Development and Use of Heuristics to Evaluate Neonatal Medical Devices for Use in Low-Resource Settings

Jake Johnston
Rice360 Institute for Global Health Technologies
Houston, TX, USA
jtj1@vt.edu

Elizabeth Asma
Rice360 Institute for Global Health Technologies
Houston, TX, USA
ea23@rice.edu

Claudia Ziegler Acemyan
Rice University
Houston, TX, USA
claudiaz@rice.edu

Maria Oden
Rice360 Institute for Global Health Technologies
Houston, TX, USA
moden@rice.edu

Rebecca Richards-Kortum
Rice360 Institute for Global Health Technologies
Houston, TX, USA
rkortum@rice.edu

Philip Kortum
Rice University
Houston, TX, USA
pkortum@rice.edu

Abstract

Previous research has shown that for several domains and environments, developing and using domain-specific heuristics can effectively supplement general heuristics to capture additional usability problems. However, no existing heuristic set specifically addresses medical devices for use in low-resource settings, such as hospitals in low- and middle-income countries. These settings are limited by a lack of critical resources such as medical consumables, equipment, and human resources; therefore, they require devices that meet unique usability challenges.

In this paper, we describe the development and use of domain-specific heuristics to evaluate neonatal medical devices intended for low-resource settings. Five additional heuristics were developed to account for specific usability needs in low-resource settings including cleanability, maintainability and reparability, low workload, minimize discomfort, and access to baby. Three human factors experts independently applied 19 heuristics (14 standard plus 5 new) to 23 neonatal medical device product options. Evaluators identified 36 (9%) additional heuristic violations when applying the newly developed heuristic set focused on low-resource usability needs. Results support the ability of domain-specific heuristics to identify potential usability problems that would not be captured using only standard heuristics.

Keywords

Usability, heuristic evaluation, domain-specific, medical devices, low-resource

Introduction

The assessment of system usability is important in understanding whether products and services can be used by their target audiences. It is important that the user can effectively and efficiently perform required tasks on any given device, and usability helps to measure the extent to which that is true. In some domains, however, the ability to accurately measure the usability of a device takes on particular importance—neonatal medical devices fall into this category.

Design-induced medical device error is a substantial cause of patient mortality in both low-resource and high-resource settings (Dain, 2002), and usability failures for these kinds of devices can lead to significant consequences for the patients who are receiving care. In low-resource hospital settings (such as those in sub-Saharan Africa), medical devices must meet demands that are even more challenging than those found in higher resource settings (McGuire & Weigl, 2014). Difficulty in using medical devices can also cause unintentional device damage or failure. Because of these additional challenges, the penalty for poor device design is much higher in locations with fewer resources, so an accurate assessment of the usability of devices deployed in these environments is essential.

Heuristic evaluation is one human factors/human-computer interaction method that is used to identify potential usability problems. This method involves multiple evaluators who inspect a product and identify aspects of a system that violate accepted usability principles (Nielsen & Molich, 1990). While Nielsen and Molich's heuristic method was originally applied to software (particularly the web), it has also been successfully utilized and modified to assess a wide variety of systems, including medical devices. For example, Zhang et al. (2003) synthesized rules and heuristics developed by Nielsen (1992) and Shneiderman (1998) and then demonstrated their suitability for evaluating medical devices with both hardware and computer interfaces to identify usability problems. Previous research has demonstrated that heuristics have been effectively adapted to capture domain-specific usability issues in many diverse areas (Quiñones & Rusu, 2017), ranging from transportation systems and the physical environment (Shanklin et al., 2020) to interactive digital television systems (Solano et al., 2011).

While these heuristics are highly versatile across different domains and product types, there are situations where unique interaction elements or environments demand additional heuristics to capture all the potential areas where usability failures may occur.

