediana	More than 10 years
Physician	P1	30's	Internal medicine & Critical care medicine	More than 5 years and less than 10 years	Philips	More than 3 years and less than 5 years
Biomedical Engineer	E1	30's	Biomedical Engineering	More than 5 years and less than 10 years	General Electric, Philips	More than 5 years and less than 10 years

* Participants are coded and recorded according to the corresponding user group

(D) Setting

The usability test was conducted in the usability testing room and observation room of the Medical Device Usability Research Center, Gangnam Severance Hospital, Yonsei University College of Medicine. A test environment was established in consideration of the actual use environment of the evaluation device. Table 30 shows the temperature, humidity, light and sound environment of the test room.

Table 30. Test Room Environment

Feature	Range
Temperature	22 ~ 24℃
Humidity	50 ~ 60 %
Light	780 ~ 850 lx
Sound	40 ~ 80dB

For the smooth progress of the evaluation, the evaluation activities were carried out by dividing the place into a device training room, a testing room, and an observation room. Figure 15 shows the Configuration of the equipment training room, test room, and observation room.

In the training room, a large monitor, a laptop, and a desk were placed for user training. In the training room, a large monitor, a laptop, and a desk were placed for user training. In the testing room, a patient monitoring device was installed next to the bed and a simulated environment was established using a dummy, a simulated patient. To simulate the patient's vital signs, we also located simulator equipment that generates signals that are monitored by patient monitoring devices. A large monitor was placed so that participants could follow the prompts and use the device to perform tasks, and a laptop was provided to evaluate their satisfaction and respond to preference surveys. The observation room was equipped with equipment that can observe and record participants' task performance in real-time in the testing room.

Figure 15. Configuration of Test Environment. (Left: Training room, middle: Testing room, right: Observation room)

(E) Procedure

The usability test moderator Introduced the usability test to the participants and obtained written informed consent to participate in the usability test. Participants were trained to use the device based on user training materials from an actual patient monitor manufacturer and tried to use the device for 1-2 minutes.

In a test session, participants used the device according to the prompts presented by the monitor, and a test observer observed all of the participants' interactions to record their level of success. Participants rated their satisfaction after completing each scenario. After the test session was over, participants filled out a questionnaire about their preference for using the patient monitor and use experience. Table 31 shows the Usability evaluation procedure in simulated environment.

Table 31. Usability Evaluation Procedure in Simulated Environment

Procedure	Description
Test Introduction and Obtain Informed Consent	Introduce the usability test to the participants and obtain consent to participate in the usability test.
Training	Conduct training on the test device for the participants.
Perform the Tasks	Perform tasks according to the prompts provided and interact with patient monitor A.
Observe Task Success	Observe and record the participant's level of task success
Satisfaction Evaluation	Evaluate the satisfaction survey of tasks of the scenario.
User Preference Survey	Fill out a questionnaire about user preference and use experience.

(F) Data Source & Analysis

(1) Task Success Rate

Observed participants' performance of 43 tasks in 8 use scenarios, recorded task success levels, and derived task success rates. The task success level was divided into task completed (C), task completed with Issue (CI), and Not Completed (NC). In the task success rate, task completion and task completion with Issue were included in successes, and Not Completed are included in failures. The success rate for each task and the success rate for each scenario were presented. The task success rate was calculated as the ratio of task completion and task completion with issue to the number of tasks performed. The scenario success rate was calculated as the ratio of task completion and task completion with issue to the total number of tasks performed.

(2) Satisfaction

The participant evaluated satisfaction after interacting with patient monitor A while performing the task of the scenario. Table 32 shows the satisfaction evaluation Questionnaire. The questionnaire consisted of a total of 23 questions designed to assess convenience and ease of use in relation to the interaction with the device required to perform each task. Satisfaction was assessed using a 7-point Likert scale. A score of 1 means very dissatisfied, and a score of 7 means very satisfied. The satisfaction evaluation result was presented as the average of each satisfaction evaluation score of the usability test participants for each of the 23 evaluation questions. Satisfaction scores for each scenario were also calculated and compared with target scores.

Table 32. Satisfaction Evaluation Questionnaire

Use Scenarios	No.	Satisfaction Questionnaire
Patient Setup	1	Is it easy to admit/discharge the patient?
	2	Is it easy to edit the patient?
Waveform/Parameter Setting	3	Is it easy to change the number of waveform setting?
	4	Is it easy to change the label of waveform?
	5	Is it easy to change waveform display?
	6	Is it easy to change the ECG waveform setting (waveform size and speed)?
	7	Is it easy to change NIBP interval?
	8	Is it easy to change RESP Source?
Basic Setting	9	Is it easy to change menu reset counter menu?

Use Scenarios	No.	Satisfaction Questionnaire
Basic Setting	10	Is it easy to change the sound setting?
Alarm Setting	11	Is it proper visual alarm display location and display methodology?
	12	Is it easy to set alarm limit setting and alarm display setting?
	13	Is it easy to approach audible alarm stop button per parameter?
Arrhythmia Alarm Setting	14	Is it easy to set arrhythmia alarm setting?
Arrhythmia Alarm Occurs	15	Is it proper to visual/audible alarm expression according to the risk of arrhythmia?
	16	Does the action of the alarm pause button appropriate according to the risk of arrhythmia?
	17	Is proper the alarm occurrence conditions and alarm message duration for each type of PVCs?
	18	Is it easy to approach the overall audible alarm off button?
Display Mode Setting	19	Do you think the main screen is organized for easy operation?
	20	Are you satisfied with the screen configuration with large numbers?
	21	Are you satisfied with the Tabular trend screen?
	22	Are you satisfied with the Graphical trend screen?
	23	Are you satisfied with the Event review screen?

In question 11, the visual alarm display of patient monitor A includes the flickering of the background color of the alarm occurrence parameter value in the parameter numerical area and the display of an alarm message in the bottom area of the screen (Figure 16).

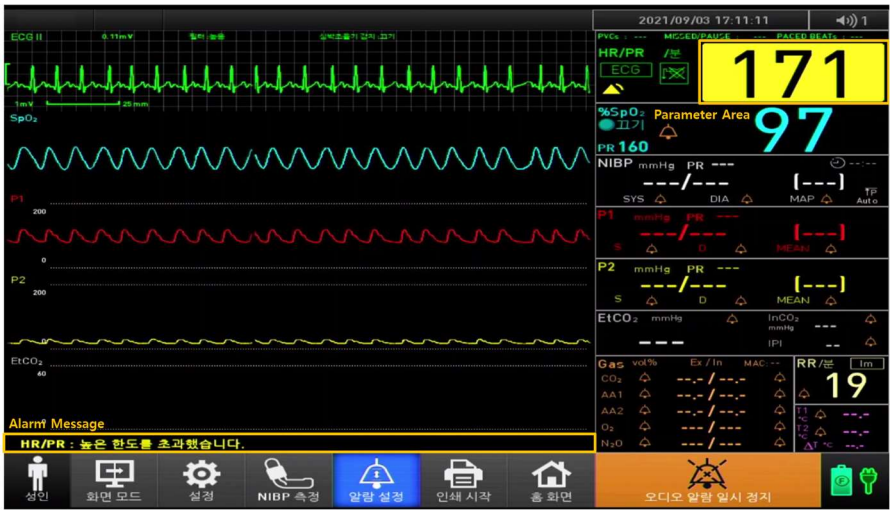


Figure 16. Visual Alarm Display of Patient Monitor A.

In question 13, The audible alarm stop button for each parameter is a bell-shaped button as shown in Figure 17. If you press the button to stop the alarm, it changes to an image with an X mark on the bell, and you can turn off the audible alarm.

Figure 17. Audible Alarm Stop Button for Each Parameter.
(left: audio alarm activated state, right: audio alarm stopped state)

In question 14, arrhythmia settings are available by pressing the arrhythmia settings menu in the ECG waveform menu (Figure 18). There are buttons to turn on/off arrhythmia alarms, and for some arrhythmias such as ventricular tachycardia, the alarm occurrence condition can be changed.

Figure 18. Arrhythmia Alarm Setup Menu.

In question 16, the action of the alarm pause button according to the risk of arrhythmia is shown in Table 33. As for the operation of the alarm pause button according to the risk of arrhythmia, only the audible alarm is paused for 60 to 120 seconds in the case of high and medium alarms. In the case of low alarms, both visual and audible alarms are turned off. In the message list window, delete the previous alarm history recorded in the message list.

Table 33. Action of the Alarm Pause Button

Category	Function of The Alarm Pause Button
Hight/Medium alarm	Pause of audible alarm for 60 to 120 seconds
Low alarm	Visual/Audible alarm OFF
Alarm message list window	Delete previous alarm history recorded in message list

In question 17, In question 17, the alarm occurrence conditions and alarm message duration for each type of PVCs are shown in Table 34. As for the alarm message duration, for Run of PVCs, Pair PVCs, and Multiform PVCs, even if the status becomes normal, the alarm is maintained for at least 1 minute. In addition, if a PVC alarm with a higher priority than the currently occurring PVC alarm does not occur, the alarm is maintained for 1 minute. Frequent PVCs Keep alarm status when PVC count is 10 or more.

