Mining MAUDE: Human Factors Perspectives on EHR and Device Design from the FDA Manufacturers and Users Device Experience Database

Steven D. Harris, Rational HealthCare, LLC
Robert A. North, Human Centered Strategies, LLC

Electronic Health Record (EHR) systems have been under development for several years, and have been in operational use by individual practices and healthcare delivery institutions since at least 2008. The federal Food and Drug Administration (FDA) maintains the Manufacturer and User Facility Device Experience (MAUDE) database to collect voluntary and mandatory reports of adverse events involving medical devices. In recent years, this system has also been used for reporting incidents involving electronic health record (EHR) systems. An examination of the FDA's MAUDE database identified recent entries that point clearly to serious human factors issues in the design of EHR systems. This paper describes an initial attempt to develop a systematic approach to the analysis of MAUDE data with the specific goal to gain insight into the nature and number of human factors issues that may be implicated in adverse events involving EHR systems. The paper presents results of preliminary analysis using these techniques. The results suggest that human factors issues are beginning to emerge in the design of EHR systems, manifesting themselves as adverse health care outcomes.

INTRODUCTION

Electronic Health Record (EHR) systems have been under development for several years, and have been in operational use by individual practices and healthcare delivery institutions since at least 2008. Based on the frequency and severity of human factors issues found in a wide variety of medical settings (cf Gawron, Drury, Fairbanks and Berger, 2006) it would seem highly likely that one or more adverse events attributable to EHR systems may have occurred in this time-frame.

There is no formal established mechanism for reporting such adverse events with human factors root cause to regulatory authorities. There is, however, an existing system, called the Manufacturer and User Facility Device Experience (MAUDE) database maintained by the federal Food and Drug Administration (FDA) and used for voluntary and mandatory reporting of adverse events involving medical devices.

Ambiguity in the relevant legislation, regulations, and policies leaves open the question whether EHR systems should be considered to be medical devices, and whether adverse events involving EHR systems are reportable in the MAUDE system.

An informal and very cursory examination of the FDA's MAUDE database identified recent entries that point clearly to serious human factors issues in the design of electronic health records (EHR) systems regarding patient safety and compromises in health care delivery. This observation piqued the curiosity of the authors, and motivated the present study.

APPROACH

Previous research (Zhang, Patel, Johnson, Chung and Turley, 2005) has employed human reviewers using an objective rating scale to review and score records from MAUDE for relevance to research questions related to medical devices as components in systems. The methodology used in that research is well-established, and has produced valuable results. But, the methodology depends on highly trained reviewers and is extremely labor-intensive and time-consuming. In addition, because it depends on human observers, the methodology still entails a certain subjective quality in the rating/scoring process that cannot be eliminated.

Moreover, the process of perusing narratives in the MAUDE database is tedious and inefficient as a means of identifying reports concerning specific topics such as EHR systems.

This paper describes an initial attempt to develop a more systematic approach to the analysis of MAUDE data with the specific goal to gain insight into the nature and number of human factors issues that may be implicated in adverse events involving EHR systems. The resulting methodology may have more general applicability to design and use of database systems, and to the design of taxonomic approaches to scientific observations.

Data sources and authorities

EHR systems are rapidly gaining wide use as an integral component of healthcare delivery in the United States, driven, in part, by recent federal legislation and regulations pertaining to implementation of EHR systems. These legislative and regulatory initiatives focus on “meaningful use” of such systems, and foster deployment and meaningful use through various financial incentives to individual practitioners and institutions involved in healthcare delivery.

The Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009 (a component of the American Recovery and Reinvestment Act of 2009) provides the Department of Health and Human Services (HHS) the authority to establish programs to improve health care quality, safety and efficiency through promotion of health information technology (HIT), including electronic health records.
legislation also established the Office of the National Coordinator (ONC) within the HHS to oversee the implementation of healthcare reform. The ONC engaged the National Institute of Standards and Technology (NIST) to help develop processes and practices related to certification of EHR systems. NIST has established such a process, and maintains a list of certified EHR systems called the Certified Health Information Technology Products List (CHPL). Lowry, Quinn, Ramaiah, Schumacher, Patterson, North, Zhang, Gibbons and Abbott (2012) recently published a valuable treatise on technical evaluation methodologies for the usability of EHR systems.