The research presented in this paper builds on the heuristics used by Zhang et al. (2003). In Zhang's work, specially chosen Nielsen-Shneiderman heuristics were used to evaluate different types of infusion pumps. This paper expands on those heuristics to create a specialized set of heuristics that can be used to evaluate a broad range of neonatal medical devices for use in low-resource medical settings. This setting was identified as part of a larger project to select and implement appropriate newborn medical devices in African countries, including Malawi, Tanzania, Kenya, and Nigeria.

Method

Development of the Neonatal Equipment for Low-Resource Settings (NELRS) Heuristic Set

We developed five additional heuristics specifically for Neonatal Equipment for Low-Resource Settings (NELRS) using the general methodology described by Quiñones and Rusu (2017), who emphasized four essential steps: 1) determine specific application features, 2) identify existing sets of heuristics, 3) specify the new heuristics set in a standard template, and 4) validate the new heuristics set. The new NELRS heuristics arose from our experiences working in low-resource hospitals as well as extensive interviews with clinicians living and working in low-resource settings. Each of the newly developed NELRS heuristics had a specific rationale for their inclusion in the expanded heuristic set as presented in the following descriptions.

Heuristic: Cleanability

Rationale: All device surfaces should be easily reachable for thorough cleaning without specialized tools or materials. Neonatal sepsis accounts for a significant percentage of newborn deaths in low-income countries, and medical devices that are difficult to clean could increase infection rates because devices are often used to treat multiple newborns (Rudd et al., 2018).

Heuristic: Maintainability and reparability

Rationale: Device maintenance and repair should require a minimal set of tools and easy-to-follow instructions. Equipment "graveyards" are commonly found in under-resourced hospitals and contain stockpiles of broken devices that failed due to exposure to harsh environmental conditions (dust, humidity, and power surges) and inaccessible spare parts (Richards-Kortum, 2017). However, it has been shown that most medical equipment in low-resource settings can be put back into service using basic repair knowledge (Malkin & Keane, 2010).

Heuristic: Low workload

Rationale: Devices should minimize clinician workload to decrease the risk of fatigue or injury given that hospitals in low-resource settings often treat a high volume of patients relative to the number of clinical staff. Users' physical capabilities and limitations should be considered assuming high patient volumes.

Heuristic: Minimize discomfort

Rationale: Devices should not cause discomfort for either the user or baby, such as repetitive movements, awkward postures, or skin irritation. Systems that cause discomfort may compromise patient care in clinical environments.

Heuristic: Access to baby

Rationale: Users should be able to quickly and easily access or remove the baby from the device. Rapid access to newborns during emergency situations is critical.

Study Design and Materials
Heuristic analysis was performed on 23 product options across eight product categories important for care of small and sick newborns, as summarized in Table 1. Product categories included phototherapy light, radiant warmer, bilirubinometer, pulse oximeter, suction pump, oxygen concentrator, flow splitter, and syringe pump. Two to five unique product options were evaluated within each product category.

Table 1. Number of Evaluated Product Options within Each Product Category

Product category	Product options
Phototherapy light	2
Radiant warmer	2
Bilirubinometer	2
Pulse oximeter	3
Suction pump	3
Oxygen concentrator	3
Flow splitter	3
Syringe pump	5
Total	**23**

Product categories and product options were carefully selected to meet target product profiles for newborn care in low-resource settings, which were defined by a large team of neonatal care experts in collaboration with UNICEF (Kirby & Palamountain, 2020).

The additional NELRS heuristics were combined with the 14 Nielsen-Shneiderman heuristics (Zhang et al., 2003) as suggested by Hermawati and Lawson (2016), resulting in a comprehensive list of heuristics for medical device evaluation (Table 2).

