Usability Study of Two Common Defibrillators Reveals Hazards

Rollin J. Fairbanks, MD, MS
Stanley H. Caplan, MS
Paul A. Bishop, MPA, EMT-P
Aaron M. Marks, BA, EMT-P
Manish N. Shah, MD, MPH

From the Departments of Emergency Medicine (Fairbanks, Shah), and Community and Preventative Medicine (Shah), University of Rochester School of Medicine, Rochester, NY; Usability Associates, Rochester, NY (Caplan); Monroe Community College, Rochester, NY (Bishop); and Monroe Ambulance, Rochester, NY (Marks).

Mr. Marks is now with Bridgeborn Government Solutions, Arlington, VA.

Study objective: This usability study evaluates the user interface of 2 common monitor-defibrillators, the Lifepak10 and Lifepak12, to identify use-related hazards.

Methods: Fourteen paramedics familiar with both devices completed 4 EMS simulator scenarios using each device. The scenarios involved “quick look” and monitoring, defibrillation, synchronized cardioversion, and replacing paper. Qualitative and quantitative data were collected, including both participant self-evaluation (scored 1 to 9) and expert observer evaluation (scored 0 to 4).

Results: Participant ratings demonstrated that for performing a quick look, the Lifepak10 was easier to use (mean 8.0 versus 7.1), and for synchronized cardioversion the Lifepak12 was easier (mean 6.7 versus 5.3). Participants performed better on the Lifepak12 than the Lifepak10 for synchronized cardioversion (mean 3.1 versus 1.6) and replacing paper (mean 3.0 versus 2.1). One participant did not complete the final questionnaire. Of the remaining 13, 11 (85%) participants preferred the Lifepak12 for use on a regular basis. Eight (62%) paramedics thought that the Lifepak12 would be more effective in an emergency; 9 (69%) believed that the Lifepak10 is quicker to learn. Paramedics reported difficulty using the devices with gloves and confusion in “sync” mode. Of note, 50% of participants inadvertently delivered an unsynchronized countershock for supraventricular tachycardia.

Conclusion: Although the Lifepak10 is easier to learn, the Lifepak12 is perceived as easier to use and more effective in emergencies. The high failure rate in synchronized cardioversion indicates a need to reevaluate the user interface design for this function. Limitations of this study include the use of simulation. [Ann Emerg Med. 2007;50:424-432.]

SEE EDITORIALS, P. 384 and 433.

INTRODUCTION

Background
Medical providers often depend on medical devices such as monitor-defibrillators in critical and time-dependent situations. It is important that these devices be designed with an emphasis on reducing the potential for adverse events. Although the traditional response to adverse events and near misses in medicine has been to blame the provider, experts in patient safety have demonstrated that there is often a deficiency in a system component, such as the user-interface design of a medical device, that is the actual root cause.

Medical providers interact with medical devices through the user interface, which typically consists of visual and auditory displays (to communicate information to the user) and controls (to communicate instructions to the device). A good user-interface design follows human-factors engineering design standards and takes into consideration the capabilities and limitations of the user, as well as any limitations imposed by the environment(s) in which the device is intended to be used.

The user interface has a surprisingly powerful ability to facilitate and avert hazards. The evaluation of user-interface design is a well-established component of safety engineering in other complex industries, but its role in the medical industry is underrecognized. Usability testing is a method used by human factors engineers to evaluate a device’s user interface and its effect on user performance and safety. Few usability studies are found in the medical literature, and we are aware of none that examine manual monitor-defibrillator devices.
Importance
Emergency medical services (EMS) providers are concerned about adverse events. Monitor-defibrillators are complex medical devices, and their use has been shown to save lives in the out-of-hospital environment, but they also have potential for harm. Identification of use-related hazards and subsequent optimization of user-interface design are essential for safe operation of these devices.

Goals of This Investigation
The objectives of this study were to evaluate and compare the usability of 2 commonly used manual monitor-defibrillators and to identify user-interface-related hazards that may lead to adverse events.

MATERIALS AND METHODS

Study Design
This is a prospective crossover study of paramedic use of manual monitor-defibrillator devices in a simulated EMS environment. In contrast to traditional research methods, in usability testing the object of the research is the device; the goal is not to assess users’ performance but rather to identify design characteristics that could lead to hazards in use. The study was approved by the University of Rochester Research Subjects Review Board.

Setting
Usability testing was conducted at the Monroe County Public Safety Training Facility’s Crime Scene Simulator, a 4-room apartment with an observation deck separated by 1-way mirrors.