Table 34. Alarm Occurrence Conditions and Alarm Message Duration for Type of PVCs

Alarm Priority	PVCs Alarm Type	Alarm occurrence condition	Alarm message duration
1	Run of PVCs	PVC occurs continuously more than a certain number of times	Even if the status becomes normal, the alarm is maintained for at least 1 minute If a PVC alarm with a higher priority than the currently occurring PVC alarm does not occur, the alarm is maintained for 1 minute.
2	Pair PVCs	PVC occurs twice in a row during normal waveforms	
3	Multiform PVCs	different types of PVC occur	
4	Frequent PVCs	More than 10 PVCs in 1 minute	Keep alarm status when PVC count is 10 or more

In question 18, The overall audible alarm off button is a button that permanently stops an audible alarm, and a display indicating permanent stop appears on the top of the screen and on the audible alarm pause button.

Figure 19. Overall Audible Alarm Off Button.

(3) User Preference

The User preference and use experience survey was conducted to evaluate the user's subjective medical device design preference and use experience after using Patient monitor A for medical staff and Biomedical engineers. As shown in Table 35, this survey consisted of 7 questions related to whether additional specific functions are required, whether there is experience in using a telemetry system, and what needs to be improved compared to the patient monitoring device in use.

Survey results for user preference and experience were expressed as the number of options selected in each question. For each option, the higher the selected number, the higher the user preference.

Table 35. User Preference and Use Experience Survey

No.	Questionnaire
1	Depending on the use environment (ex. ICU, NICU), Please write which one you prefer and why between the patient type (ex. adult, neonatal) settings that are maintained or changed every time the power is turned on.
2	Please write whether you need the Wave Freeze feature and why.
3	Please write whether the alarm reset function is necessary and why.
4	Please write whether you need the Alarm Off function and why.
5	Please write whether Standby mode is necessary and why.
6	Have you ever used a telemetry system that measures the patient's ECG signal while the patient is moving, transmits the data to the central station through the wireless network, and also analyzes the ECG data?
7	If you have any suggestions for improvement of the patient monitor A compared to other intensive care unit patient monitoring devices, please write.

C. Result of Patient Monitor Clinical Evaluation

(A) False Alarm Rate

Table 36 presents the count, expected count and ratio of false alarms and correct alarms generated by patient monitors A and B. A total of 387 alarms occurred in patient monitor A, of which 344 were false alarms. A total of 33 alarms occurred in patient monitor B, and 29 of them corresponded to false alarms. The false alarm rates of patient monitor A and patient monitor B were 89.1% and 87.9%, respectively, showing no significant difference.

Since the expected count of one cell was 3.7, which was less than 5, Fisher's exact test was performed to confirm whether there was a difference in the occurrence of false alarms according to the use of patient monitors A and B. As a result of Fisher's exact test, there was no significant association between patient monitor A and B false alarm occurrence ($p=0.777 > 0.05$).

Table 36. Result of False Alarm Occurrence

	Patient Monitor A			**Patient Monitor B**			**p***
	Count	**Expected Count**	**(%) of Group**	**Count**	**Expected Count**	**(%) of Group**	
False Alarm	344	343.7	89.1	29	29.3	87.9	0.777
Correct Alarm	43	43.3	11.1	4	3.7	12.1	

* p-values were determined with use of fisher's exact test for categorical variable

(B) System Usability

Table 37 presents the descriptive statistics (mean, standard deviation, median, first quartile, third quartile) of the PSSUQ scores for system usability of the patient monitors A and B. The average scores of the system usefulness (questionnaires 1 - 6) of patient monitor A and B were 2.86 and 2.25, respectively, and the average scores of the information quality (questionnaire 7 -1 2) of patient monitor A and B were 2.87 and 2.47, respectively, the average scores of the interface quality (questionnaire 13-15) of patient monitor A and B were 3.12 and 2.17, respectively. The lower the PSSUQ score, the higher the perceived satisfaction after using the two devices. In terms of user's system usability, information quality, interface quality, and satisfaction, scores generally ranged from 2 to 3 points.

Table 37. Descriptive Statistics of PSSUQ Scores

	SEVERANCE CCU(n=39)	
	Patient Monitor A	**Patient Monitor B**
1	2.68±0.96 (2.50, 2.00–3.00)	2.21±0.78 (2.00, 2.00–2.25)
2	2.69±0.91 (2.70, 1.93–3.08)	2.29±0.98 (2.00, 2.00–3.00)
3	2.95±0.97 (2.85, 2.00–3.35)	2.13±0.99 (2.00, 1.75–2.25)
4	2.99±1.10 (2.70, 2.00–4.00)	2.32±1.07 (2.00, 2.00–3.00)
5	2.83±1.02 (2.70, 2.00–3.55)	2.37±1.08 (2.00, 2.00–3.00)
6	3.04±1.03 (9.00, 2.30–3.70)	2.24±1.05 (2.00, 1.75–3.00)
System usefulness	2.86±0.91 (2.70, 2.10–3.35)	2.25±0.88 (2.00, 1.83–2.50)
7	3.19±1.16 (3.00, 2.30–4.00)	2.61±1.31 (2.50, 2.00–3.00)
8	3.02±0.92 (3.00, 2.30–3.70)	2.47±1.11 (2.50, 2.00–3.00)

	SEVERANCE CCU(n=39)	
	Patient Monitor A	**Patient Monitor B**
9	3.01±1.14 (2.70, 2.00–3.70)	2.55±1.20 (3.00, 1.00–3.00)
10	2.84±0.98 (3.00, 2.00–3.30)	2.71±1.18 (2.50, 2.00–3.00)
11	2.83±0.86 (2.70, 2.00–3.30)	2.34±0.85 (2.00, 2.00–3.00)
12	2.33±0.84 (2.15, 1.70–3.00)	2.16±0.79 (2.00, 2.00–3.00)
Information quality	2.87±0.88 (2.80, 2.16–3.39)	2.47±0.84 (2.33, 2.00–2.83)
13	2.21±1.16 (3.00, 2.30–4.08)	1.84±0.86 (2.00, 1.00–2.25)
14	2.07±1.12 (3.00, 2.23–3.78)	2.11±0.95 (2.00, 1.00–3.00)
15	2.07±1.00 (3.00, 2.30–3.78)	2.61±1.08 (2.00, 2.00–3.00)
Interface quality	3.12±0.98 (3.02, 2.41–3.30)	2.17±0.79 (2.00, 1.67–2.67)
16	2.82±0.90 (2.70, 2.30–3.30)	2.11±0.80 (2.00, 2.00–2.00)
PSSUQ	2.91±0.86 (2.84, 2.20–3.29)	2.31±0.77 (2.19, 1.88–2.56)

Table 38 presents the statistical comparison of PSSUQ scores for system usability for patient monitors A and B. In the case of system usefulness and interface quality, because the data were not normally distributed, the Mann-Whitney's U test was conducted to compare the PSSUQ scores of the two devices. The results of the Mann-Whitney's U test show that there was a statistically significant difference in system usefulness ($p=0.001<0.05$) and interface quality ($p<0.001$). In the case of information quality, because the data were normally distributed, the independent two-sample t-test was performed, and there was a significant difference in the PSSUQ scores of the patient monitors A and B ($p=0.044<0.05$). The Overall PSSUQ score was 2.19 (1.88-2.56) for patient monitor A and

2.84 (2.20-3.29) for patient monitor B, showing a statistically significant difference as a result of comparing the two devices with Mann-Whitney's U test($p=0.001<0.05$).

The lower the PSSUQ score, the higher the perceived satisfaction after using the two devices. Therefore, it can be confirmed that the satisfaction of patient monitor B was higher than that of patient monitor A in terms of user's system usefulness, information quality, interface quality, and Overall system usability.

Table 38. Statistical Comparison of PSSUQ Scores

	SEVERANCE, CCU(n=39)				
	Patent Monitor A	**Patent Monitor B**	**t***	**U****	**P-value#**
System Usefulness	2.70 (2.10-3.35)	2.00 (1.83-2.50)		422	0.001
Information Quality	2.87±0.88	2.47±0.84	2.049		0.044
Interface Quality	3.02 (2.41-3.57)	2.00 (1.67-2.67)		332.5	<0.001
PSSUQ (Overall)	2.84 (2.20-3.29)	2.19 (1.88-2.56)		422	0.001

* Because the data were normally distributed, the independent two sample t-test was performed. **Because the data were not normally distributed, the Mann-Whitney's U test was performed. #P-values were determined with the independent two sample t-test and Mann-Whitney's U test for continuous variables.

(C) Workload

Table 39 presents the descriptive statistics (mean, standard deviation, median, first quartile, third quartile) of the TLX scores for Workload of the patient monitors A and B. The average TLX scores of patient monitor A and B were 31.79 and 24.62 in the mental requirement category, 27.23 and 23.72 in the physical requirement category, and 60.46 and

55.26 in the temporal requirement category, respectively. The average score of the performance item was 48.75 and 37.69, respectively, the effort item was 33.13 and 28.33, and the dissatisfaction item was 25.53 and 23.85, respectively.

