Heuristic Evaluation of Personal Health Records Systems

Leslie S. Liu
Department of Informatics
University of California, Irvine
lsliu@ics.uci.edu

Gillian R. Hayes
Department of Informatics
University of California, Irvine
gillianrh@ics.uci.edu

Abstract
Evaluation of Health Information Technology can be expensive and complex. In this work, we explored the use of heuristic evaluation as a low-cost evaluation method for Personal Health Records (PHR). We evaluate three PHR using both a traditional and a modified heuristic evaluation. The results of these evaluations demonstrate that heuristic evaluation can be used successfully to find many of the challenges to use of particular healthcare technologies.

Keywords
Personal health records, usability, discount evaluation

ACM Classification Keywords
J.3 Computer Applications: Life and Medical Sciences; H.4 Information Systems Applications.

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.
ACM 978-1-60558-930-5/10/04.

Introduction
Personal health records (PHR) maintained by patients and their families offer one solution for ensuring accurate and comprehensive health records even in health care systems that lack comprehensive interoperable standards for data storage and exchange. They can be desktop-based, Internet-based, or mobile (e.g., phone-based or located entirely on portable storage). They can be linked to Electronic Medical Record (EMR) systems, which are provider-owned and operated or they can be stand-alone. These records are created and monitored by the patients themselves, typically without the substantial restrictions and limitations of paper-based, centrally located medical records. If properly maintained and up-to-date, these systems can help a patient communicate important health information with clinicians, saving both patient and clinician time.

In this work, we set out to understand the particular challenges limiting the adoption and use of PHR through a systematic evaluation of three Internet-based PHR systems. Recognizing, however, that many hospitals and clinics lack the resources to conduct substantial usability evaluations in practice, we explored the use of heuristic evaluation (HE) [2] for PHR through the use of two separate heuristic evaluations, one using Nielsen's original ten heuristics and one with a modified set created by us for HIT

specifically. We compare the results using our modified heuristics with the traditional set.

Related Work

Usability testing of information systems has been a known benefit for decades. The cost of conducting extensive usability tests depends on the number of end-users, the site, and the process of the evaluation. This cost can grow quite large when considering complex medical systems. Thus, it is important to consider methods that are lower-cost, such as "discount" usability methods. Discount methods, as originally articulated by Nielsen and others, include scenarios, the think-aloud method, and heuristic evaluations. In this work, we were particularly interested in the potential applicability of heuristic evaluation (HE) to HIT. HE is a "discount" method for low-cost evaluation of an interface. The standard method involves a small set of evaluators (typically 3 to 5) who "inspect" the interface for issues of non-compliance (bugs) with a set of standard known usability principles (heuristics) [6]. A benefit of performing heuristic evaluations rather than other discount usability techniques is that these evaluations do not require end-user participation. Interfacing with potential end users—including both patients and clinicians—can be a substantial burden for HIT, thereby making HE an appealing choice.

Although heuristic evaluation has been shown repeatedly to be useful for uncovering usability problems for many types of applications with only a minimal number of evaluators, the original ten recommended heuristics may not always be adequately applied to particular systems. Such is the case for both Baker et al. and Mankoff et al., who not only applied heuristic evaluations to computer supported cooperative work (CSCW) and ambient displays, respectively, but also modified and used a different set of heuristics than Nielsen's original ten to fit the evaluated system [1, 4].

Methods

We evaluated three specific PHR systems: Microsoft HealthVault[1], Google Health[2], and WorldMedCard, now known as WorldHealthRecord[3]. These were chosen from a field of dozens based on the following criteria:

- Available to the public, not just patients of a particular healthcare system (e.g, Blue Cross or Veterans Affairs)
- Internet-based
- Mix of familiar and less known corporations

Following the work of Baker *et al.* who applied heuristic evaluation to CSCW [1] and Mankoff *et al.* who applied the technique to ambient displays [4], we were interested in whether and how heuristic evaluation might be used for HIT, in particular for PHR.

While heuristic evaluations are a useful means for obtaining feedback about interfaces, the general heuristics originally provided by Nielsen[4] do not necessarily apply in the HIT realm when considering PHR. For example, there are multiple simultaneous functions and users of a PHR system. Furthermore,

---

[1] http://www.healthvault.com
[2] http://google.com/health
[3] http://www.worldhealthrecord.com
[4] http://www.useit.com/papers/heuristic/heuristic_list.html, Retrieved March 2009

multiple individuals—parents of a sick child for example—may modify records for a single patient at once. Ensuring that all information is synchronized across users and views, however, is imperative to the usability and usefulness of these systems as well as to the health outcomes of individual patients. As another example, Nielsen's original heuristics emphasize the interface notifying users of any errors generated through the system. This heuristic, when applied to PHR, tended to uncover only technological errors.

In medical systems, however, many errors are technologically acceptable but medically problematic. For example, Koppel *et al.* found in their study of computerized physician order entry (CPOE) systems, systems in which physicians can enter medication orders into a computer, that medical personnel regularly depended on CPOE displays to indicate what the effective minimum dosage of medication should be for the patient. However, CPOE displays would show dosages that were not based on clinical guidelines, but rather pharmacy warehouse purchasing decisions, leading to errors [3]. To address the need for medically relevant usability evaluation, we developed a set of modified heuristics to be used in our evaluation of PHR systems. We first modified Nielsen's canonical heuristics and then added two more to fit the needs of health technologies.