Selection of Participants
Fourteen EMS provider participants with experience using both devices were recruited from the local EMS community.

Methods of Measurement and Data Collection and Processing
A detailed description of study procedures has been previously reported, and is summarized in Figure 4.
Participants were familiarized with the simulator and defibrillator devices and instructed to follow the local treatment protocols and to wear gloves.

Participants were presented with 4 scenarios typical for out-of-hospital care: ECG monitoring using a “quick look” technique (through paddles or pads), defibrillation, synchronized cardioversion, and replacing the paper. Quick look is a technique used in EMS to allow a rapid initial assessment of the ECG rhythm, accomplished by placing the paddles or pads on the patient’s chest. Defibrillation is the delivery of a shock that occurs immediately, in contrast to synchronized cardioversion, which delivers a shock at a specific time in the cardiac cycle.

Participants completed all of the following tasks with 1 device and then repeated all tasks on the other. The model used first was randomly assigned.

Task 1: Quick Look and Routine Monitoring: Perform a quick look on an unresponsive patient with a pulse and then monitor the ECG using the chest leads.

Task 2: Defibrillation: Perform a quick look, confirm the presence of ventricular fibrillation, and then deliver 2 defibrillations.

Task 3: Synchronized Cardioversion: Perform 2 sequential synchronized cardioversions on a patient with unstable tachycardia. During the first attempt, the monitor is set to display an exceedingly low R-wave amplitude.

Task 4: Paper Change: Print a 10-second rhythm strip (device is preset to be out of paper).

During the scenarios, participants were asked to “think aloud,” a technique to help observers gain insight into the thought process of participants.

Outcome Measures
Immediately after each task, participants assigned a numeric rating (1 to 9) in response to the question, How would you rate the ease or difficulty of doing this task? (in which 1 is “very difficult” and 9 is “very easy”). Participants were then asked why they chose this rating, and responses were recorded and transcribed. Once all 4 tasks had been completed on 1 device, a postdevice questionnaire was completed before tasks on the next device were begun. This questionnaire collected qualitative data with open-ended questions such as what the participant specifically liked and disliked about using the device, quantitative data assessing the participants’ level of confidence in their ability to effectively use the device for the tasks in the simulation, and an overall device use rating. At the end of the entire session, a final questionnaire was administered to assess perceptions of comparison between devices.

Objective observer ratings were obtained by an EMS physician investigator with human factors engineering training (R.J.F.), who assigned a numeric rating for the success level of each task, 0 (failed) to 4 (excellent). Qualitative data were collected through direct observation, think-aloud comments during the tasks, and follow-up questions during interviews.

Primary Data Analysis
Standard qualitative and usability testing analysis techniques were used.27,28 Recorded interviews, questionnaire data, and observer comments were transcribed, thematically coded, and sorted into groups of similar issues. An inductive content analysis was conducted by a committee of 4 investigators (R.J.F., S.H.C., A.M.M., M.N.S.) and emerging themes were identified. Mean ratings were calculated for the participant rating scales and the observer scales for each task and device model, and confidence intervals around the differences of these means were calculated.

RESULTS
The 14 participants were 21% women and had an average of 9 years’ experience as advanced life support EMS providers (range 1 to 24).
Figure 4. Study procedure.

Random Assignment of First Device

Brief Familiarization of Device

Perform Scenario #1 Quick Look

Task Rating Questionnaire and Interview

Perform Scenario #2 Defibrillation

Task Rating Questionnaire & Interview

Perform Scenario #3 Cardioversion

Task Rating Questionnaire and Interview

Perform Scenario #4 Paper Change

Task Rating Questionnaire and Interview

Post Device Questionnaire

(After all tasks have been completed on both devices)

Final Questionnaire (Comparison of Devices)

Repeat all tasks on second device
Participants’ ease of use ratings are summarized in the Table. There were similar ratings between devices for tasks 2 and 4. The newer Lifepak12 was deemed less easy to use for the routine monitoring task yet easier to use for the more complex synchronized cardioversion task. Observer ratings, also shown in the Table, revealed similar success between devices for the routine monitoring and defibrillation tasks but better success on the newer Lifepak12 for the synchronized cardioversion and the paper-change tasks. The overall ease of use ratings for each device are shown in Figure 5.

