Human factors engineering: A tool for medical device evaluation in hospital procurement decision-making

Gill Ginsburg*

Biomedical Engineering, Trillium Health Centre, Mississauga, Ont., Canada L5B 1B8
ARAMARK CTS, Toronto, Ont., Canada, M8Z, 5X9

Received 6 October 2004
Available online 8 December 2004

Abstract

A human factors evaluation was conducted to inform hospital procurement decision-making in selecting a general-purpose infusion pump to be used hospital-wide. Three infusion pumps from different vendors were involved in the evaluation, which consisted of two phases: a human factors heuristic assessment of the pumps according to several criteria, and user testing in five clinical areas. The clinical areas were: Oncology, Medical/Surgical, Pediatric, ICU, and Anaesthesiology. Fourteen nurses and three anaesthetists participated in the user testing. Reasonable agreement was observed between results of both phases of the evaluation, and overall results clearly favoured one of the infusion pumps over the others. It is recommended that a human factors evaluation should be performed to influence all hospital procurement decisions when purchasing medical devices, to ensure the best devices are selected for the end users and to ultimately enhance patient safety.

Keywords: Infusion pump; Human factors; Hospital procurement; User testing; Medical error; Patient safety; Equipment evaluation

1. Introduction

According to the recent Canadian Adverse Events Study [1], 7.5% of patients admitted to acute care hospitals in Canada experienced at least one adverse event during the year 2000, with drug- or fluid-related events being the most common type of adverse events, next to those related to surgical procedures. The study also found that 36.9% of those patients had adverse events that could have been prevented, and that between 9250 and 23,750 deaths from adverse events could have been prevented in the year 2000. Human factors engineering (HFE) is frequently being cited as an important method to reduce medical error and adverse events and to increase patient safety, when it is applied to the design and evaluation of medical devices, including infusion pumps [2–10]. HFE is starting to be used to influence medical device procurement decisions in hospitals, to ensure that the safest and most efficient and effective devices are purchased. The University Health Network in Toronto, Ontario used HFE to evaluate electrosurgical units on the market [7]. This process influenced the purchasing decision, and in the end the chosen product was the oldest on the market and had the fewest new features, but it was deemed to be the most usable and had the highest acceptance by clinical users. HFE was also used by the Veterans Health Administration to compare the usability of various infusion pumps in order to inform procurement decisions [11]. Recently, Kelsman et al. [12] analysed the decision-making process for infusion pump selection in a large hospital, and they emphasized the need to incorporate HFE into procurement decisions. Health Canada recently issued a notice on infusion pumps to hospitals, recommending that before selecting a pump, the hospital should determine...
that the manufacturer meets human factors standards [13], to reduce the major safety concerns associated with infusion pumps. The notice indicates that between 1987 and 2003, 291 incidents associated with general-purpose infusion pumps were reported in Canada. This number is likely lower than the actual number of incidents due to under-reporting.

2. Background

Trillium Health Centre is one of Canada’s leading community hospitals, serving over one million residents of Mississauga, south Etobicoke, and surrounding areas across two sites. In March 2004, Trillium’s existing contract with a general-purpose infusion pump vendor was to expire. Several months prior to this date, it was identified that a new contract would be sought for general-purpose infusion pumps that incorporated safety features and utilized the latest “smart pump technology,” such as dose error reduction systems and automated programming. In addition, it was deemed important, both from a patient safety and cost effectiveness perspective, that pump usage be standardized across the organization so that one vendor’s product be used in all clinical areas. This would ensure that intravenous sets and pumps would not have to be switched when patients are transferred to and from certain units. To this end, a Request For Proposal (RFP) was issued to replace over 500 general-purpose infusion pumps in the hospital, and three vendors were subsequently identified as meeting the RFP requirements (hereon in referred to as Vendors A, B, and C). Interestingly, Trillium’s current contract is with a prior model of Vendor C’s infusion pump (approximately 450 such pumps are used in the hospital) and a prior model of Vendor B’s infusion pump (approximately 70 such pumps are used in the hospital). All three pumps consisted of human–computer interfaces, requiring human interaction. All pumps had drug libraries and dose-error reduction systems.

Following the identification of the three vendors, each vendor was requested to demonstrate their pump technology, including dose-error reduction systems and automated programming if available, according to numerous clinical simulations that were developed by Pharmacy and various Clinical Leaders throughout the hospital. Staff and Management were invited to the demonstrations, following which they rated the usability of various features of the systems that were demonstrated. Two such demonstration sessions were given per vendor, one of which included basic, unsupervised and unstructured hands-on trial by members of the audience. While user participation in hospital medical device selection decision-making is both important and encouraged, there was no strong preference that came out of the demonstration and usability rating process. Each user did not have enough hands-on experience with the features to be able to accurately judge their usability, and the clinical simulations were not necessarily applicable to every user and did not cover each user’s most common and problematic tasks. Also, a structured observation of the users’ hands-on interaction with the pumps was not conducted, which could have revealed valuable information on the errors that were made and what users found frustrating.