The Food and Drug Administration (FDA), another component of HHS, has had longstanding authority and responsibility for regulation of medical devices. Pursuant to that mission, FDA maintains the MAUDE database as a repository for narrative descriptions and related information about adverse events involving medical devices.

The authority for FDA to collect these data is derived from 21 USC 803.

MAUDE is the repository for both mandatory and voluntary reports of adverse incidents that reflect safety of medical devices. The data in MAUDE consists of reports filed by individuals, by user institutions, and by device manufacturers. The database is not intended to be used either to evaluate rates of adverse events or to compare adverse event occurrence rates across medical devices. To be useful for such purposes, technical changes and improvements are required in the database system. However, many and valuable insights can be gleaned from examination of the MAUDE data with focus on individual products, and with a broader view as to trends in the frequency and severity of adverse events involving those products or their manufacturers.

ONC has not established a reporting system for adverse events related to EHR systems analogous to MAUDE; nor has it been clearly required to do so by its enabling legislation. Despite the existence of these two databases – MAUDE and CHPL – available to the public, there is no repository designated specifically for reporting issues with EHR systems. Yet, some reports of incidents concerning EHR systems with negative patient outcomes have been filed in the MAUDE system by individuals and by institutions. (This is determined by analysis of narrative incident descriptions that incorporated explicit reference to a certified EHR system.) Hence, MAUDE is emerging as the de facto national repository of information concerning adverse events that involve EHR systems.

Methodology

The MAUDE database is accessible using an online search utility at an Internet website maintained by the FDA. The authors’ initial cursory search mentioned earlier was conducted using this online utility. Although the utility is useful for some purposes, it is not capable nor was it intended to support extensive analytical studies. For purposes of the present effort, a copy of relevant portions of the MAUDE database was obtained, reconstructed and hosted on a MySQL database platform. MySQL is a free, open-source relational database system that is widely used in both academic and industrial settings. As one might expect, MySQL was originally conceived and designed to facilitate information retrieval using Structured Query Language (SQL).

The entire MAUDE database is available for download from the same FDA server that supports the online search utility. Download is accomplished in several segments – depending on the period of time of interest, and on the specific details of the data that one wishes to examine. For example, narrative descriptions of events are maintained in separate files from details concerning manufacturers, or patients that were adversely affected by the event.

MAUDE data integrity

Upon initiation of a report concerning a new event, a new unique Medical Device Report (MDR) index key is assigned automatically by the MAUDE system. All subsequent records entered into the database pertaining to that same event are supposed to be identified with this MDR index key.

Perusal of MAUDE reveals that there are numerous event narratives that clearly pertain to the same event, but have been entered at different times, perhaps by different reporting individuals, and thus were assigned different MDR index keys. Moreover, there is no cross-correlation or compilation of data within MAUDE. So, for example, two different users may enter a vendor name slightly differently. These variations in data entry propagate complexity downstream in any attempt to analyze the MAUDE data.

Research dataset

For this present study, a subset of MAUDE data spanning approximately the period 1 January 2007 through 31 December 2011 was downloaded from the FDA website as compressed text files, and imported into the MySQL database system.

This period 2007-2011 is approximate, because there are several dates associated with each event – date of first report, date of the event, date of follow-up report, etc. The authors selected an MDR representing the earliest event that occurred in 2007. This MDR was then used to prune records from the dataset that had earlier MDR keys.

The period 2007-2011, inclusive, was chosen as representative of the time-frame during which EHR systems began to emerge in the healthcare IT industry leading up to the passage of the legislation designed specifically to foster development and adoption of those systems, and for a period of time afterward.

The dataset constructed for this study comprises approximately 2,454,904 reports representing ~843,333 unique reportable events (MDRs). We refer to this dataset herein as the Research Dataset. These numbers are approximate because the database engine uses an algorithm to estimate the size of large datasets, rather than counting the individual records. There were between 1 and 255 MDR
Overview of two-phase approach

The original research question arose from an informal, and somewhat ad hoc search conducted using the online search facility of the MAUDE database. This approach identified several records of potential interest – the first of which was deemed by the authors to be precisely on target. Excerpts of this record are presented in Figure 1.