The postdevice questionnaires revealed several findings. For both defibrillator models, the synchronized cardioversion and paper-change tasks were rated as the more difficult tasks. Three participants thought that defibrillation was the most difficult task on the Lifepak12, in contrast to none for the Lifepak10. Visibility of information on the display and replacing paper ease of use were both rated higher for the Lifepak12 than for the Lifepak10. Understanding the status information of the display received a favorable rating 3 times for the Lifepak10 compared with 8 times for the Lifepak12, though the means were similar (3.9 and 4.3, respectively).

Figure 5. Participant ratings on “using the overall defibrillator” acquired after all 4 tasks were performed on each device model.

And although 8 (62%) said the Lifepak12 would be more effective in an emergency, 9 (69%) believed the Lifepak10 is easier to learn.

Several themes emerged from the qualitative data and are presented below, organized by topic.

Seven of 14 (50%) participants performed at least 1 unsynchronized defibrillation when they intended to perform a synchronized cardioversion on the patient with SVT. Five of the 7 events occurred on the model that the participant prospectively stated they used most often in their practice. This event occurred only with the Lifepak10 for 4 participants, only on the Lifepak12 for 1 participant, and on both machines for 2 participants, and it occurred most often (but not always) during the second shock. In 5 of the 7 cases, the provider never recognized the mistake. Both units passively reset out of synchronized mode after a synchronized shock is delivered.

Several instances of delayed cardioversion were observed with the Lifepak10 when the participant was not aware that the signal gain was too low to facilitate “marking” of the ECG R wave. Observers noted related problems with the synchronized cardioversion mode feedback. After the participant pushed the “sync” button, the Lifepak10 displayed the word “SYNC” in steady state (eg, not flashing), which seemed to indicate that the machine was in synchronized mode. In fact, a constant display of the word “sync” indicates that the device is in synchronized mode but not ready to deliver a shock, whereas a flashing

<table>
<thead>
<tr>
<th>Device</th>
<th>Lifepak10</th>
<th>Lifepak12</th>
<th>Difference (95% CI)</th>
<th>Lifepak10</th>
<th>Lifepak12</th>
<th>Difference (95% CI)</th>
<th>Lifepak10</th>
<th>Lifepak12</th>
<th>Difference (95% CI)</th>
<th>Lifepak10</th>
<th>Lifepak12</th>
<th>Difference (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant rating “How would you rate the ease or difficulty of doing this task?” 1 (very difficult) to 9 (very easy) Rating (mean)</td>
<td>8.0</td>
<td>7.1</td>
<td>0.9 (0.04-1.7)</td>
<td>6.9</td>
<td>6.9</td>
<td>0.0 (–1.2-1.1)</td>
<td>5.3</td>
<td>6.7</td>
<td>1.4 (0.3-2.6)</td>
<td>6.4</td>
<td>6.9</td>
<td>0.4 (–1.4-0.6)</td>
</tr>
<tr>
<td>Observer rating of task success: 0 (failed) to 4 (excellent) Rating (mean)</td>
<td>3.4</td>
<td>3.1</td>
<td>0.3 (–0.4-1.0)</td>
<td>3.2</td>
<td>3.2</td>
<td>0.0 (–0.8-0.8)</td>
<td>1.6</td>
<td>3.1</td>
<td>1.5 (0.5-2.5)</td>
<td>2.1</td>
<td>3.0</td>
<td>0.9 (0.1-1.7)</td>
</tr>
</tbody>
</table>
display indicates a state of readiness, which caused confusion among some participants, one of whom even attempted to deliver a shock before the device was ready, thus learning by trial that the device was not ready.

Participants made multiple comments about the buttons, suggesting that they sometimes lead to confusion. Comments included reports of difficulty finding the right button and that they look too similar, particularly on the Lifepak12. One participant stated, “It is a busy display, and you have to look around at a lot of buttons to figure out which one you want to push.” Another commented that the controls for the ECG amplitude and the QRS beep volume were easy to confuse.

Participants noted another difficulty with the button configuration: “With the gloves on, it is hard to hit the button in the right place because they are fairly flat and close together.” This was sometimes noted to cause a hazard, such as inadvertent increase of the defibrillator energy level: “One time my thumb slipped and I hit the energy button instead of the charge button.”

During the paper-change task on the Lifepak10, some participants pressed the “record” button repeatedly when the strip did not print. The Lifepak10 has no formal feedback mechanism to notify the user when it is out of paper, so the user must figure this out by trial and error. Also, the button is soft and gives no tactile, auditory, or visual feedback to the user to acknowledge that the input has been received (though paper printing is feedback in normal use). But the Lifepak12 avoided this problem because when the device was out of paper and the user pressed the “print” button, the device emitted a beep and displayed the words “check printer.” An observer noted the following when watching a participant try to figure out why the Lifepak10 was not printing: “He presses the button repeatedly, stares at the screen, and says ‘I’m trying to record here.’” In contrast, the effectiveness of the feedback provided by the Lifepak12 is demonstrated by this think-aloud comment, overheard as the participant was attempting to print a strip: “It’s telling me to check printer . . . Oh, the paper is out.”