Table 39. Descriptive Statistics of TLX Scores

	SEVERANCE CCU(n=39)	
	Patent Monitor A	**Patent Monitor B**
Mental Demand	31.79±14.44 (30.00, 20.00–41.25)	24.62±16.40 (20.00, 10.00–30.00)
Physical Demand	27.23±13.14 (26.70, 18.10–33.30)	23.72±17.69 (20.00, 10.00–30.00)
Temporal Demand	60.46±14.05 (60.85, 49.58–70.00)	55.26±27.00 (60.00, 30.00–80.00)
Performance	48.75±17.05 (49.15, 34.58–63.73)	37.69±22.71 (30.00, 20.00–50.00)
Effort	33.81±13.78 (34.15, 26.28–40.43)	28.33±18.00 (25.00, 15.00–40.00)
Frustration	25.53±16.33 (25.00, 12.48–33.73)	23.85±20.53 (20.00, 10.00–30.00)
TLX (Overall)	37.93±8.35 (39.17, 34.79–43.40)	32.24±12.79 (35.00, 26.67–39.17)

Table 40 presents the statistical comparison of TLX scores for system usability for patient monitors A and B. In all six subscales of mental demand, physical demand, time demand, performance, effort, and dissatisfaction, because the data were not normally distributed, the Mann-Whitney's U test was conducted to compare the TLX scores of the two devices. The results of the Mann-Whitney's U test comparing the TLX scores of the patient monitor A and B including physical demand ($P=0.119>0.05$), temporal demand ($P=0.870>0.05$), and frustration ($P =0.0.355>0.05$) did not show a significant difference in the scores. However, there was a statistically significant difference in the scores for mental demand ($p = 0.019<0.05$), performance ($p = 0.009<0.05$) and effort ($p=0.045<0.05$).

In conclusion, the Overall TLX score of patient monitor B was 35.00 (6.67-39.17), which was statistically significantly lower than the patient monitor A score of 39.17 (34.79-43.40) (P=0.005<0.05). Through the results, It was confirmed that there was no significant difference in the degree of physical activity required, the feeling of time pressure, and frustration when users performed tasks of the two medical devices.

Table 40. Statistical Comparison of TLX Scores

	SEVERANCE, CCU(n=39)			
	Patent Monitor A	**Patent Monitor B**	**Mann-Whitney's U***	**P-value****
Mental Demand	30.00(20.00-41.25)	20.00(10.00-30.00)	512.0	0.019
Physical Demand	26.70(18.10-33.30)	20.00(10.00-30.00)	588.5	0.119
Temporal Demand	60.85(49.58-70.00)	60.00(30.00-80.00)	725.0	0.870
Performance	49.15(34.58-63.73)	30.00(20.00-50.00)	484.0	0.009
Effort	34.15(26.28-40.43)	25.00(15.00-40.00)	544.5	0.045
Frustration	25.00(12.48-33.73)	20.00(10.00-30.00)	650.5	0.355
TLX (Overall)	39.17(34.79-43.40)	35.00(6.67-39.17)	468.5	0.005

*Because the data were not normally distributed, the Mann-Whitney's U test was performed. **P-values were determined with the Mann-Whitney U test for continuous variables.

In the mental demand, performance, and effort that showed a statistically significant difference in scores between the two devices, the average score of patient monitor A was higher than that of patient monitor B. This means that when the user performed the task using patient monitor B, activities such as thinking, decision-making, and memory were less demanding than patient monitor A, so it was less mentally demanding, and the task was performed successfully and accurately without much effort.

(D) User Preference form Nurses

Table 41 shows the results of a user preference survey to evaluate users' subjective medical device design preference after using Patient monitors A and B targeting intensive care unit nurses.

Table 41. Results of the User Preference Survey

No.	User Preference Questionnaire			Number of Responses	
1	When using a patient monitoring device, which language do you prefer, English or Korean?	1-1	High use frequency	English	30
				Korean	9
		1-2	Preference	English	28
				Korean	12
2	Do you think that the wave freeze function that keeps the previously measured parameter waveform on the screen is necessary when module is detached?			Yes	25
				No	14
3	What is your preference for screen composition change when module is detached?			Automatic change	24
				Holding the previous state	17
				User's setting	1
				Others	0
4	Do you prefer to have a not-measured/unused menu displayed on the screen?			Yes	30
				No	11
5	Do you think you need standby mode?			Yes	36
				No	1
6	Do you think you need an Early Warning Scoring (EW) function that provides a signal for a patient's deterioration in health?			Yes	32
				No	7
7	Do you think the Drug Calculation function is necessary for the patient monitoring system?			Yes	12
				No	25
8	Where would you like the alarm message to be displayed on the screen? (Refer to the figure. Duplicate selection is possible.)			Parameter waveform area	22
				Parameter numerical area	24

No.	User Preference Questionnaire			Number of Responses	
8	Where would you like the alarm message to be displayed on the screen? (Refer to the figure. Duplicate selection is possible.)			Area at the top of the screen	11
				Area at the bottom of the screen	4
9	Which audible alarm pattern do you prefer?			Patient monitor A	6
				Patient monitor B	32
10	Which one do you prefer when changing the pulse tone pitch according to the SpO2 value change?			Change in tone pitch according to SpO2 figure change during ECG and SpO2 measurement	19
				No change in tone pitch when measuring ECG, but change in tone pitch according to change in SpO2 figure when measuring only SpO2	18
11	When an alarm occurs, do you think it is necessary to have an acknowledging alarm function that can permanently turn off the audible alarm and alarm lamp for the alarm that has occurred even if the corresponding alarm is still occurring?			Yes	36
				No	1
12	What is your alarm preference for tachycardia and bradycardia?	12-1	Occurrence condition	Alarm occurs immediately when getting out the alarm limit range.	20
				Alarm triggers in the event of alarm limit +/- 20 bpm	19
		12-2	Alarm level	High alarm	30
				Low alarm	6
13-1	What is your preferred setting for ASYSTOLE alarm activation (on/off)?			On	35
				Off	4
13-2	What is your preferred setting for VENT FIB/TACH alarm activation (on/off)?			On	36
				Off	2
13-3	What is your preferred setting for VTACH alarm activation (on/off)?			On	37
				Off	1
13-4	What is your preferred setting for VENT RHYTHM alarm activation (on/off)?			On	26
				Off	11

No.	User Preference Questionnaire	Number of Responses	
13-5	What is your preferred setting for EXTREME TACHY alarm activation (on/off)?	On	32
		Off	5
13-6	What is your preferred setting for EXTREME BRADY alarm activation (on/off)?	On	32
		Off	5
13-7	What is your preferred setting for HR HIGH alarm activation (on/off)?	On	36
		Off	1
13-8	What is your preferred setting for HR LOW alarm activation (on/off)?	On	36
		Off	1
13-9	What is your preferred setting for VENT BIGEMINY alarm activation (on/off)?	On	32
		Off	8
13-10	What is your preferred setting for VENT TRIGEMINY alarm activation (on/off)?	On	32
		Off	7
13-11	What is your preferred setting for MISSED BEAT alarm activation (on/off)?	On	19
		Off	19
13-12	What is your preferred setting for PAUSE alarm activation (on/off)?	On	35
		Off	1
13-13	What is your preferred setting for R-ON-T PVCs alarm activation (on/off)?	On	29
		Off	10
13-14	What is your preferred setting for PVCs/min HIGH alarm activation (on/off)?	On	28
		Off	9
13-15	What is your preferred setting for PAIR PVCs alarm activation (on/off)?	On	23
		Off	15
13-16	What is your preferred setting for MULTIFORM PVCs alarm activation (on/off)?	On	29
		Off	10
13-17	What is your preferred setting for RUN PVCs HIGH alarm activation (on/off)?	On	31
		Off	6
13-18	What is your preferred setting for NON-SUSTAIN VT alarm activation (on/off)?	On	37
		Off	1
13-19	What is your preferred setting for ST HIGH alarm activation (on/off)?	On	19
		Off	19

No.	User Preference Questionnaire	Number of Responses	
13-20	What is your preferred setting for ST LOW alarm activation (on/off)?	On	13
		Off	25
13-21	What is your preferred setting for ST MULTI alarm activation (on/off)?	On	12
		Off	24
13-22	What is your preferred setting for SVT alarm activation (on/off)?	On	30
		Off	9
13-23	What is your preferred setting for IRREGULAR HR alarm activation (on/off)?	On	24
		Off	16

Figure 20 shows the results of a user preference survey for patient monitor display settings and alarm. Based on the results of the patient monitor display setting preference survey, English was preferred for the language setting of the patient monitor, and the frequency of use was also higher in English. For the change of the screen composition according to the detachment of the patient monitor module, it was preferred to change automatically. The most preferred location for the alarm message was the parameter area, followed by the waveform area, top of the screen, and bottom of the screen.

Based on the results of the patient monitor alarm preference survey, for the audible alarm pattern, the pattern of patient monitor B was preferred to patient monitor A. Pulse tone pitch was preferred to change according to SPO2 change, and in the case of tachycardia and bradycardia alarms, it was preferred to sound immediately out of the alarm limit range. Users preferred that the tachycardia and bradycardia alarm levels were high-level alarms rather than low-level alarms.