With this interesting result, the authors decided to return to the question with a more systematic and rigorous approach. The dataset was derived in a two-phase strategy. Phase I entailed an effort to develop a comprehensive approach to identifying relevant MDR records in the dataset. Phase II entailed an effort to develop and demonstrate an innovative approach to categorizing the resulting records in order to “make sense” of the data. This two-phase approach was necessary since the specific research question we were confronted with – Are EMR system related adverse events reported in MDRs? – was not envisioned when the MAUDE database system was developed and deployed. Indeed, EMR systems did not exist at that time. Hence, the fact that there appear to be relevant records in MAUDE is somewhat surprising and deserves attention in its own right. But, the focus of this paper is on what the data represent, and not on how the data came to be there.

Phase I: Dataset construction

Phase I of the approach was conducted using a straightforward, if somewhat ungainly, Structured Query Language (SQL) query of the research data set. The query was developed in an iterative trial-and-error approach in which an initial set of query terms were tried, and the resulting datasets evaluated for relevance by the authors. We then refined the query to expand or contract the resulting set until we had a dataset that we had confidence was (a) repeatable, and (b) would withstand scrutiny from a disinterested third party. That dataset consisted of some 280 MDRs that, the authors concluded, were representative of EMR systems.

Phase II: UPCARE evaluation

UPCARE is an acronym that refers to – Unmet user needs, Perception, Cognition, Action, Results, and Evaluation. The term was introduced by Kaye, North and Peterson (2003), and has found some utility in description, analysis and classification of use error in medical devices. “Use errors” occur in the interaction process between human users and medical devices. Other disciplines employ other terms – e.g., pilot error in aviation; operator error in broader defense systems. In the defense realm, in particular, a broad perspective on the interaction of human operators and their systems is encompassed in the term “human-system integration.” Both of the present authors are experienced in defense and aerospace systems applications, and realized that complex interactions among human users and their information and decision support systems can be very difficult to identify and understand. The authors presumed that a comprehensive approach that encompasses all aspects of human-system integration (such as outlined and made explicit in the defense systems acquisition process) would be (or become) important.

Phase II of the approach then entailed evaluating MDR records from the UPCARE perspective. We did this in two sub-phases. First, we looked at all ~2.5 million records using the UPCARE methodology outlined above.

This process entailed calculating the MySql FULLTEXT index on the narrative descriptions in these MDRs, and then calculating the distance between each of the 280 records and individual documents that comprised the verbatim definition of each of the UPCARE categories. Herein below we refer to these individual documents as the UPCARE probes. These documents were constructed by a direct copy/paste operation from the UPCARE paper into the FULLTEXT query. An example of such a query is illustrated in Figure 2.

In Fig. 2, the language from the first taxonomic category of the UPCARE model is evident; lifted verbatim from the source and inserted into the FULLTEXT query. In the figure, the terms “fused312,” IN NATURAL LANGUAGE MODE, and Sc312 refer to the source table in the database, the mode for executing a FULLTEXT query (as opposed to a BOOLEAN mode), and the resulting Score, respectively.
Following the UPCARE analysis on devices, we repeated the approach using the 280 MDR records on EMR systems. We anticipate that we will publish the results of the medical device phase in a separate venue; the focus of this paper is on the EMR results.

RESULTS

Each of the six UPCARE categories was separately evaluated against the database of 280 EMR records in the dataset. This process entailed calculating the FULLTEXT index on the dataset, and then calculating the distance between each of these records and each UPCARE probe. The result was 280 records with a score representing the distance. These 280 records were then sorted into descending order by distance (UPCARE score). The top document in each set was then deemed to be the closest to the UPCARE category. That document was examined for clues as to the reason for appearing at the top.

Various statistics could be calculated on these results – including Spearman rho, for example. These statistics were not calculated explicitly because of the caveats in the MAUDE database system itself, and the exploratory nature of the research presented here. Caveats entailed admonition that statistical analysis of MAUDE data is problematic because of inherent redundancies and errors in the database. For example, as mentioned above, each MDR supposedly represents a single event. However, in some cases, including the first MDR examined in this study, there was no single event. That record concerned multiple events; it reported on a trend that spanned several months. In other cases, as cited above, there were between 1 and 255 MDRs for a given event; often duplicate entries. It is not clear how such entries arise in the database, but it is clear that such entries would skew any attempt at statistical rigor.