Participants had difficulty changing paper on both devices, particularly with gloves on. There were multiple occurrences of participants placing the paper with incorrect orientation (eg, putting it in backwards). Participants noted that the instructional diagram was not helpful, that the device accepted the paper even when it was oriented in the wrong direction, and that once paper was incorrectly placed, recognizing the problem took trial and error. When trying to correct the problem, participants had difficulty removing the incorrectly placed paper, especially with gloves on. Participants said, “It’s difficult for me to get the paper out of there,” and “You have to look down inside to see which way the paper goes, and it’s not real visible,” and “I couldn’t get my fingers in to grasp the paper.” Most participants had to take their gloves off to finish the paper-change task.

Participants found that several functions were difficult with gloves on. In addition to difficulty retrieving a wrongly placed roll (as discussed above), few participants were able to remove the wrapper from the new paper roll with their gloves on. When describing her problems changing paper, one participant said, “The difficulty increased because I had to take my gloves off to get to the paper and open it.” Participants also had difficulty grasping the zipper tabs on the paper storage compartment when they had gloves on. Some of the participants’ gloves actually became entangled in the zipper mechanism. In addition, participants reported that it was difficult to operate the defibrillator paddle controls when wearing gloves.

Participants commented on the difference between the devices in the selection of energy levels. The Lifepak10 easily progressed through each adult dose, but with the Lifepak12, participants had to toggle through several pediatric doses, resulting in the need for several inputs to go from one level to the next during the defibrillation and cardioversion tasks. One participant expressed a typical concern: “I really disliked having to step through energy levels . . . so there is a lot of pushing buttons needlessly.”

Observers noted that the Lifepak10 does not go automatically into paddles mode when the paddles are removed from the case. When the device is powered on, it is programmed to default to lead II (a programmable selection that makes sense for EMS because the majority of use is for simple monitoring). In some cases, the participants did not initially select the paddles mode when they attempted to perform a quick look. Because artifact produced when the device is set to lead II but not attached to the patient can mimic asystole or ventricular fibrillation, the potential for identification of the wrong rhythm exists. This problem is avoided in the Lifepak12 model, which recognizes when the leads are not connected to the patient and displays a message.

Participants and observers noted a predominant problem with tangled leads on both devices, including frustration due to perceived time delays.

Defibrillation is normally accomplished one of 2 ways, by manually applying paddles to the patient’s skin or through hands-free multipurpose patches, which are also capable of monitoring and pacing. Many of the paramedics stated that the ability to control the defibrillator from the paddles is a beneficial tool. However, participants had trouble using the energy select feature, especially with gloves. This problem was exacerbated because the control is found on the left side, which is nondominant for most providers. The hands-free mechanism was praised by participants for allowing the operator to be located remotely from the patient when performing countershocks, likely to decrease the chance of an accidental shock.

LIMITATIONS

Several limitations to our study must be recognized. First, the devices are capable of performing more functions than were tested in our scenarios. However, we chose routine and emergency tasks that are typical of those performed by paramedics and some that incorporate multiple functions. We
did not control for previous experience with the individual
devices in our analyses. But all participants reported familiarity
(and field experience) with both models, and the need to use
both models is a valid reflection on actual practice in our region,
where paramedics frequently work for more than 1 agency and
share equipment during multiple agency responses.

One of the key requirements of usability testing is to observe
the device being used in true environmental conditions.
Although we realistically simulated typical EMS scenarios, we
were not able to conduct the test in all possible environments,
such as inside a helicopter or ambulance or outside in the sun.
For example, the displays were not tested in a high-ambient-
light environment. Using context-appropriate stress levels in
usability testing is also important. Although this study could not
reproduce the stress involved with a true patient care situation,
the more stressful real environment is likely to reveal even more
hazards than we found.

Because it was important to select device models that
participants would be familiar with and because almost all EMS
agencies in our local region use Medtronic/Physio-control
products, we did not test products from other manufacturers.
However, it is likely that interface design issues are present in
other brands of devices, and it was not our intent to compare
usability between manufacturers.

Finally, the usability test scenarios did not involve simulated
pediatric patients. Therefore, our results may not have identified
issues that arise when tasks are performed on children.