As such, another phase of the selection process was instated, wherein a thorough human factors evaluation of the three infusion pumps was conducted by a Human Factors Engineer, so that the best infusion pump system for the end user could be recommended, and in so-doing patient safety could be enhanced. Although many hospitals face competing financial constraints, pressures from different stakeholder groups, various organizational considerations, and other factors during their decision-making processes [12], the outcome of the human factors evaluation formed the basis for the final decision of which system to purchase, as Trillium Health Centre is committed first and foremost to patient safety and patient-centred care. The purpose of this paper is to share methods used in the human factors evaluation and to illustrate how these methods led to results which clearly showed which system would have the largest impact on improving patient safety. Human factors medical device evaluations, using similar methods to the ones described in this paper, should form the basis of all hospital procurement decisions, to mitigate risk, reduce medical errors, and enhance patient safety.

3. Methods

Two phases of the human factors evaluation of the infusion pumps were conducted. The first phase involved a heuristic evaluation of each pump according to four sets of criteria. Fifteen human factors principles formed the basis of the first set of criteria. These principles were adapted from a set of human factors principles developed by Nielson [14], a set of rules developed by Schneiderman [15], as well as from specific considerations in this study, such as alarm audibility. The 15 human factors principles were:

- Visibility of system status,
- Match between system and world,
- User control and freedom,
- Consistency and standards,
- Error prevention,
- Recognition rather than recall,
- Flexibility and ease-of-use,
- Aesthetic and minimalist design,
- Help users diagnose and recover from errors,
- Help and documentation,
• User efficiency,
• Anticipate user action,
• Color blindness,
• Readability,
• Audibility.

The second set of criteria included standards, related to human factors, established by regulatory bodies, such as the Food Drug Administration (FDA) [2,3], Institute for Safe Medication Practices (ISMP) [16], Health Canada [13], and ECRI [17] for medical devices in general or specifically for infusion pumps. These standards included the usability and content of the drug libraries, the information that is displayed by the devices, and the ability of users to override various limits and warnings.

The third set of criteria was adapted from a usability rating form developed in the initial phases of the infusion pump selection process, and contained characteristics such as:

• Set easy to prime and load,
• Easy to titrate flow rates,
• Key information visible on one screen,
• Visual indicators to show pump is infusing,
• Easy to view battery level,
• Easy to view pressure level,
• Pump lightweight and easy to transport.

The fourth and final set of criteria was developed from the RFP requirements that were sent out to all the vendors before the selection process was commenced, and included requirements such as:

• Pump has large graphics to indicate flow rate and volume,
• Pump keypad has tactile feedback,
• Pump has an audible indication when a key has been activated successfully,
• Pump has a visible channel indicator for alarm conditions, and
• Pump occlusion threshold is easily user adjustable in mmHg.

The Human Factors Engineer rated each infusion pump according to each criterion in the four sets of criteria on a scale of one to five, with one being unacceptable from a usability perspective, and five being excellent. These were overall ratings based on the usability violations that were found to each criterion and the severities of those violations. The scores within each set of criteria were then added up for each pump, as well as the total score of each set of criteria to give each pump an overall usability score. In addition, to the usability ratings, the Human Factors Engineer also performed a task analysis [3,18] for several common tasks that a user would perform with an infusion pump. These included:

• Turning on and off the pump,
• Loading and unloading a set,
• Programming a basic infusion,
• Programming a drug infusion from scratch,
• Programming a drug infusion from a drug library, and
• Programming an infusion on a secondary line.

The task analysis consisted of an assessment of the user’s action to perform each step of the task, the information provided by the device, the information not provided by the device, the device response, the observed problems and the severity of those problems (low, medium, or high from a patient safety perspective). Together, the information gained from the usability ratings and the task analyses not only served to illustrate the usability strengths and weaknesses of the pumps, but also helped to shed light on why users made certain errors during the user testing phase of the human factors evaluation.