Table 2. Description of Nielsen-Shneiderman/NELRS Heuristics Used in this Study

Heuristic	Description
Consistency and standards	Users should not have to wonder whether different words, situations, or actions mean the same thing. Standards and conventions in product design should be followed.
Visibility of system state	Users should be informed about what is going on with the system through appropriate feedback and/or display of information.
Match between system and world	The image of the system perceived by users should match the model users have about the system.
Minimalist	Any extraneous information is a distraction and a slow-down.
Minimize memory load	Users should not be required to memorize a lot of information to carry out tasks. Memory load reduces users' capacity to carry out the main tasks.
Informative feedback	Users should be given prompt and informative feedback about their actions.
Flexibility and efficiency	Users always learn and users are always different. Give users the flexibility of creating customization and shortcuts to accelerate their performance.
Good error messages	Messages should be informative enough such that users can understand the nature of errors, learn, and recover from errors.
Prevent errors	It is always better to design interfaces that prevent errors from happening in the first place.
Clear closure	Every task has a beginning and an end. Users should be clearly notified about the completion of a task.
Reversible actions	Users should be allowed to recover from errors. Reversible actions also encourage exploratory learning.
Use users' language	The language should be always presented in a form understandable by the intended users.
Users in control	Do not give users the impression they are controlled by the systems.
Help and documentation	Always provide help when needed.
Cleanability [NEW]	Users should be able to easily clean the device with standard cleaning materials available in low-resource settings.
Maintainability and reparability [NEW]	Users should be able to maintain and repair the device with a minimal set of basic tools and instructions.
Low workload [NEW]	Risk of fatigue or injury from physical demands should remain minimal for the users.
Minimize discomfort [NEW]	Device should avoid any type of pain or physical irritation for both users and patients.
Access to baby [NEW]	Users should be able to quickly and easily access or remove baby at any time.

Note: Gray shading indicates Nielsen-Shneiderman heuristics, while no shading indicates newly added NELRS heuristics to address the unique usability challenges of hospitals in low-resource settings.

To assess the severity of each heuristic violation, evaluators used the severity rating scale suggested by Nielsen (1994). Severity differences between medical devices allowed evaluators to rank the devices and decide which ones would continue for further acceptance testing. See Table 3 for severity ratings and their descriptions.

Table 3. Heuristic Violation Severity Ratings and Their Descriptions

Severity rating	Description
0	Not a usability problem at all.
1	Cosmetic problem only. Need not be fixed unless extra time is available.
2	Minor usability problem. Fixing this should be given low priority.
3	Major usability problem. Important to fix. Should be given high priority.
4	Usability catastrophe. Imperative to fix this before product can be released.

Procedure

Three human factors experts independently applied the 19 heuristics (Table 2) to all 23 product options. Evaluators had experience and were familiar with the constraints of low-resource environments. Each of the three evaluators considered functions such as set-up, calibration, clinical use, basic repair troubleshooting, and interactions between medical devices.

We created a heuristics spreadsheet for each evaluator's findings for each device. This data included the location of the deficiency, the heuristic evaluated, a description of the violation, the severity rating, as well as any additional notes needed to describe the deficiency. Post analysis, evaluators made a single comprehensive list of heuristic violations per product option and averaged the severity scores of repeat violations.