DISCUSSION

Traditionally in medicine when an adverse event occurs the
natural reaction is to assign fault to a person. The systems
approach to reducing adverse events emphasizes that latent
factors exist that either facilitate hazards or fail to protect
patients from the effects of hazards. Failures in medical device
user-interface design serve as an example of latent errors that
can lead to adverse events.

Although usability testing has been recognized as an
important component of medical device development, some
experts in medical device usability assert that medical product
manufacturers do not always conduct usability testing early
enough in the development process and instead often identify
usability problems after the final product has been produced,
when design changes are prohibitively expensive.16,29

Published standards for the design of medical devices now
specify a need for usability testing before US Food and Drug
Administration approval, but there is no requirement to make
results available to the consumer.14,30,31 A higher level of
awareness of the existence of usability testing might cause
medical device consumers to request usability testing results
when considering a new product.32 This might in turn drive
change in the industry, which presently has no motivation from
the consumer to produce usability testing results.

One of the key principles in user-interface design is
consideration for the actual conditions under which the
device is intended to be used. Several issues identified by this
study show room for improvement in designing devices for
the EMS environment. For example, EMS providers are
usually wearing gloves during patient care, and tasks such as
changing paper and selecting defibrillation energy levels were
found to be difficult with gloves on, which is significant
because the need for new paper often arises during high-task
times, which use larger amounts of paper, such as during
resuscitations.

The most significant and potentially harmful use-related
hazard noted by observers was the inadvertent failure to perform
synchronized countershocks. This failure occurred most often
during the second shock in the sequence, which suggests that
the devices’ passive reset out of synchronized mode was a
contributing factor (eg, the only indication of mode change is
the silent disappearance of the word “sync” on the display). Our
data demonstrate that the device changes modes (from
synchronized to unsynchronized) without effectively
communicating this change to the user. A solution to this might
be an audible alert or a message on the display that must be
acknowledged by the user before proceeding. This hazard is not
insignificant because unsynchronized countershock has the
potential to cause ventricular fibrillation.33

A related user-interface design factor that confused several
participants was that a constant display of the word “sync” on
the Lifepak10 (and the indicator light on the Lifepak12)
depicted that the device was in synchronized mode but not
ready (eg, a problem state) while flashing indicated a state of
readiness. This display is in conflict with population stereotypes
in our culture, in which a flashing display indicates a problem
condition.34

The monitor-defibrillator devices used in this study are also
commonly found in the emergency medicine setting, where it is
likely that similar problems occur. In contrast to EMS providers
who use the monitor-defibrillator devices for daily ECG
monitoring, most emergency departments (EDs) use installed
bedside telemetry devices for this function and so use the
portable devices only in emergencies. A low frequency of use
can lead to decreased familiarity, so it is possible that some of
the issues identified by this study are magnified in the ED
setting.

The results of this study highlight several successes of the
Lifepak12 user-interface design. The Lifepak12 is the newer-
generation replacement for the Lifepak10, and because
monitoring technology has advanced considerably, the
Lifepak12 includes more monitoring functions than the
Lifepak10. As a result, the user interface is necessarily much
more involved, and without careful human factors engineering,
this could lead to a higher incidence of use-related hazards. The
most significant hazard we observed, inadvertent unsynchronized
defibrillation, occurred much less often on the Lifepak12.
Although participants perceived that the Lifepak10 was easier to
learn, they thought the Lifepak12 had better visibility of
information, better communicated status, and an easier paper-
change mechanism. The majority of participants said they
would prefer to use the Lifepak12 on a regular basis, and most believed it was more effective in an emergency. Thus, the Lifepak12 user-interface design not only eliminated several problems encountered on the older Lifepak10 but also was the preferred device by participants, despite its increased complexity. Although proprietary companies do not generally publish the results of their own usability studies in the medical literature, it is evident that the manufacturer has given attention to the user-interface design of the Lifepak12.

Finally, our data highlight the need for further study of hazards in EMS, beyond that of event reporting, one traditional method of evaluating hazards in an environment. Five of the episodes of inadvertent defibrillation were completely unrecognized by the participants. To learn about adverse events that occur during actual patient care, it will be necessary to use simulation or real-time data collection methods such as videotaping and analysis of event and vital sign data stored by the monitor-defibrillator devices.

In conclusion, the Lifepak10 seems easier to learn, but the Lifepak12 is perceived as easier to use continually and more effective in emergencies. Several use-related hazards were identified, suggesting a need for further focus on user-interface design of medical devices used in the emergency medicine setting. Additional training of medical personnel to compensate for user-interface design problems is not, by itself, the answer to avoiding adverse events in medicine.