The second phase of the human factors evaluation consisted of user testing in which the Human Factors Engineer visited different clinical areas with the pumps and observed users as they performed realistic clinical scenarios with each infusion pump. In total, five clinical areas participated in the user testing, and there were 17 participants in total: four nurses from Oncology, four nurses from Medical/Surgical, three nurses from Pediatrics, three nurses from Intensive Care (ICU), and three anaesthetists from Anaesthesiology. All 17 participants had prior experience in programming a previous but very similar version of Vendor C’s product, and all clinical areas except for Anaesthesiology currently use Vendor C’s previous product. The only difference between the previous version of Vendor C’s product and the version under consideration, in terms of programming, is the addition of barcode scanning capability. Furthermore, one participant from the ICU and all anaesthetist participants had prior experience in programming a previous, but very similar, version of Vendor B’s infusion pump. There is no significant difference between programming the prior version of Vendor B’s product and the current version under consideration. None of the participants had any previous training or experience in programming Vendor A’s pump.

The clinical scenarios were developed by each clinical leader specifically for his/her own area, based on the most common tasks that users in that area perform with infusion pumps. These tasks included a combination of programming basic infusions, piggyback infusions, and drug infusions, as well as other tasks. Each user participated in three sessions, one for each pump. The same scenarios were used for each pump, and the order of the pumps across the clinical units was counterbalanced.
All nurse participants completed all sessions within a day and a half, with either their first session in the afternoon of one day and their second and third sessions in the morning and afternoon of the next day respectively, or their first two sessions in the morning and afternoon of one day and their third session in the morning of the next day. Each session took approximately 40 min. The anaesthetists completed all three sessions consecutively, within approximately 90 min, due to scheduling difficulties. At the beginning of each session, participants received approximately 10 min of basic training on the pump to be programmed in that session. At the end of each session, participants completed a form to rate the usability of a set of characteristics related to the tasks they had performed, and at the end of the third session participants filled out a comparison form to indicate which pump they preferred based on each characteristic, as well as overall.

Due to scheduling challenges as well as the nature of work that the participants had to put on hold while participating in the sessions, it was not possible to strictly control the test environment. The Oncology sessions were conducted in a quiet patient lounge, although the paging system was still audible and the participants were occasionally interrupted by nurses who were covering for them. In the Medical/Surgical and Pediatric areas, the sessions were conducted in the nursing lounges, which occasionally saw some traffic by other nurses. The ICU testing conditions were the most representative of actual conditions, as the participants performed the scenarios either inside or right outside of the patient rooms, always keeping an eye on their patient and sometimes having to go and attend to them during a scenario. In Anaesthesiology, the sessions were conducted in the Anaesthesia Office, in which other Anaesthetists were usually on break or in conversation, or monitoring their patients in the operating room via remote patient monitors.

While the participants performed the scenarios, the Human Factors Engineer observed and recorded in writing the usability errors that were committed. A usability error was defined as any action deviating from the correct or most accurate or efficient programming sequence. In the data analysis, usability errors were further broken up into critical and non-critical errors. A critical error was defined as an error that could have led to a negative consequence for a patient (such as selecting the wrong concentration of a drug from the drug library), whereas a non-critical error was defined as an error that would not have led to a negative consequence for a patient (such as trying to enter a value before selecting the parameter). Critical errors were also further divided into detected and undetected errors. A detected error could have resulted from an infusion pump alerting the user to the error via a visual alert or audible alarm, or by the user simply realizing the mis-

4. Results and discussion

Table 1 summarizes the results of the human factors heuristic phase of the study. There were various usability issues associated with each infusion pump. Vendors A and B seemed to have comparable total scores, with Vendor A slightly higher and Vendor C lagging behind.

Table 2 illustrates examples of the usability strengths and weaknesses that were found with each pump. The heuristic assessment revealed that certain strengths of one pump were weaknesses of another, and vice versa. At first glance, some of the weaknesses illustrated in Table 2 may actually seem like strengths, but it depends on the perspective from which one views them, because sometimes there is a trade-off between safety and efficiency. For example, although having the new location/new patient screen automatically disappear in Vendor B’s product is more efficient in that a user does not have to make a selection at all if they are programming for the same patient in the same location, it is easy to miss this screen altogether if one is occupied with something else after turning on the pump (such as priming the set) or if one gets interrupted [8], and hence the wrong drugs and dosing limits could be loaded into the drug library.