Excerpts from each of the top UPCARE categories are presented in Figures 3-8. It is noteworthy that one MDR in particular rated highest on two different UPCARE categories. That would suggest that this particular event might deserve further scrutiny for evidence of human factors deficiencies and failures in human-system integration.

Figure 3. Excerpts from MDR from EMR record scoring highest on Unmet User Needs.

The authors chose to examine the language in first record in the results – that is, the MDR record calculated using the UPCARE/FULLTEXT algorithm as being the closest to the UPCARE category in semantic content. It is important to note that the records generated by this algorithm are “untouched by human hands” – that is, the algorithm is completely transparent and repeatable. Development of an objective and repeatable methodology was an important scientific goal of this study. The data and algorithms are available to any interested reader. Requests may be directed by email to the second author.

Fig. 3 indicates that the system failed to meet user needs by failing to send required information from the source to the user's workstation. It is not absolutely clear that this failure is due to a human factors issue, however. The failure may have been due to software. More on this question in the discussion section, below.

Figure 4. Excerpts from UPCARE category: Perception.

Results in Figure 4, as with Figure 2, clearly show a failure in the system, but, again, as with Figure 2, it is not...
obvious whether the problem is one of failure in software or human-system integration. Closer study of this particular narrative indicated that it was an “electromagnetic interference” (EMI) problem. New laptops in the system operated on the same radio frequencies as existing equipment, causing improper information flow between devices and the EMR system.

While this may not be considered a conventional, traditional area for human factors analysis, the problem of interactions among information technologies is especially difficult to identify and trace to human or system failure. For example, if the systems in question are taken out of service and returned to the factory, they will pass bench tests. It is only in the operational milieu that the problems arose, and only after attention to the details of the problem, and not simply classifying the problem as one of “human error.”

The incident depicted in Figure 6 illustrates an interesting and all too common failure in human-system integration. In this case, “additional training” is called for. In this case, “human error” was in using a bar-code reader that would misread patient identification numbers if held too far (more than ~4 in.) from the patient’s wrist-band. In the complex and busy, interruption-filled milieu in which this technology is employed, it is probably unreasonable to expect that the bar-code reader would be held no more than 4 in from the wrist-band on every patient every time. Such precision is simply not within the scope of human capability.

In this case, introduction of a new technology could solve the problem. The interesting result here was, again, one of failure in effective human-system integration.

The incident in Figure 7 may look familiar. In fact, it is the same incident that scored closest to UPCARE category Cognition, depicted in Figure 5. The fact that the same incident (out of ~280) scored highest on these two different UPCARE categories deserves some attention, and will be addressed in the discussion section.

Figure 6. Excerpts from UPCARE category: Action.

The incident depicted in Figure 5 illustrates another example of the challenge of identifying and categorizing issues with EMR systems in real-world settings. In this case, the system performed exactly as advertised – hence there was no malfunction to report. The problem is more subtle. Information was not presented to the critical decision-maker in a manner that supported decision-making.

Figure 7. Excerpt from UPCARE category: Results.

The incident illustrated in Figure 8 clearly identified “user error” as the root cause. This is the “gold standard” toward which we are striving in the search for evidence in MAUDE of human factors deficiencies. In this case, however, the ‘user’ was not a nurse or a technician, but an information technology (IT) professional who configured the system in such a manner as to cause subtle conflicts in data among various devices attached to the EMR system. In short, this incident reflects a failure of human-system integration at a level not unlike that identified in many aerospace and defense systems – that is, user error may be induced by poor system design, poor or inadequate training, distractions, difficulties in operating within tolerances, incomplete or poorly organized and presented technical documentation, and many other issues that are not readily identified as hardware or software failures. Yet, such errors and failures in systems have been documented in MAUDE reports.

Figure 8. Excerpt from UPCARE category: Evaluation
**Human error rate**

One other result bears mentioning. Of the 280 MDR records that were identified as reflecting EMR deficiencies, 44 specifically mentioned the term “error.” Of those, 4 were not clearly relevant to the study at hand. For example, one of those four concerned human error in manufacturing of a device. In addition, 1 of the 44 attributed error to both human and system – a compounding of error. Removing those 5 from the 44 records leaves 39. Of those, 12 (or approximately 30%) were clearly concerned with use error (in the opinion of the authors). This result is consistent with error rates found in other domains such as aerospace, defense and transportation systems. (cf Gawron, et al)

**Analysis of medical devices**

An earlier study of MAUDE data conducted by Zhang, et al (2003), indicated that human error in medical device use accounts for a large portion of medical errors. In addition to the analysis of EMR data, a similar analysis was conducted of medical device data. Figure 9 illustrates the results of a single analysis to show the general shape of the data. A number of conclusions can be reached about the relationship between UPCARE categories and the medical device data, but such results and conclusions are beyond the scope of the present study.