We conducted two distinct heuristic evaluations. Five HCI experts evaluated the PHR systems based solely on Nielsen's set of heuristics. Five other HCI experts evaluated the PHR systems based on the modified set of heuristics. In both cases, each HCI expert also had the end-user task list for reference. The expert logged into each of the PHR systems to evaluate the specific system based on their assigned heuristics—either a modified or the original set of heuristics. For each heuristic, experts documented how they believed the heuristic and the actual system matched up. The evaluations lasted approximately 90 minutes. At the end of each evaluation session, the HCI experts were encouraged to give general feedback regarding their perceptions of PHR systems. After scoring the results for each heuristic evaluation, we further assessed the differences between those bugs and usability problems uncovered with each set of heuristics.

Results

We describe our Heuristic Evaluation comparison results that indicate that, although HE with Nielsen's ten recommended heuristics does in fact uncover many usability bugs, slight modifications to these heuristics perform better with the same number of evaluators.

To assess the ability of our heuristics to uncover usability bugs in PHR, we calculated the number of unique bugs found with the addition of each evaluator. For each case, beginning with the first evaluator until all evaluators were summed together, we calculated the number of distinct usability bugs—that is, only bugs that had not already been accounted for by a previous evaluator were added to the total. Five expert evaluators using the modified set of heuristics were able to identify more unique usability bugs (45) than eighteen end-users in lab testing (44). Although the modified set of heuristics were built upon the original Nielsen ten and explicitly only modified slightly, there were still usability problems found by the modified HE that were not found with the original set (see Figure 1).

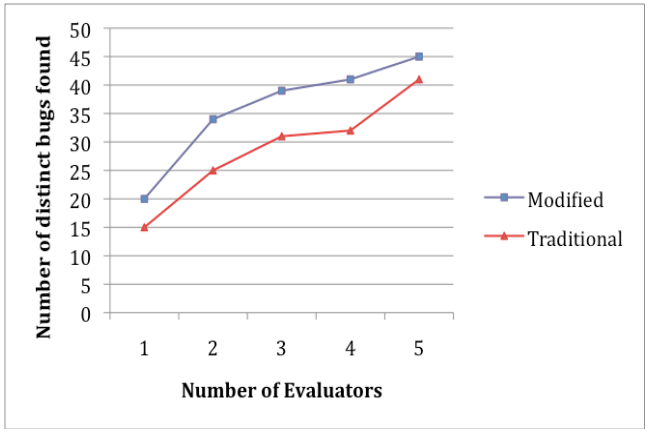

Figure 1. Comparison of unique bugs found across all systems per quantity of evaluators for each set of heuristics.

Conclusions and Future Work

PHR have enormous potential for empowering patients, increasing patient-provider communication, and tracking and monitoring health and wellness over time. Tracking these data over extended periods of time have the potential to provide baseline data, making detection of decline more efficient [5]. Furthermore, regular use of PHR can be beneficial in critical care situations, such as a trip to Emergency care, providing up-to-date information for a patient who may not normally be seen at the hospital that is currently providing treatment. Like any novel interactive system, before successful adoption can occur, however, the substantially challenges to usability and usefulness present in these systems must be addressed. Our results reinforce the need for integrating traditional and more recently developed methods from the HCI literature into the field of medical informatics, in particular in light of designing online systems for the general user population, such as PHR.

Our use of multiple heuristic evaluation techniques, using both traditional and modified heuristics, provide an example of how HIT may be effectively evaluated bearing in mind both standard usability and usefulness criteria as well as medically-relevant concerns. Furthermore, this approach of using modified heuristics can aid in evaluation of other HIT systems, an area for future work.

Acknowledgements

We thank all evaluators who participated in this study.

References

[1] Baker, K., Greenberg, S., Gutwin, C. Heuristic Evaluation of Groupware Based on the Mechanics of Collaboration, *Springer 2006.*

[2] Friedman, C., Wyatt, J., Ash, J. Evaluation Methods in Biomedical Informatics. *Springer, 2006.*

[3] Koppel, R., Metlay, J., Cohen, A., Abaluck, B., Localio, R., Kimmel, S., Strom, B., Role of Computerized Physician Order Entry Systems in Facilitating Medical Errors. *JAMA,* 1677-1682. 2005.

[4] Mankoff, J., Dey, A., Hsieh, G., Kientz, J., Lederer, S., Ames, M. Heuristic Evaluation of Ambient Displays. In *Proc. CHI'03*, 69-176.

[5] Morris, M., Intille, S., Beaudin, J. Embedded Assessment: Overcoming Barriers to Early Detection with Pervasive Computing. In *Proc. Pervasive'05*, 333-346.

[6] Nielsen, J., Mack, R. Usability Inspection Methods. *John Wiley & Sons, 1994.*

[7] Tang, P., Ash, J., Bates, D., Overhage, M., Sands D. Personal Health Records: Definitions, Benefits, and Strategies for Overcoming Barriers to Adoption. *JAMIA*, 121-127. 2005.