The authors wish to thank Jennifer Williams, Sarah Campbell, MS, Peter Bonadonna, EMT-P, and the paramedic participants for their assistance with this study, as well as the Monroe Community College and the Monroe County Public Safety Training Facility for the use of equipment and facilities.

The authors wish to thank Jennifer Williams, Sarah Campbell, MS, Peter Bonadonna, EMT-P, and the paramedic participants for their assistance with this study, as well as the Monroe Community College and the Monroe County Public Safety Training Facility for the use of equipment and facilities.

Supervising editor: Robert L. Wears, MD, MS

Author contributions: RJF and SHC conceived the study, designed the study, and obtained research funding. RJF, SHC, and MNS supervised the conduct of the trial and data collection. RJF, PAB, AMM, and MNS undertook recruitment of participating paramedics. RJF, SHC, and MNS managed the data. All investigators participated in the qualitative analysis. MNS provided statistical advice and RJF, SHC, and MNS analyzed the data. RJF and SHC drafted the article, and all authors contributed substantially to its revision. RJF takes responsibility for the paper as a whole.

Funding and support: By Annals policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article, that may create any potential conflict of interest. See the Manuscript Submission Agreement in this issue for examples of specific conflicts covered by this statement. This research was supported by the University of Rochester Department of Emergency Medicine research seed funding. At the time of the study, Drs. Fairbanks and Shah were supported by funding from the National Institutes of Health (1R41NR009592) and Agency for Healthcare Research and Quality (1U18HS015818), and Dr. Shah, by funding from Health Resources and Services Administration.


Presented in part at the Human Factors and Ergonomics Society 48th annual meeting, September 2004, New Orleans, LA; and the National Association of EMS Physicians meeting, January 2006, Tucson, AZ.

Reprints not available from the authors.

Address for correspondence: Rollin J. (Terry) Fairbanks, MD, MS, NREMTP, Department of Emergency Medicine, University of Rochester Medical Center, 601 Elmwood Avenue, Box 655, Rochester, NY 14642; 585-273-3849, fax 585-275-2092; E-mail Terry.Fairbanks@Rochester.edu.

REFERENCES

15. Wichansky AM. Usability testing in 2000 and beyond. 
16. Welch DL. Human factors analysis and design support in medical 
testing in healthcare organizations. Paper presented at: 
Proceedings of the Human Factors and Ergonomics Society 45th 
annual meeting, Minneapolis, MN. 2001.
error and usability: the relationship between usability problems 
and prescription errors when using a handheld application. Int 
20. Obradovich JH, Woods DD. Users as designers: how people cope 
with poor HCI design in computer-based medical devices. Hum 
successful lay person automated external defibrillator use during 
22. Eames P, Larsen PD, Galletly DC. Comparison of ease of use of 
three automated external defibrillators by untrained lay people. 
23. Fairbanks RJ, Crittenden CN, O’Gara KG, et al. The nature of 
adult and pediatric adverse events and near misses in EMS 
cardiac arrest survival through the inexpensive optimization of an 
existing defibrillation program: OPALS Study Phase II. Ontario 
25. Faulkner L. Beyond the five-user assumption: benefits of 
increased sample sizes in usability testing. Behav Res Methods 
study among paramedics. Proceedings of the Human Factors and 
Ergonomics Society 48th annual meeting; September 24, 2004; 
New Orleans, LA.
27. Pope C, Ziebland S, Mays N. Qualitative research in health care. 
Francis; 1998.
29. Wiklund ME. Usability tests of medical products as prelude to the 
1993.
31. Sawyer D. Do it by design. The FDA Center for Devices and 
Radiological Health (CDRH). Available at: http://www.fda.gov/ 
for medical device purchasing: evaluating patient safety. Medinfo. 
33. Ebrahimi P, Rubin SA. Electrical cardioversion resulting in death 
34. Department of Defense. Design criteria standard for human 
engineering. MIL-STD 1472F, August 1999, Paragraph 5.2.2.3.3.
When Is a Defibrillator Not a Defibrillator? When It’s Like a Clock Radio . . . . The Challenge of Usability and Patient Safety in the Real World

Ben-Tzion Karsh, PhD
Matthew Scanlon, MD
From the Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, WI (Karsh); and the Department of Pediatrics, Medical College of Wisconsin, Milwaukee, WI (Scanlon).