Fig. 1 illustrates the total number of usability errors committed by users in each clinical area for each infusion pump, and shows that substantially fewer errors were made in the Oncology, Medical/Surgical, and Pediatric areas with Vendor A’s infusion pump compared to the other two vendors’ pumps. In all three of these units, nurses program a previous but very similar version of Vendor C’s infusion pump multiple times per shift, and had no previous experience in programming Vendor

<table>
<thead>
<tr>
<th>Set of criteria</th>
<th>Vendor A</th>
<th>Vendor B</th>
<th>Vendor C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human factors principles (15a, 75b)</td>
<td>56</td>
<td>57</td>
<td>47</td>
</tr>
<tr>
<td>ECRI, FDA, ISMP, and Health Canada standards (11, 55)</td>
<td>48</td>
<td>48</td>
<td>38</td>
</tr>
<tr>
<td>Usability rating form (18, 90)</td>
<td>80</td>
<td>78</td>
<td>70</td>
</tr>
<tr>
<td>RFP requirements (17, 85)</td>
<td>76</td>
<td>69</td>
<td>73</td>
</tr>
<tr>
<td>Total usability score (61, 305)</td>
<td>260</td>
<td>252</td>
<td>228</td>
</tr>
</tbody>
</table>

a The number of criteria in the set.

b The maximum possible score if the pump received a five on each criterion in the set.
A's pump. This shows that user experience cannot overcome poor device design. Similar numbers of usability errors were made with the pumps in ICU and in Anaesthesiology.

As shown in Fig. 2, a similar trend was observed for the total number of critical usability errors committed by participants in the Oncology, Medical/Surgical, and Pediatric areas, and as well, slightly fewer critical errors were made in ICU and Anaesthesiology with Vendor A's product versus the other two infusion pumps.

The number of undetected critical errors made in each clinical area is shown in Fig. 3, and again a similar trend is observed. Fig. 4 illustrates the total number of usability, critical, and undetected critical errors committed for each infusion pump across all clinical areas. Overall, performance was better in all error categories with Vendor A's device than with the other vendors' infusion pumps.

Table 3 shows overall user preferences across all clinical areas for various infusion pump tasks or characteristics. It is interesting that these preferences do not always reflect user performance. In this study, this phenomenon may be attributed to a few factors: (a) all users were already familiar with Vendor C's infusion pump,
and some had experience with Vendor B’s infusion pump, (b) there are usability strengths and weaknesses associated with each device, and (c) the tasks or characteristics for which users were stating their device preferences were related to ease-of-use rather than safety. This clearly illustrates the need to perform user testing in which errors are observed and analyzed.

5. Conclusions

There was general agreement between results of both phases of the human factors evaluation. Overall, results favoured Vendor A’s infusion pump over Vendor B or Vendor C’s infusion pumps. This study also illustrates the value of conducting both phases in a human factors evaluation of medical devices, a heuristic assessment phase and a user testing phase, as both phases complement each other and reveal important information. The heuristic assessment phase reveals information on aspects of the design that could potentially be problematic for users and lead to errors or frustrations, but the likelihood and magnitude of those errors can only be seen through user testing. Furthermore, user testing takes into account the specific skills and competencies of users in an organization, especially when conducted with members from all user groups. The errors made by users during the user testing phase can be better understood by the detailed information gained in the heuristic evaluation phase, and that information can then be incorporated in a proactive manner into user training programs on the selected product. Data obtained from the heuristic evaluation phase can also be used to guide design changes and modifications of the devices to improve their usability and safety.

6. Recommendations

It is recommended that the human factors evaluation that was conducted in this study be applied in future medical device procurement decisions conducted at Trillium Health Centre, to ensure that the best products for the end user are selected and to enhance patient safety. Other health care facilities are also encouraged to use this process. Furthermore, it is also recommended that the results of this study for each vendor be made available to that specific vendor, in an effort to influence future product design. Finally, all three pumps had human factors usability issues that caused certain errors to be committed. When user training is conducted throughout the hospital on the selected pump, the training program should be tailored to address these problems so that users become aware of them up front and are better prepared to deal with them.

7. Limitations

There were several limitations to this study, most revolving around time and resource constraints, however it is unlikely that the strong trends observed in the results between Vendor A and Vendors B and C would change if these limitations were to be addressed in future studies. First of all, no novice users (users who had no previous experience in programming infusion pumps) participated in the user testing, as no such novices were available to participate in the study. Also, there was a small sample size of participants tested in each clinical area, and the clinical scenarios did not include all tasks that users perform with the pumps, but was rather a representation of the most common tasks and focused mainly on pump programming. Pump order was not counterbalanced across participants within each area, and testing conditions were not strictly controlled across participants or clinical areas. Finally, observed errors were recorded by hand, but more accurate observation equipment, such as video recording, or an analysis of the key press logs on the pumps, may have revealed errors that could have been missed by the observer.

Acknowledgments

Many thanks to Anu Tuli and Murray Greenwood for their assistance and direction, and to Elizabeth Ever- son, Zelia Campos, Anne-Marie Harper, Jo Forbell, and Dr. Charles Cruise, for their assistance in recruiting participants, scheduling the sessions, and developing the clinical test scenarios. Gratitude is also extended to the participants who took time out of their busy schedules to participate in the sessions.
References