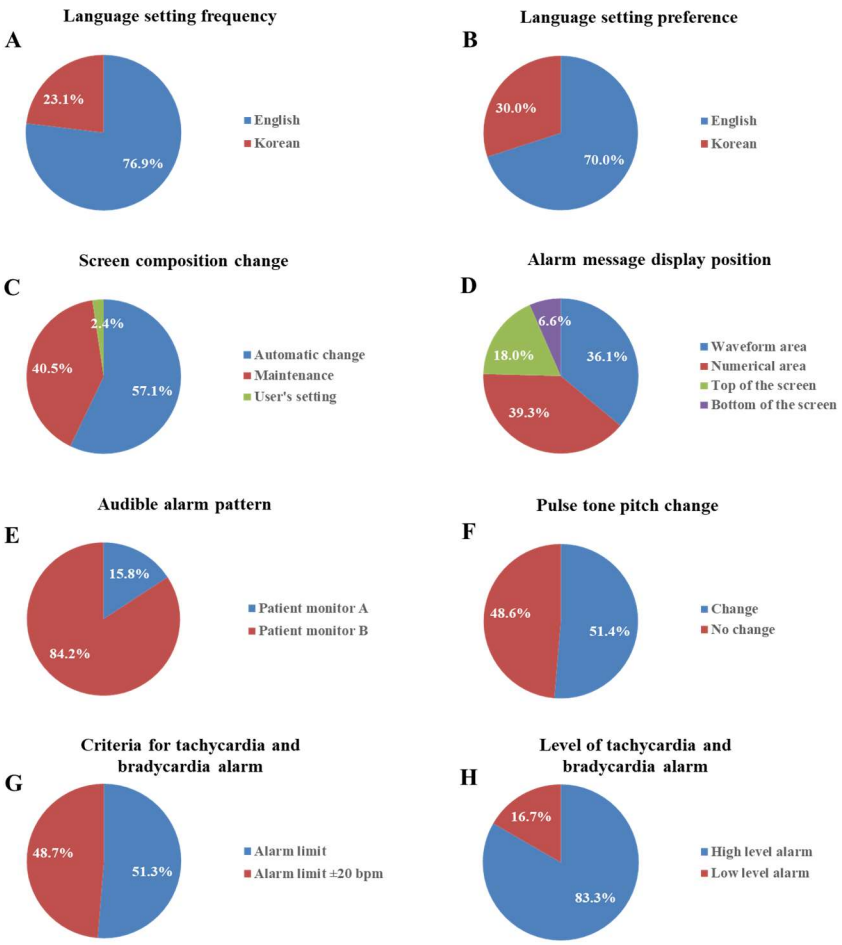


Figure 20. User Preference Survey Results for Display Setting and Alarm.

Figure 21 shows the preference survey results on the need for 6 specific functions of the patient monitor. User preference was found to require the wave freeze function, the display of menus that are not measured/used on the screen, the standby mode function, the Early Warning Scoring function, and the acknowledgment alarm function, except for the drug calculation function.

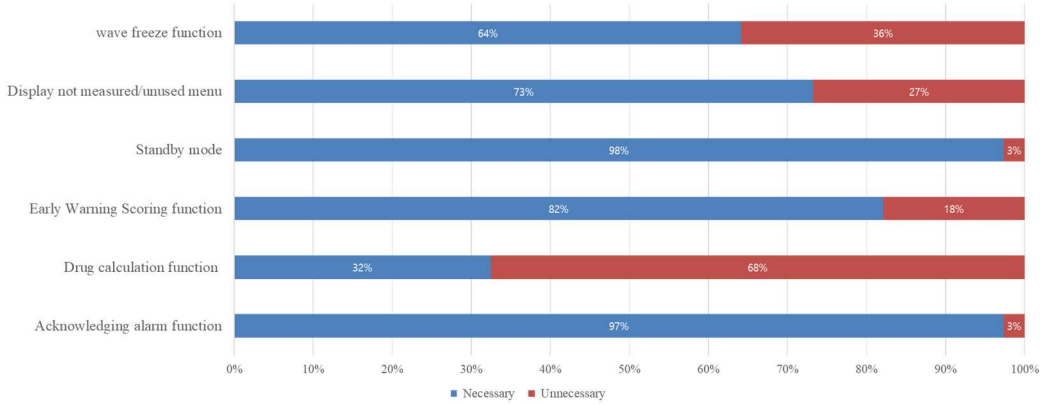


Figure 21. User Preference Survey Results on the Need for Specific Functions.

Figure 22 shows the results of the preference survey for the arrhythmia alarm activation setting of the patient monitor. For the preferred settings for activating the arrhythmia alarm, users preferred its activation for alarms except for missed beat, ST high, ST low, and ST multi-alarm, and missed beat and ST high showed the same user preference for its activation and deactivation.

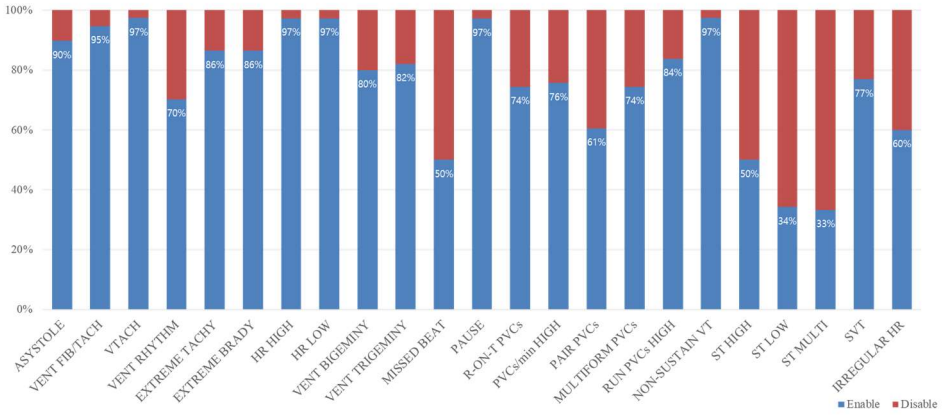


Figure 22. Results of a Preference Survey for Arrhythmia Alarm Activation Settings.

(E) Task Success Rate

The task success level, task success rate, and scenario success rate of the six participants in the usability test are shown in Table 42. The success rate for each use scenario was 83% for patient setup, 79% for waveform/parameter settings, 67% for basic settings, 83% for alarm setting, 78% for arrhythmia alarm settings, 90% for arrhythmia alarms. In addition, It showed a success rate of 88% in screen mode setting and 100% in discharge.

Among the 43 patient monitor tasks performed in the usability test, the task of changing the number of waveforms on the main screen. (Task 3) and the task of deleting the arrhythmia alarm history from the alarm message list (Task 29) were observed with the lowest success rate. It is possible to change the number of waveforms in the display setting of the setup menu, but there were many cases in which the task could not be completed while trying to change the number by pressing the waveform on the main screen. In addition, in order to delete the arrhythmia alarm message list, you need to press the alarm pause button, but many participants thought that pressing the alarm message several times would delete it, and participants felt unfamiliar with deleting the message by pressing the alarm pause button.

Table 42. Success Rate for Each Task and Scenario

Use Scenario	Task Number	Level of success			Success Rate for Each Task	Success Rate for Each Scenario
		C	CI	NC		
Patient Setup	Task1	2	3	1	83%	83%
	Task2	3	2	1	83%	

Use Scenario	Task Number	Level of success			Success Rate for Each Task	Success Rate for Each Scenario
		C	CI	NC		
Waveform/Parameter Setting	Task3	0	2	4	33%	79%
	Task4	3	1	2	67%	
	Task5	3	3	0	100%	
	Task6	5	0	1	83%	
	Task7	6	0	0	100%	
	Task8	4	2	0	100%	
	Task9	1	3	2	67%	
Basic Setting	Task10	2	2	2	67%	67%
	Task11	3	0	3	50%	
	Task12	5	0	1	83%	
Alarm Setting	Task13	2	3	1	83%	83%
	Task14	6	0	0	100%	
	Task15	5	0	1	83%	
	Task16	6	0	0	100%	
	Task17	6	0	0	100%	
	Task18	4	0	2	67%	
	Task19	3	1	2	67%	
	Task20	2	2	2	67%	
Arrhythmia Alarm Setting	Task21	2	3	1	83%	78%
	Task22	5	1	0	100%	
	Task23	3	0	3	50%	
Arrhythmia Alarm Occurs	Task24	6	0	0	100%	90%
	Task25	6	0	0	100%	
	Task26	6	0	0	100%	
	Task27	6	0	0	100%	
	Task28	6	0	0	100%	
	Task29	1	1	4	33%	
	Task30	6	0	0	100%	

Use Scenario	Task Number	Level of success			Success Rate for Each Task	Success Rate for Each Scenario
		C	CI	NC		
Arrhythmia Alarm Occurs	Task31	6	0	0	100%	90%
	Task32	6	0	0	100%	
	Task33	6	0	0	100%	
	Task34	4	0	2	67%	
	Task35	4	1	1	83%	
Display Mode Setting	Task36	5	1	0	100%	88%
	Task37	6	0	0	100%	
	Task38	3	1	2	67%	
	Task39	6	0	0	100%	
	Task40	6	0	0	100%	
	Task41	4	1	1	83%	
	Task42	3	1	2	67%	
Discharge	Task43	6	0	0	100%	100%

(F) Satisfaction

The satisfaction results for each use scenario of the participants are shown in Table 42. The function that received the highest satisfaction score was to change the settings such as the speed and size of the waveform in the waveform/parameter settings, with a score of 6.33. The function that received the lowest satisfaction score was 4.50 for the alarm occurrence condition and alarm message retention period for each type of PVC and Graphical trend screen. There were many opinions that it would be nice if the alarm message matched the rhythm displayed on the current Wave. There were many opinions that the graphic trend screen would not be used because it was easy to operate, but the graph was not easy to see.