Results

In total, the Nielsen-Shneiderman/Neonatal Equipment for Low-Resource Settings (NELRS) heuristic evaluation identified 404 non-duplicative errors across the 23 product options. Of these, 340 (84%) were Nielsen-Shneiderman violations, 36 (9%) were solely NELRS violations, and 28 (7%) were both (see Figure 1).
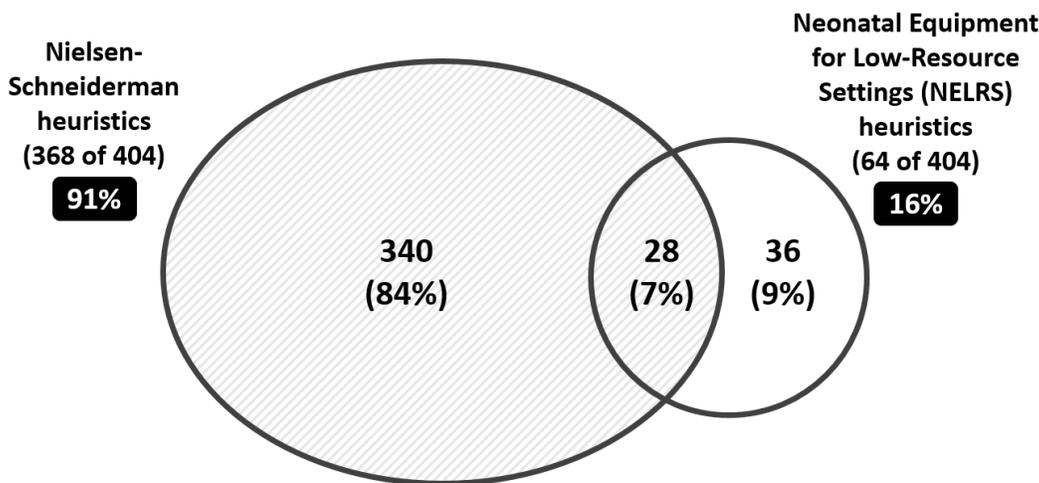

Figure 1. Overlap between heuristic violations using Nielsen's heuristics and the NELRS heuristics.

The number of heuristic violations for each product category are shown in Table 4. Because each product category had a varying number of product options, ranging from two to five, evaluators divided the total violations by number of product options to calculate average violations per product option. Depending on the product category, NELRS heuristic violations accounted for between 7% and 35% of the total identified violations.

Overall, 13% of all identified heuristics were identified by all three evaluators. This is consistent with standard inter-observer agreement (Hertzum et al., 2002; Molich & Jeffries, 2003).

Table 4. Heuristic Violation Counts Per Product Category, Ordered by Average Violations

Product category	# Of product options	# Of violations of only N-S[a] heuristics	# Of violations involving NELRS heuristics	Total violations	Average violations per product option
Flow splitter	3	17 (65%)	9 (35%)	26	8.6
Phototherapy lights	2	20 (87%)	3 (13%)	23	11.5
Oxygen concentrator	3	39 (85%)	7 (15%)	46	15.3
Radiant warmer	2	22 (71%)	9 (29%)	31	15.5
Suction pump	3	36 (77%)	11 (23%)	47	15.6
Bilirubinometer	2	30 (91%)	3 (9%)	33	16.5
Syringe pump	5	114 (93%)	9 (7%)	123	24.6
Pulse oximeter	3	62 (83%)	13 (17%)	75	25.0
Total	23	340 (84%)	64 (16%)	404	-

[a] N-S signifies the Nielsen-Shneiderman heuristic set.

The average severity rating for Nielsen-Shneiderman heuristics was found to be significantly higher (M = 2.3, SD = .84) than the average severity rating for NERLS heuristics (M = 2.1, SD = .79), $t(402) = 2.47$, $p = .01$; $d = .34$. Overall, the addition of domain specific heuristics did not lead to discovering more severe usability problems, and violations of the added heuristics were categorized as less severe on average than those of standard heuristics.

Discussion

Given the large number of usability issues discovered during the heuristic evaluation process, heuristic assessment remains an important tool in uncovering potential difficulties users might encounter with a given device. While some of the items identified by the newly added heuristics were also captured using the Nielsen-Shneiderman heuristics, the addition of the specific heuristics identified 36 unique problems (9%) that would *not* have been discovered had only standard heuristics been employed. This suggests additional NELRS heuristics were necessary to form a complete picture of usability deficiencies present in neonatal medical devices intended for use in low-resource settings.