![Figure 9. Preliminary look at Frequency of adverse events by type of Medical Device.](image)

The results in Figure 9 illustrate that while the vast majority of devices show a nominal rate of reporting in MAUDE, certain types of devices (specifically infusion pumps and fluoroscope x-ray devices) are considerably more likely to be implicated in MDRs. Similar data can be plotted by calculating frequencies of devices grouped in various ways.

**DISCUSSION**

The results of this study are limited to only six of the approximately 280 MDR records identified by the MySql FULLTEXT algorithm as being related in some way to EMR systems. Even cursory examination of the result narratives clearly and unambiguously show that the approach outlined here, relying on techniques from computational linguistics combined with accepted practice in human factors analysis can discover interesting and relevant information in the MAUDE database concerning human factors deficiencies and issues in deployed EMR systems.

This approach shows promise, but it is far from conclusive. Several aspects of the approach bear further study and scientific scrutiny. First, the definitions of the UPCARE categories could be further refined in an iterative process that entails testing the definitions against the calculated datasets from MAUDE. The similarities between category definitions are discovered when the same narratives are identified by the algorithm as being high on the list in two or more categories. It may be that these narratives depict problems that are, in fact, complex and may entail more than one UPCARE category. On the other hand, it may also be true that UPCARE language could be refined to further improve distinctions. In other words, it would be reasonable to consider using the approach outlined herein to facilitate the design of taxonomic categories such as UPCARE. The MAUDE database may be flawed in many respects, but it is a valuable corpus of natural language descriptions of problems; many of them involving human factors issues.

Furthermore, the results show that there is evidence that EMR systems have been implicated in adverse medical events.

**CONCLUSIONS**

**EMR systems are implicated in adverse medical events**

Perhaps the most important outcome of this study is the overwhelming and somewhat disturbing sense that EMR Systems have been directly implicated in adverse patient outcomes. The number of reports is relatively low, but the impact in some cases was quite remarkable. In at least one reported incident, the rate of medication errors increased by 50-fold. In other cases, death and injury were the result, caused in some cases by faulty decisions made by healthcare practitioners – but those decisions were predicated on faulty or incomplete or untimely information provided by the EMR systems.

**Potential Use in EMR usability studies**

It would be imminently feasible to use the approach outlined herein to evaluate error categories identified in the recent NIST technical report on usability and safety factors of EHR/EMR systems by Lowry et al (2012). In many cases, this report uses examples in its taxonomic definitions. The problem with examples is they are misleading for machines, which tend to weight the detailed and unique aspects of examples, and tend not to be able to draw conclusions and generalize from the examples. This is not surprising for fundamental reasons (cf Harris, et al, 1993).
Implications for operational decision-makers

One of the more interesting aspects of this work is the notion that so-called “natural language” processing may be able to facilitate development and refinement of scientific taxonomies. In this case, the individuals who actually filed the reports and generated the narratives were not (typically) trained in human factors. It would be useful to focus on training EMR professionals in the fundamentals of human factors problems – specifically the “informatics” and patient safety professionals in hospitals, clinics and practices.

It is reasonable to expect that if the reporting individuals were trained to recognize human factors issues, many incidents that are not now reported would begin to show up in MAUDE reports. Consider that in other settings, it is routinely estimated that “human error” is the “cause” of between 30 and 50% of accidents and adverse events. This is known to be true in other settings, and it held true in the present EMR dataset (12 of 39, or 30%). In short, it is reasonable to presume that so-called “human error” (i.e., failures in human-system integration) will arise in the use of EMR systems at rates that are common in other settings.

It remains to be determined how and whether an effective reporting system designed specifically to encourage and facilitate reporting of adverse events involving EMR systems can improve the effectiveness of EMR systems, and realize the promises of improved medical care.

References