Table 43. Satisfaction Result of Participants

Use Scenario	Satisfaction Score							
	N1	N2	N3	N4	P1	E1	Mean	Satisfaction
Patient Setup	6	5	5	3	6	4	4.83	4.92
	6	4	5	3	7	5	5.00	
Waveform/Parameter Setting	6	4	6	6	5	4	5.17	5.64
	6	5	7	5	6	5	5.67	
	6	5	7	5	7	5	5.83	
	6	6	7	6	7	6	6.33	
	6	6	7	3	6	6	5.67	
	6	6	4	6	4	5	5.17	
Basic Setting	7	6	5	3	3	4	4.67	5.08
	6	5	5	5	7	5	5.50	
Alarm Setting	6	6	6	6	6	4	5.67	5.61
	6	5	5	6	6	5	5.50	
	6	6	5	7	6	4	5.67	
Arrhythmia Alarm Setting	6	6	5	5	6	5	5.50	5.00
Arrhythmia Alarm Occurs	7	4	3	5	6	5	5.00	4.96
	6	4	4	4	5	6	4.83	
	3	6	2	6	5	5	4.50	
	6	6	5	5	7	4	5.50	
Display Mode Setting	6	5	4	5	7	6	5.50	5.07
	6	4	3	6	6	6	5.17	
	6	6	3	6	6	4	5.17	
	6	3	2	5	5	6	4.50	
	6	4	3	4	7	6	5.00	

The satisfaction results of 6 people on the 23 detailed satisfaction questionnaires for each use scenario showed less satisfaction than the target in 5 cases, as shown in (Figure 23), when compared with the target score of 5 points.

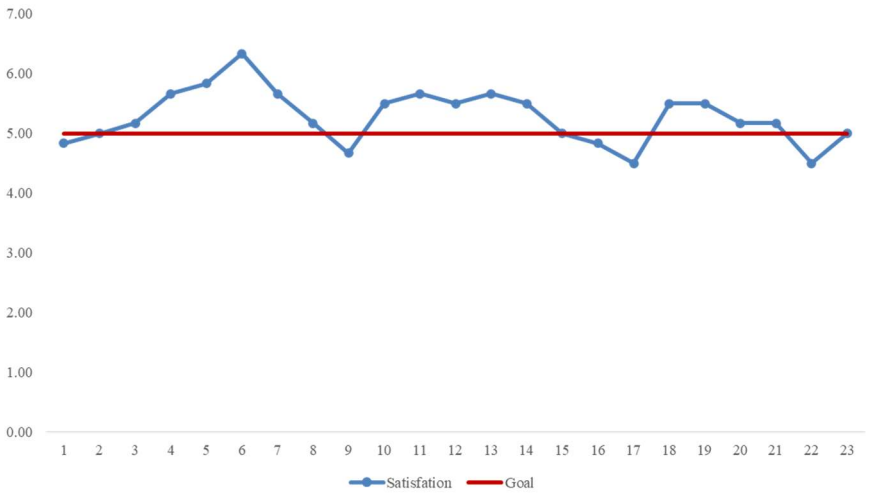


Figure 23. Results of 23 Satisfaction Questionnaires of Participants.

When the satisfaction results for each use scenario were analyzed, patient setup satisfaction was 4.92, and arrhythmia alarm occurs satisfaction was 4.96, lower than 5.00 (Figure 24). Since the satisfaction with patient admit/discharge was 4.83 points, the satisfaction result of the patient setting scenario was lower than the goal. Satisfaction with the operation of the pause button according to the risk of arrhythmia was 4.83 points, and satisfaction with the alarm occurrence condition and alarm messages for PVC type was 4.5 points, so the satisfaction result of the arrhythmia alarm occurrence scenario was also lower than the target.

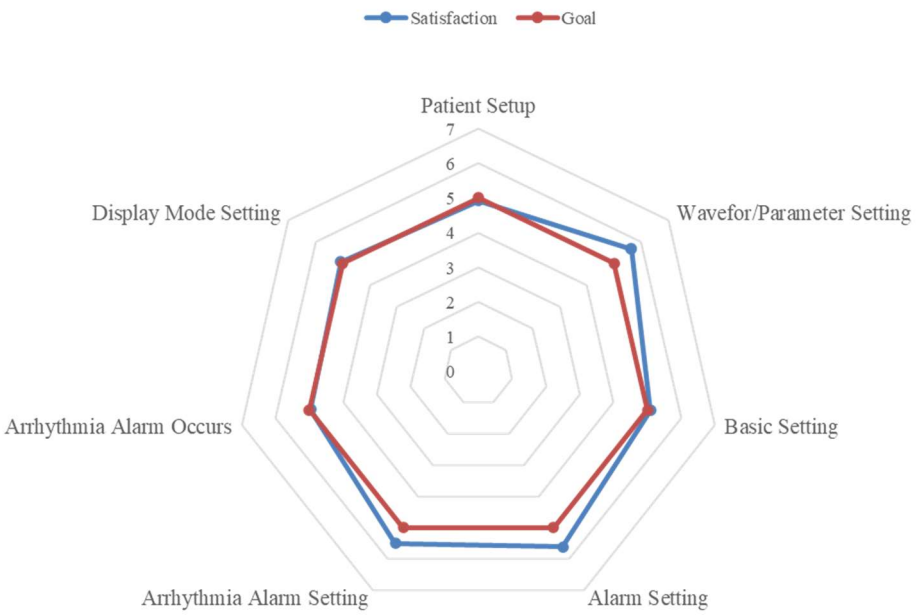


Figure 24. Satisfaction Result of Participants for Use Scenarios.

As a result of collecting participants' opinions on each satisfaction questionnaire for the analysis of satisfaction results, when modifying the contents during the patient registration process, partial modification is not possible and must be filled in again from the beginning, and when entering the date of birth, the dot is automatically created, but there was an opinion that it was confusing because it was possible to enter it directly. In addition, if another part of the screen is touched during patient setup, it is canceled and it is necessary to proceed again from the beginning, so it takes a lot of time and satisfaction seems to be low. In the case of the alarm pause button, the high alarm and medium alarm pause the audible alarm for 60 to 120 seconds, and the low alarm visual and audible alarms turn off. It also has the function of deleting the alarm record except for the currently occurring alarm

in the alarm message window. There was an opinion that an alarm record should not disappear when you press pause alarm, and that a hurdle is needed for deleting the list because it can be erased even when pressed accidentally.

(G) User Preference from Medical Staff and Biomedical Engineer

Table 44 shows the results of a user preference survey to evaluate the user's subjective medical device design preference and use experience after using Patient monitor A for medical staff and Biomedical engineers. All six participants preferred to keep the patient type setting (adult or neonate) between changing every time and maintaining the initial setting according to the use environment (ex. ICU, NICU) the power is turned on.

There were more responses saying that Wave Freeze was not necessary, but there were opinions that it would be good when the medical staff needs the previous data when the module is attached or detached for a short time while moving the patient or when the module is detached due to poor contact.

There were many opinions that the alarm reset function was necessary because new patients require initial settings, and they are modified for each patient from the default. In addition, there was an opinion about the function that asks whether to keep the alarm set value or reset it. Some said that an Alarm off is a function to turn off the audible alarm, and it is necessary because there are cases where unnecessary alarms continuously occur due to noise and there was an opinion that it would be nice to inform the medical staff that it is turned off. In the case of standby mode, all 6 patients responded that it is necessary because

it is a function that is used very often in the intensive care unit because there are cases where patients are away for reasons such as examinations.

The improvement of patient monitor A over other patient monitoring devices used in the intensive care unit was that there were too many items displayed on the main display and they were unfamiliar with terms different from the medical terminology usually used in hospitals. In addition, although alarm sounds are very important due to the noisy nature of the intensive care unit, there was an opinion that the awareness of patient monitor A alarm sounds was low.

Table 44. Result of User Preference and Use Experience Survey

No.	Questionnaire	Number of Responses	
1	Depending on the use environment (ex. ICU, NICU), Please write which one you prefer and why between the patient type (ex. adult, neonatal) settings that are maintained or changed every time the power is turned on.	maintain	6
		change	0
2	Please write whether you need the Wave Freeze feature and why.	necessary	2
		unnecessary	4
3	Please write whether the alarm reset function is necessary and why.	necessary	4
		unnecessary	2
4	Please write whether you need the Alarm Off function and why.	necessary	4
		unnecessary	2
5	Please write whether Standby mode is necessary and why.	necessary	6
		unnecessary	0
6	Have you ever used a telemetry system that measures the patient's ECG signal while the patient is moving, transmits the data to the central station through the wireless network, and analyzes the ECG data?	experienced	3
		never experienced	3
7	If you have any suggestions for improvement of the patient monitor A compared to other intensive care unit patient monitoring devices, please write.		

III. RESULT

1. Optimization of Clinical Evaluation Design

The clinical evaluation model as shown in Figure 25 based on the results of designing and evaluating clinical evaluation in an actual use environment and clinical evaluation in a simulated environment by carrying out a case study of a patient monitoring device to derive a clinical evaluation design optimization model has been configured.

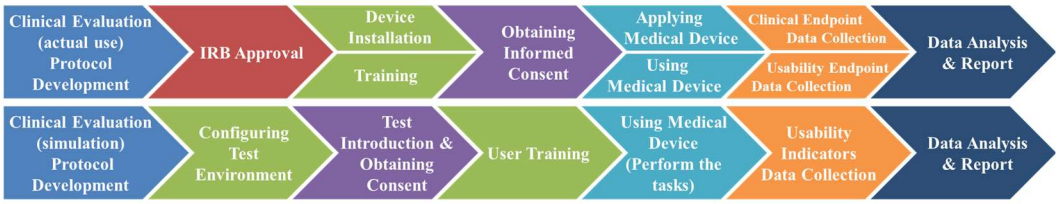


Figure 25. Process of Clinical Evaluation Design Optimization Model.