It might be tempting to believe that because the additional heuristics focused on critical areas specific to the technology, severity ratings associated with identified deficiencies might be higher than severity ratings found using the more general Nielsen-Shneiderman heuristics. Clearly, this was not the case, as the average severity ratings for the added NELRS heuristics were lower than Nielsen-Shneiderman heuristics. This is an important finding because it suggests that the added heuristics were not generated to identify known high-severity errors post hoc, but rather add dimensions along which important deficiencies can be identified, without necessarily increasing the severity of those discovered deficiencies. This finding may be due to the relatively high-severity nature of violations of general heuristics such as "prevent error" in the medical domain.

The average number of deficiencies identified by the NELRS heuristic set showed large variations across device types, ranging from 8.6 violations per flow splitter to 25 violations per pulse oximeter and 24.6 for the syringe pump. These variations are expected and reflect the complexity of the devices evaluated and interface elements (e.g., a display screen with menus) that are present or absent in each of the different device types.

Previous research has demonstrated that specialized heuristics are common (Quiñones & Rusu, 2017) and have been developed in those cases where there are unique interactions or elements of the designs that fall outside the original heuristics proposed by Nielsen and Molich (1990), Nielsen (1992), and Shneiderman (1998). This expansion of heuristics to include additional elements to account for more novel kinds of interfaces is not unexpected, since the original heuristics lists were developed almost exclusively with an eye toward the internet, which was an emerging technology at

the time. The heuristics developed here follow that pattern, with additional evaluation heuristics being added to account for medical device use in novel settings in ways that the baseline heuristics could not capture.

The research described here reinforces the idea that specialized heuristics can be beneficial in identifying usability deficiencies that may have an impact on the ability of the device to be used effectively and efficiently for its intended purpose. If additional heuristic elements are developed properly, using appropriate techniques (e.g., Quiñones & Rusu, 2017), then they are likely to add to the thoroughness of the evaluation and help ensure relevant deficiencies have been identified. That said, the addition of heuristics should not be taken lightly. Evidence regarding the need for enhanced heuristics should be considered, and the list of heuristics should be kept to a manageable number to facilitate expeditious reviews and maintain interrater reliability in the categorization of deficiencies.

Conclusion

These findings support the ability of tailored heuristics to produce value for the evaluation of systems used in particular circumstances and environments. Specifically, the extra benefit gained from using the NELRS heuristics resulted in a more comprehensive usability analysis that considered the specific low-resource settings and unique use environments where these neonatal devices will ultimately be used.

Tips for Usability Practitioners

We offer the following suggestions for practitioners based on the results of the study:

- The addition of specialized heuristics, when developed properly, can be useful in identifying important additional design deficiencies and potential usability issues of unique systems.
- The NELRS heuristics presented here should be used in the evaluation of neonatal medical devices that will be deployed in low-resource settings throughout the world because they find additional system deficiencies not identified by traditional heuristics.
- Specialized heuristics identify additional, but not necessarily more severe, design deficiencies for medical devices.

Acknowledgements

References

Dain, S. (2002). Normal accidents: Human error and medical equipment design. In *Heart Surg Forum, 5*(3), 254–257.

Hermawati, S., & Lawson, G. (2016). Establishing usability heuristics for heuristics evaluation in a specific domain: Is there a consensus? *Applied Ergonomics, 56*, 34–51.

Hertzum, M., Jacobsen, N. E., & Molich, R. (2002). Usability inspections by groups of specialists: Perceived agreement in spite of disparate observations. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems* (pp. 662–663). https://doi.org/10.1145/506443.506534

Kirby, R., & Palamountain, K. (2020). *Target product profiles for newborn care in low-resource settings* (v1.0). UNICEF. https://www.unicef.org/supply/media/2556/file/TPP-newborn-care-final-report-v1-2.pdf

Malkin, R., & Keane, A. (2010). Evidence-based approach to the maintenance of laboratory and medical equipment in resource-poor settings. *Medical & Biological Engineering & Computing*, *48*, 721–726. https://doi.org/10.1007/s11517-010-0630-1

McGuire, H., & Weigl, B. H. (2014). Medical devices and diagnostics for cardiovascular diseases in low-resource settings. *Journal of Cardiovascular Translational Research, 7*(8), 737–748.