SEE RELATED ARTICLE, P. 424.


Fairbanks et al1 describe how usability testing can be used to uncover medical device design flaws that compromise the safe and efficient delivery of care. Their experience is consistent with other published reports,2-5 which tested readily available technologies and found dramatic human-automation interactions problems that could be linked to poor device design. These findings are likely the tip of a proverbial iceberg of medical technology and devices, that, by design, have limited usability. And yet health care delivery organizations continue to purchase devices and information technologies that lack usability.

This brings up many questions. Why do hospitals and health care providers continue to purchase devices and technology with poor usability? Why don’t manufacturers address these shortcomings? And how, if at all, are regulatory agencies involved in this problem? After all, the US Federal Aviation Administration, Department of Defense, Department of Transportation, Nuclear Regulatory Commission, Department of Energy, and the National Aviation and Space Administration use usability and human-factors engineering design. The full answers to these questions are the subject of textbooks, PhD theses, and international meetings. We examine 3 of the major issues that we believe in part address those questions: lack of understanding usability science, lack of good usability testing practices, and lack of market demand for usability.

UNDERSTANDING OF USABILITY SCIENCE

First, providers, many users, manufacturers, and even regulatory agencies do not necessarily know or understand usability. Anyone who has struggled with programming a DVD player or setting the alarm on a novel clock radio in a hotel room inherently understands the problem of usability but may be unaware that usability science exists. Usability is grounded in the field of human factors engineering,6,7 which is the science that studies human performance capabilities and limitations and designs built systems (eg, medical devices, information technology, scalpels, computer monitors) to support performance needs. Unfortunately, many decisionmakers and designers erroneously believe that “identifying performance needs,” also referred to as “user-centered design,” is common sense. Users themselves often share the same erroneous assumption; this leads to blaming (I can’t believe that person can’t figure it out) and even ironically leads to much self-blaming (Why can’t I figure this out? or If only I had been paying more attention, this wouldn’t have happened) when in fact the real problem was poor design.

Consider a defibrillator; the obvious or common-sense need is for the user to be able to save the life of a patient by correctly diagnosing and treating a life-threatening rhythm disturbance. But a human factors engineer or usability scientist analyzes usability at a much deeper level. He or she realizes there are physical and cognitive performance needs that must be met for all the different environments of use and possible users. The physical performance needs include having buttons, knobs, and dials designed to accommodate the hand and finger sizes of different users to maximize correct usage and minimize incorrect usage or accidental activation. The device should also accommodate lifting and carrying needs by having correctly designed handles and a weight that can be carried by the majority of users. The device needs to be designed so that it can be read (visual sensation) in dark or bright light and the auditory cues (alarms or confirmatory sounds) can be heard in environments with varying noise and with the possibility of vibration in a transport vehicle. Cognitive performance needs may include planning, decisionmaking, attention focusing, pattern matching, problem solving, and many others. The defibrillator can be designed to make all of these tasks easier, and Fairbanks et al clearly demonstrate how readily available devices failed to meet many of these design requirements. Fortunately, dozens of scientifically validated design guidelines
exist to support each of those tasks.8-12 The bad news is that it is not clear how much of this available guidance is used.

Additionally, many people, including health care administrators, clinicians, and manufacturers, believe proper training and compliance with correct use protocols are sufficient for avoiding errors when automation is used. This problem is not a lack of awareness of the existence of a science behind usability but a lack of awareness of the concept itself. This belief, which is not evidence based, brings with it a host of problems. Users get blamed (think user error) for mistakes caused by bad design, and users themselves even think all bad outcomes stemming from human automation interactions are their fault! Norman13 explains that this self-blame phenomenon is the result of misunderstanding in causality, as well as both a learned and taught helplessness. Although training is crucial for effective use of automation, it cannot completely compensate for poor design. Training is unlikely to overcome interfaces that do not conform to population stereotypes for where information is located or what colors mean or what “enter” or “return” means. And all the training in the world will not help someone hear an alert in a noisy environment or read a small display in a vibrating ambulance, especially during a time-critical task. Each of these cases requires better system design.6,14