The derived clinical evaluation optimization model consists of clinical evaluation in an actual use environment and clinical evaluation in a simulated environment. Clinical evaluation in an actual use environment is a model that integrates clinical effectiveness evaluation and usability evaluation and proceeds simultaneously.

In order to conduct clinical evaluation in an actual use environment, it is necessary to develop a clinical evaluation protocol first. The protocol includes a description of the purpose of the evaluation, medical device information, criteria for inclusion/exclusion of participants, evaluation period, and endpoints to evaluate effectiveness and usability.

Participants include both medical staff who use medical devices to perform clinical examinations, treatments, and surgeries, and patients who receive examinations, treatments, and surgeries using medical devices. In other words, from a clinical investigation point of view, the investigator becomes the target for usability endpoints to be evaluated, and the patient becomes the target for clinical effectiveness to be evaluated. Therefore, all criteria for selection/exclusion of the two participants should be written.

Since it is an investigation conducted on human subjects, the clinical trial protocol is developed and then reviewed and approved by the Institutional Review Board (IRB). Documents submitted to the Institutional Review Board include not only protocols, but also subject statements and consent forms. Since there are two subject groups, the subject statement and informed consent form should be developed for both subjects.

In the following procedure, along with medical device installation, clinical evaluation protocols and procedures, and training on how to use the medical device are conducted. Since usability evaluation participants are recruited from investigators conducting clinical effectiveness evaluation, additional device use training for usability evaluation is not required.

Both groups of participants read and sign an informed consent form prior to participating in the assessment. In the evaluation, patients are applied with medical devices, and medical staff uses them. Endpoint data to evaluate clinical effectiveness are collected from patients and endpoint data to evaluate usability are collected from medical staff.

Endpoint data for evaluating clinical effectiveness are determined based on the function

of the medical device. Endpoint data for evaluating usability can use widely used standard usability evaluation indexes such as PSSUQ and NASA-TLX or use developed questionnaires and interviews. Clinical evaluation in actual use environment is completed by analyzing the collected data and writing a report.

Clinical evaluation in a simulated environment evaluates indicators that are difficult to evaluate in an actual use environment through a usability test. The clinical evaluation procedure in the simulated environment is the same as the usability testing procedure. The usability test allows users to evaluate interactions with functions of interest by performing tasks according to use scenarios and can evaluate usability effectiveness and efficiency, such as task success rate and task execution time. In addition, you can perform a satisfaction rating for all functions used while performing the task. Preference surveys can also be conducted in the same way as clinical evaluation in actual use environments.

2. Clinical Evaluation Design Optimization Model Application Criteria

Through the analysis of clinical evaluation case studies of the patient monitor in the cardiology intensive care unit and characteristics of the patient monitor in the cardiology intensive care unit, the application criteria for medical devices to which the clinical evaluation design optimization model can be applied were derived. Figure 26 shows the decision flow chart for applying the clinical evaluation optimization model.

First, determine whether the medical device to be evaluated is a multifunctional medical device. Multifunctional medical devices are complicated to use and error-prone. Use errors

often occur on equipment with complex user interfaces. Intensive care units commonly use multifunctional medical devices. Medical devices used in intensive care units include patient monitors, ventilators. Anesthesia machine is one of the multifunctional devices in the operating room. Second, determine whether urgent decision-making and fast-paced use are required when using the device. In environments such as several types of intensive care units, operating rooms, emergency departments, and ambulances,[10] response times to changes in a patient's status must be prompt.[15] Examples of medical devices that require fast-paced use include defibrillators and automated external defibrillators. In particular, critical care is fast-paced, complex and usually requires urgent high-risk decision-making.[18] Third, determine whether the environment is one where medical staff is treating multiple patients at the same time. The intensive care unit usually cares for 6 to 12 critical illness patients and the emergency room treats 2 to 4 critically ill, emergency, and acutely ill patients per hour.[15,63]

This environment causes a high workload of medical personnel, which can be the root cause of medical errors.[18] Finally, determine whether it is used in an environment with a variety of alarms, lots of noise, and flashing lights. These environments can increase the potential for errors.[15] In conclusion, the clinical evaluation optimization design model can be applied to cases that are complex devices to use and the environment includes fast-paced care, multiple patients, and various alarm noises that can disturb users. Even if the above criteria are not met if it is not possible to simulate key aspects of actual patient treatment in a simulated environment or to simulate some risk-related use scenarios, the clinical

evaluation optimization design model can be used.[10,64] For example, Variability in human physiology can affect the way and technique the user applies the device to the body, and the patient's response to treatment can affect the user's interaction with the device, but due to the use of mannequins, it is not simulated in the general clinical evaluation design model. If all of the above is not applicable, the general clinical investigation design model is used. The general clinical investigation design model means a model that separately performs medical device clinical investigation and usability test.

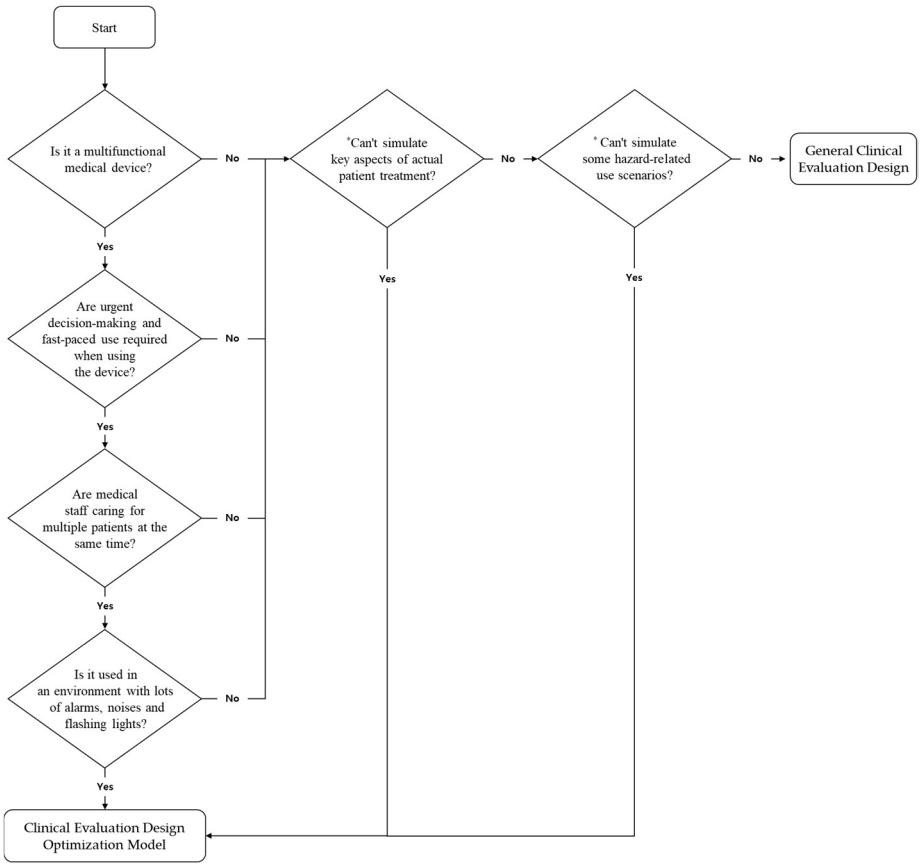


Figure 26. Clinical Evaluation Optimization Model Application Decision Flow Chart.
(*IEC 62366-2:2016 Guidance on the application of usability engineering to medical devices)

IV. DISCUSSION

For a patient monitor case study to derive a clinical evaluation design optimization model, clinical Trials related to clinical effectiveness evaluation and clinical effectiveness endpoints were investigated, patient monitor use scenarios and critical tasks were analyzed, and usability evaluation index was investigated. In the 8 clinical trial cases investigated, it was confirmed that the number of false alarms, false alarm rate, positive predictive value (PPV), negative predictive value (NPV), sensitivity and specificity were evaluated as clinical effectiveness endpoints in previous studies. In this study, we evaluated the false alarm rate, which is the probability that an alarm sounded does not have the condition of interest, but other clinical efficacy endpoints mentioned are also available. In the case of false alarm rates and positive predictions, the arrhythmia alarm that occurred and the electrocardiogram data at the time the alarm occurred are collected and analyzed. However, in order to use sensitivity, specificity, and negative predictive value as effectiveness evaluation endpoints, all electrocardiograms must be collected during the entire patient monitoring period, including the time when no arrhythmia alarm occurred.

As a result of analyzing patient monitor use scenarios and important tasks, 8 usage scenarios and 43 tasks performed in each usage scenario were divided, and tasks in alarm-related usage scenarios were selected as important tasks. The investigated usability indicators included task success rate and satisfaction, system usability, workload, and user preference questionnaire. The task success rate is a usability indicator that can be used for usability testing in simulated environments rather than actual use environments where it is

challenging to observe users, since completion, user errors, and close calls observed while performing tasks should be recorded. Other indicators can be used in evaluation methods other than usability tests, and standard indicators can be used, or new questionnaires can be developed and used.