Molich, R., & Jeffries, R. (2003). Comparative expert reviews. In *CHI'03 Extended Abstracts on Human Factors in Computing Systems* (pp. 1060–1061). https://doi.org/10.1145/765891.766148

Nielsen, J. (1992, June). Finding usability problems through heuristic evaluation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 373–380). https://doi.org/10.1145/142750.142834

Nielsen, J. (1994). Heuristic evaluation. In J. Nielsen &, R. L. Mack (Eds.), *Usability inspection methods*. John Wiley & Sons.

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 249–256). https://doi.org/10.1145/97243.97281

Quiñones, D., & Rusu, C. (2017). How to develop usability heuristics: A systematic literature review. *Computer Standards & Interfaces*, *53*, 89–122.

Richards-Kortum, R. R. (2017). Tools to reduce newborn deaths in Africa. *Health Affairs*, *36*(11), 2019–2022.

Rudd, K. E., Kissoon, N., Limmathurotsakul, D., Bory, S., Mutahunga, B., Seymour, C. W., ... & West, T. E. (2018). The global burden of sepsis: Barriers and potential solutions. *Critical Care*, *22*(1), 1–11.

Shneiderman, B. (1998). *Designing the user interface: Strategies for effective human computer interaction* (3rd Ed.). Addison-Wesley.

Shanklin, R., Kortum, P., & Acemyan, C. Z. (2020). Adaptation of heuristic evaluations for the physical environment. In *Proceedings of the Human Factors and Ergonomics Society*, *64*(1), (pp. 1135–1139). https://doi.org/10.1177%2F1071181320641272.

Solano, A., Rusu, C., Collazos, C., Roncagliolo, S., Arciniegas, J. L., & Rusu, V. (2011). Usability heuristics for interactive digital television. In *The Third International Conference on Advances in Future Internet* (pp. 60–63).

Zhang, J., Johnson, T. R., Patel, V. L., Paige, D. L., & Kubose, T. (2003). Using usability heuristics to evaluate patient safety of medical devices. *Journal of Biomedical Informatics*, *36*(1–2), 23–30.

About the Authors

Jake Johnston
Mr. Johnston completed a human factors fellowship at Rice360 Institute for Global Health Technologies at Rice University and is now pursuing a PhD in Human Factors of Systems and Product Design at Virginia Polytechnic Institute and State University. His work focuses on user-centered design, medical human factors, and safety.

Elizabeth Asma, MSc
Ms. Asma is Associate Director of Technology Optimization at Rice360 Institute for Global Health Technologies at Rice University. Her work focuses on selecting and developing appropriate technologies for use in low-resource settings.

Claudia Ziegler Acemyan, PhD
Dr. Acemyan is an Adjunct Assistant Professor in the Department of Psychological Sciences at Rice University and the founding principal of Post Hoc, LLC, a human factors and safety consulting company. Her work focuses on system usability, safety, human error, warnings, medical human factors, assessment tools, and user-centered system development.

Maria Oden, PhD
Dr. Oden is a Professor of Bioengineering and Director of the Oshman Engineering Design Kitchen at Rice University and co-director of Rice360 Institute for Global Health Technologies. Her work focuses on strengthening engineering education techniques to help design teams better address healthcare challenges around the world.

Rebecca Richards-Kortum, PhD
Dr. Richards-Kortum is the Rice University Malcolm Gillis University Professor of Bioengineering and Electrical and Computer Engineering, and the founding Director of Rice360 Institute for Global Health Technologies. Her work is focused on developing medical technologies for use in low resource settings.

Philip Kortum, PhD
Dr. Kortum is an Associate Professor in the Department of Psychological Sciences at Rice University in Houston, Texas. His research is focused on the development and refinement of measures of usability and trust, and on creating highly usable systems in the global health, mobile, and voting system domains.