**USABILITY TESTING PROCEDURES**

A second major issue is the challenge of performing usability testing. Even if one possesses a robust understanding of usability science, good usability testing is not ensured. Dozens of resources describe the plethora of usability methods,15-18 including many specific to medical devices or health information technology.5,19-24 Bad conclusions result when the methods are incorrectly applied. One important consideration for testing requires knowing how to select representative end users and representative environments of use. Whether the studies are conducted by manufacturers or hospitals, if the sample includes only “expert users” or “well-trained users,” the results will likely not be generalizable. Similarly, if the device or information technology will be used in environments with different lighting intensities and sources (fluorescent tube versus sunlight), different levels of noise (intensive care unit versus surgical suite) and even different levels of distraction (emergency department versus general medicine ward), then testing must be conducted in those different environments to determine if and when the device is usable. Putting a handful of subjects in a nice, clean, simulated patient care room and having them use the device there may not simulate the real environments of use, in which the alarms might not be audible, the displays easily visible, or buttons easily activated. The key point here is that usability is not proven by demonstrating that a handful of people can use the device in a given environment. Usability is determined by the interaction among users, the technology, the environment (lighting, noise, vibration, distractions), the task characteristics (time pressure, need for concentration) and the organization (culture, policies). Good usability testing must attempt to mimic these interactions.

Understanding the many human biases that can lead to wrong conclusions during testing is another important consideration for good usability. For example, hindsight bias may lead a tester to conclude that a user error during testing could “obviously” be corrected with better training. The fundamental attribution error may lead to conclusions that it was just the “ignorant” subjects who did not understand how to use the device. These biases and many more exist, and a well-trained tester must be aware of them and not fall victim to them. The pervasive blame culture in health care, which is likely rooted in the fundamental attribution error and similar biases, may reduce the understanding of the contribution of bad design in the same way. These biases even exist in heuristic evaluations and expert testing. Experts in a laboratory setting do not view or use technology the same way as novice users in a real-world environment. Consequently, even if end users, regulatory agencies, and manufacturers all understood the issues of usability and more usability testing was mandated, there is no assurance that proper testing would occur.

**MARKET DEMAND AND REGULATION**

The final major issue affecting usability of medical devices is weak market demand for improvements, in part attributable to the lack of understanding of the importance of usable technology and devices. However, this issue also illustrates the role health care providers and organizations play in the problem of poorly usable devices by largely demanding and purchasing devices that function at the lowest price possible. At the same time, many providers insist on added features that may increase complexity and decrease usability. The combination of purchasing for price while requesting added features, without any requirement for usability, creates a potentially dangerous combination. Although some organizations request highly usable products, they are a minority, and thus, their numbers are insufficient to change device and technology design.

Alternatively, some consumers of these products, primarily health care delivery organizations, may assume they are purchasing “tested” products. The US Food and Drug Administration requires manufacturers to “address the intended use of the device, including the needs of the user and patient,” whereas information technology vendors have no such requirements.25 However, compliance with this regulation may yield very different results, according to their understanding and execution of usability principles. At best, this means that the device does work if used exactly as intended under ideal circumstances. Perfect conditions rarely exist in health care delivery organizations, rendering the possibility that a device will function in its intended manner largely futile.

The lack of current market demand takes on added significance in light of the fact that those most likely to be hurt by poorly usable devices and technology, ie, patients, are not involved in the selection and purchasing processes. Meanwhile,
it is unlikely that providers, clinicians, and even manufacturers will see the direct consequence (patient harm) of their decisions. Ironically, in an environment of increased transparency of medical errors and harm, manufacturers might leverage enhanced safety through improved usability for a market advantage.

What does this all mean? Devices and technology with poor usability are endemic in health care and will remain so until there is a fundamental shift in knowledge that leads to changes in purchasing. Groups such as the US Food and Drug Administration and the Joint Commission probably should demand better product usability at both the manufacturer and consumer levels, but improvements will not happen unless good usability testing practices are implemented. Until then, the next time you see a clock radio, think of the ventilators, defibrillators, and pumps on which your patients depend.

The authors would like to thank Arielle Silver, BA, for her editorial assistance.

Supervising editor: Robert L. Wears, MD, MS

Funding and support: By Annals policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article, that might create any potential conflict of interest. See the Manuscript Submission Agreement in this issue for examples of specific conflicts covered by this statement. Dr. Karsh is funded by the Agency for Healthcare Research and Quality, National Library of Medicine, and the Robert Woods Johnson Foundation. Dr. Scanlon is funded by the Agency for Healthcare Research and Quality, National Library of Medicine.


Reprints not available from the authors.

Address for correspondence: Ben-Tzion Karsh, PhD, Department of Industrial and Systems Engineering, University of Wisconsin–Madison, 1513 University Avenue, Room 4155, Madison, WI 53706; 608-262-3002, fax 608-262-8454; E-mail bkarsh@engr.wisc.edu.

REFERENCES