Clinical evaluation of patient monitor was designed based on clinical investigation case analysis of patient monitors, clinical effectiveness endpoint investigations, use scenarios and task analysis, usability evaluation indicators investigations, and standards related to clinical evaluation such as ISO 14155:2020 and IEC 62366-2:2016. The clinical evaluation consists of clinical evaluation in an actual use environment and in a simulated environment. Clinical evaluation in an actual use environment is an evaluation that combines clinical investigation and usability evaluation. The Guidance on the application of usability engineering to medical devices notes that supplemental user interface evaluation may be required in an actual use environment when simulated use is not adequate to explore some hazard-related use scenarios.

In the study, clinical evaluation was designed to evaluate clinical effectiveness and usability at the same time by integrating usability evaluation of medical devices, which requires evaluation of user interfaces in actual use environments, with the clinical investigation. Clinical evaluation of the simulated use environment is evaluated by performing usability tests on usability indicators that cannot be evaluated in the actual use environment. As a result of conducting a clinical evaluation to confirm the applicability of the derived clinical evaluation design, it was proved that there was no difference in the effectiveness of false alarms between the two patient monitors and that the system usability

and workload of patient monitor B were superior. In addition, it was possible to evaluate the task success rate and satisfaction of patient monitor device A, and to collect opinions on the design preferred by users. Errors and problems that occur during actual use were also identified. Problems identified include small alarm text that makes it difficult to quickly determine when the alarm rings, short pause time for the alarm, patient monitor taking up too much space, the patient monitor does not recognize the touch well, the patient movement causing a lot of alarms, sometimes SPO_2 is not detected correctly, and false apnea alarm frequently occurs.

In this clinical evaluation design optimization model, clinical evaluation in an actual environment can save time and resources by combining usability and clinical trials to proceed simultaneously.[64] For example, since IRB approval is obtained at once, the period can be shortened rather than proceeding with IRB approval for each evaluation. In addition, if medical staff participating in clinical research as researchers agree to participate in usability evaluation, they participate in usability evaluation and perform tasks, so there is no separate recruitment process for usability evaluation, which shortens the period. It can also save time and resources required to build a simulation environment to realize the actual use environment.

As a result of the study, a flow chart for determining medical devices to which the clinical evaluation design optimization model can be applied is presented so that the model can be used to evaluate the effectiveness and usability of medical devices. The application of the clinical evaluation design optimization model is premised on the need for usability evaluation in an actual use environment. Medical devices used in operating rooms, various

types of intensive care units, emergency rooms, and ambulances have limited assessment of several risk-related use scenarios, including the ability of the alarm system to draw users' attention in noisy and busy use environments.[64] The proposed clinical evaluation model is expected to be applied to various medical devices such as patient monitors, ventilators, infusion pumps, anesthesia machines, defibrillators, and automated external defibrillators used in environments such as operating rooms, various types of intensive care units, emergency rooms, and ambulances to evaluate effectiveness and usability. The clinical evaluation design optimization model proposed in this study can perform evaluation considering the main aspects of patient treatment or hazard-related use scenarios. Therefore, errors and problems occurring during actual use can be identified, and more accurate clinical effectiveness and usability evaluations will be possible.

V. CONCLUSION

In this study, a clinical evaluation design optimization model and criteria for determining medical devices applying the optimization model were proposed through a case study of the clinical evaluation design of a cardiology intensive care unit patient monitor. For the clinical evaluation design of the cardiology intensive care unit patient monitor, similar clinical studies were investigated to derive the clinical trial design and clinical effectiveness evaluation variables, analyze the use scenarios and tasks of the CCU patient monitor, and investigate usability evaluation indicators. Based on this, the clinical evaluation of the CCU patient monitoring device was designed and evaluated, and a clinical

evaluation design optimization model was proposed. In addition, I presented criteria for determining medical devices that can utilize clinical evaluation design optimization model.

The clinical evaluation design optimization model proposes to perform usability evaluation in the clinical investigation environment. Since the evaluation performed so far could not simulate all hazard-related scenarios associated with medical devices, complete assurance of acceptable risk could not be provided, and adverse events still existed. Therefore, it is necessary to conduct the clinical evaluation in the actual use environment, which is the usability evaluation in the clinical investigation environment. The clinical evaluation design optimization model proposed in this study is more accurate because evaluation can be performed by considering the main aspects of patient treatment or hazard-related use scenarios. In addition, it can save time and resources required for evaluation.

The proposed clinical evaluation design optimization model can be used to evaluate the clinical effectiveness and usability of medical devices used in environments that are complex to use, fast-paced care, multiple patients, and a variety of alarm noises. It is expected to be used for the clinical evaluation of various medical devices such as patient monitors, ventilators, infusion pumps, anesthesia machines, defibrillators, and automatic in vitro defibrillators to evaluate more accurate clinical effects and usability.

REFERENCES

1. Clinical Evaluation: A Guide for Manufactures and Notified Bodies Under Directives 93/42/EEC and 90/385/EEC, MEDDEV 2.7/rev.4. Health Technology and Cosmetics, European commission 2016.
2. Pritchard G. Clinical Evaluation Reports from the medical writer's perspective Medical Writing 2017;26:14-9.
3. Burches E, Burches M. Efficacy, effectiveness and efficiency in the health care: the need for an agreement to clarify its meaning. Int. Arch. Public Health Community Med 2020;4:35.
4. ISO 14155: 2020 Clinical investigation of medical devices for human subjects-Good clinical practice. International Organization for Standardization; 2020.
5. Medical Device Regulatory Science (RA) Expert Vol. 4 Clinical Trials: National Institute of Medical Device Safety Information; 2021.
6. Medical Device Clinical Trial Guidance Medical Device Clinical Trial Guidance 2020.
7. ISO 9241-11: 2018 Ergonomics of human-system interaction - Part 11: Usability: Definitions and concepts. International Organization for Standardization; 2018.
8. IEC 62366-1: 2015 Medical devices - Part 1: Application of usability engineering to medical devices. International Electrotechnical Commission; 2015.
9. Guidance for Common Standards and Specifications on Electromechanical Safety of Medical Devices - Usability. Ministry of Food and Drug Safety; 2017.
10. IEC TR 62366-2: 2016 Medical devices - Part 2: Guidance on the application of usability engineering to medical devices. International Electrotechnical Commission; 2016.
11. Privitera MB. Applied human factors in Medical Device design: Academic Press; 2019.
12. Michael Wiklund LB, Stephanie Alpert Larsen. Writing Human Factors Plans & Reports for Medical Technology Development: AAMI; 2018.
13. Andrade E, Quinlan L, Harte R, Byrne D, Fallon E, Kelly M, et al. Novel Interface Designs for Patient Monitoring Applications in Critical Care Medicine: Human Factors Review. JMIR Hum Factors 2020;7:e15052.

14. Regulation on Medical Device Groups and Class by Group. Medical Device Products and their Assigned Classes. Ministry of Food and Drug Safety; 2022.
15. Donchin Y, Seagull FJ. The hostile environment of the intensive care unit. Current Opinion in Critical Care 2002;8:316-20.
16. Kam P, Kam A, Thompson J. Noise pollution in the anaesthetic and intensive care environment. Anaesthesia 1994;49:982-6.
17. Effken JA, Loeb RG, Kang Y, Lin ZC. Clinical information displays to improve ICU outcomes. Int J Med Inform 2008;77:765-77.
18. Rothschild JM, Landrigan CP, Cronin JW, Kaushal R, Lockley SW, Burdick E, et al. The Critical Care Safety Study: The incidence and nature of adverse events and serious medical errors in intensive care. Crit Care Med 2005;33:1694-700.
19. Lawless ST. Crying wolf: false alarms in a pediatric intensive care unit. Crit Care Med 1994;22:981-5.
20. Tsien CL, Fackler JC. Poor prognosis for existing monitors in the intensive care unit. Critical care medicine 1997;25:614-9.
21. Aboukhalil A, Nielsen L, Saeed M, Mark RG, Clifford GD. Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform. J Biomed Inform 2008;41:442-51.
22. Tibby SM, Murdoch IA. Monitoring cardiac function in intensive care. Arch Dis Child 2003;88:46-52.
23. Serhani MA, H TEK, Ismail H, Nujum Navaz A. ECG Monitoring Systems: Review, Architecture, Processes, and Key Challenges. Sensors (Basel) 2020;20.
24. Schmid F, Goepfert MS, Reuter DA. Patient monitoring alarms in the ICU and in the operating room. Critical Care 2013;17:216.
25. Chambrin M-C. Alarms in the intensive care unit: how can the number of false alarms be reduced? Critical Care 2001;5:1-5.
26. O'Carroll TM. Survey of alarms in an intensive therapy unit. Anaesthesia 1986;41:742-4.
27. Siebig S, Kuhls S, Imhoff M, Langgartner J, Reng M, Schölmerich J, et al. Collection of annotated data in a clinical validation study for alarm algorithms in intensive care--a methodologic framework. J Crit Care 2010;25:128-35.

28. Chambrin MC, Ravaux P, Calvelo-Aros D, Jaborska A, Chopin C, Boniface B. Multicentric study of monitoring alarms in the adult intensive care unit (ICU): a descriptive analysis. Intensive Care Med 1999;25:1360-6.
29. Muroi C, Meier S, De Luca V, Mack DJ, Strässle C, Schwab P, et al. Automated False Alarm Reduction in a Real-Life Intensive Care Setting Using Motion Detection. Neurocrit Care 2020;32:419-26.
30. Follmann DA. Primary efficacy endpoint. Methods and Applications of Statistics in Clinical Trials: Concepts, Principles, Trials, and Design 2014:779-88.
31. Singh P, Singh N, Singh KK, Singh A. Diagnosing of disease using machine learning. Machine learning and the internet of medical things in healthcare: Elsevier; 2021. p.89-111.
32. Sharma DK, Chatterjee M, Kaur G, Vavilala S. Deep learning applications for disease diagnosis. Deep Learning for Medical Applications with Unique Data: Elsevier; 2022. p.31-51.
33. Guidelines for Clinical Efficacy Evaluation of Artificial Intelligence-based Medical Devices. Ministry of Food and Drug Safety; 2019.
34. Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. Frontiers in public health 2017;5:307.
35. Arjaria SK, Rathore AS, Cherian JS. Kidney disease prediction using a machine learning approach: A comparative and comprehensive analysis. Demystifying Big Data, Machine Learning, and Deep Learning for Healthcare Analytics: Elsevier; 2021. p.307-33.
36. Surgenor SD, Blike GT, Corwin HL. Teamwork and collaboration in critical care: Lessons from the cockpit. Critical Care Medicine 2003;31:992-3.
37. Israelski EW, Muto WH. Human factors risk management as a way to improve medical device safety: a case study of the therac 25 radiation therapy system. Jt Comm J Qual Saf 2004;30:689-95.
38. Crowley JJ, Kaye RD. Identifying and understanding medical device use errors. Journal of Clinical Engineering 2002;27:188-93.
39. Perry SJ. An overlooked alliance: using human factors engineering to reduce patient harm. Jt Comm J Qual Saf 2004;30:455-9.
40. PE MEW, Kendler J, Strochlic AY. Usability testing of medical devices. 2nd ed: CRC press; 2016.

41. Williams M. Radiotherapy near misses, incidents and errors: radiotherapy incident at Glasgow. Clinical Oncology 2007;19:1-3.

42. Wiklund. Usability testing of medical devices. Florida: Taylor & Framcos; 2011.

43. Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. Journal of the American Medical Informatics Association 2016;23:5-11.

44. Albert B, Tullis T. Measuring the user experience: collecting, analyzing, and presenting usability metrics: Newnes; 2013.

45. Finstad K. Response interpolation and scale sensitivity: Evidence against 5-point scales. Journal of usability studies 2010;5:104-10.

46. Lewis JR. Psychometric evaluation of the PSSUQ using data from five years of usability studies. International Journal of Human-Computer Interaction 2002;14:463-88.

47. Savoy A, Patel H, Flanagan ME, Daggy JK, Russ AL, Weiner M. Comparative usability evaluation of consultation order templates in a simulated primary care environment. Applied Ergonomics 2018;73:22-32.

48. Görges M, Winton P, Koval V, Lim J, Stinson J, Choi PT, et al. An evaluation of an expert system for detecting critical events during anesthesia in a human patient simulator: a prospective randomized controlled study. Anesth Analg 2013;117:380-91.

49. Schutte J, Gales S, Filippone A, Saptono A, Parmanto B, McCue M. Evaluation of a telerehabilitation system for community-based rehabilitation. International Journal of Telerehabilitation 2012;4:15.

50. Jiang M, Liu S, Gao J, Feng Q, Zhang Q. A Usability Study of 3 Radiotherapy Systems: A Comparative Evaluation Based on Expert Evaluation and User Experience. Med Sci Monit 2019;25:578-89.

51. Dosani M, Hunc K, Dumont GA, Dunsmuir D, Barralon P, Schwarz SK, et al. A vibro-tactile display for clinical monitoring: real-time evaluation. Anesthesia & Analgesia 2012;115:588-94.

52. Morita PP, Weinstein PB, Flewwelling CJ, Bañez CA, Chiu TA, Iannuzzi M, et al. The usability of ventilators: a comparative evaluation of use safety and user experience. Crit Care 2016;20:263.

53. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in psychology: Elsevier; 1988. p.139-83.

54. Hart SG. NASA-task load index (NASA-TLX); 20 years later. Proceedings of the human factors and ergonomics society annual meeting: Sage publications Sage CA: Los Angeles, CA; 2006. p.904-8.

55. Karsh B, Holden R, Alper S, Or C. A human factors engineering paradigm for patient safety: designing to support the performance of the healthcare professional. BMJ Quality & Safety 2006;15:i59-i65.

56. Carayon P. Handbook of human factors and ergonomics in patient safety. Mahwah, NJ: Lawrence Erlbaum; 2012.

57. NASA TLX: TASK LOAD INDEX. National Aeronautics and Space Administration; 2022.

58. Carayon P, Wetterneck TB, Hundt AS, Ozkaynak M, Ram P, DeSilvey J, et al. Advances in Patient Safety Observing Nurse Interaction with Infusion Pump Technologies. In: Henriksen K, Battles JB, Marks ES, Lewin DI, editors. Advances in Patient Safety: From Research to Implementation (Volume 2: Concepts and Methodology). Rockville (MD): Agency for Healthcare Research and Quality (US); 2005.

59. Görges M, Staggers N. Evaluations of physiological monitoring displays: a systematic review. J Clin Monit Comput 2008;22:45-66.

60. Galy E, Paxion J, Berthelon C. Measuring mental workload with the NASA-TLX needs to examine each dimension rather than relying on the global score: an example with driving. Ergonomics 2018;61:517-27.

61. Jiang M, Liu S, Gao J, Feng Q, Zhang Q. Comprehensive Evaluation of User Interface for Ventilators Based on Respiratory Therapists' Performance, Workload, and User Experience. Med Sci Monit 2018;24:9090-101.

62. Jung W-D, Ko J-H, Park J-K, Kwak S-L, Lim S-S. Task Load Analysis of KTX Operation by Using NASA-TLX Method. Proceedings of the KSR Conference: The Korean Society for Railway; 2006. p.1082-7.

63. Jeeraaumponwat T, Srisang N. Emergency Room Physician Staffing to Patient Ratio. 2015.

64. Clark S, Forkey PED. Clinical Actual Use Human Factors Testing. Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care 2018;7:152-4.

ABSTRACT (IN KOREAN)

임상 유효성과 사용적합성 분석을 통한 의료기기 임상평가 설계 최적화

<지도교수 구성욱, 장원석>

연세대학교 일반대학원 의료기기산업학과

김유림

임상평가는 기기가 안전 및 성능에 대한 요구 사항을 준수하는지 확인하기 위한 임상 증거가 충분한지를 평가하는 것이다. 평가에는 기기가 의도한 성능을 달성한다는 것과 디자인이 사용 오류의 위험을 줄이고 의도된 사용자에게 적합하다는 사용성적합성 확인이 포함된다. 이는 임상시험과 사용적합성 평가로 평가되는데 임상평가의 올바른 수행과 의료기기의 안전성과 사용적합성, 유효성을 입증하기 위해 규정이 마련되어 활용되고 있다. 그러나 여전히 실제 의료 환경에서는 의료기기의 문제로 인한 이상사례들이 발생하고 있으며, 임상평가에서 긍정적인 결과를 보이더라도 완전한 안전을 보장할 수 없음을 보여준다. 따라서 본 연구에서는 최적화된 임상평가 설계를 통해 의료기기의 임상 유효성 및 사용적합성을 평가하는 방법을 제안하고 의료기기 임상평가에 이를 적용하기위한 결정 기준을 제시하고자 한다. 임상평가 설계 최적화 모델을 도출하기 위해 중환자실의 중요 구성요소 중 하나인 환자감시장치에

대한 사례연구를 수행하여 임상평가를 설계 및 수행하고 최종적으로 임상평가 설계 최적화 모델과 해당 모델의 적용 기준을 제시하였다. 제시된 임상평가 설계 최적화 모델은 실제 사용 환경에서의 임상평가와 시뮬레이션 환경에서의 임상평가로 구성된다. 실제 사용 환경에서의 임상평가는 임상 유효성 평가와 사용성 평가를 통합하여 동시에 진행하는 모델이다. 임상시험을 수행하면서 사용적합성의 평가도 함께 진행하여 환자의 의료기기 적용과 사용자의 의료기기사용에 따른 임상적 유효성 평가 변수와 사용적합성 평가 변수 데이터를 모두 수집하게 된다. 시뮬레이션 환경에서의 임상 평가는 사용적합성 테스트를 통해 사용적합성을 정성적, 정량적으로 평가한다. 의료기기의 임상평가 시 최적화 설계 모델의 적용 여부를 결정하기 위해서는 의료기기가 다기능의 사용하기 복잡한 기기인지, 의료기기 사용 시 긴급한 의사결정 및 신속한 사용이 필요한지, 의료진이 동시에 여러 환자를 진료하는지, 많은 경보와 소음 및 점멸하는 표시등이 있는지를 고려해야 한다. 제안된 임상평가 모델은 수술실, 중환자실, 응급실, 구급차 등 다양한 환경에서 사용되는 의료기기의 임상평가에 활용되어 평가에 소요되는 시간과 자원을 절약하고 실제 환자 치료의 주요 측면이나 위해관련 사용이 고려된 평가를 수행할 수 있다. 이를 통해 실제 사용 중 발생하는 오류 및 사용상의 문제를 파악할 수 있어 보다 정확한 임상 유효성과 사용적합성 평가가 수행될 수 있을 것으로 기대된다.

핵심어: 임상평가, 임상 유효성, 사용적합성, 의료기기, 환자감시장치