

Rochester Institute of Technology

## RIT Digital Institutional Repository

---

Theses

---

5-10-2016

### Healthcare Incident Reporting: The Impacts of Usability of Input Interfaces, Usability of Resulting Data, and Attitudes Towards Reporting

Paul Wiele  
pw1408@rit.edu

Follow this and additional works at: <https://repository.rit.edu/theses>

---

#### Recommended Citation

Wiele, Paul, "Healthcare Incident Reporting: The Impacts of Usability of Input Interfaces, Usability of Resulting Data, and Attitudes Towards Reporting" (2016). Thesis. Rochester Institute of Technology. Accessed from

This Thesis is brought to you for free and open access by the RIT Libraries. For more information, please contact [repository@rit.edu](mailto:repository@rit.edu).

Healthcare Incident Reporting:  
The Impacts of Usability of Input Interfaces, Usability of  
Resulting Data, and Attitudes Towards Reporting

A Thesis

By

Paul Wiele

B.S., Psychology, Syracuse University, 2011

Submitted in partial fulfillment of the requirements for the degree of

Master of Science in Experimental Psychology

in the Department of Psychology,

College of Liberal Arts,

Rochester Institute of Technology, Rochester, NY

May 10, 2016

We approve the thesis of Paul Wiele:

---

Esa Rantanen, Ph.D.

Date

Associate Professor of Psychology, Rochester Institute of Technology

Faculty Adviser and Chair of the Thesis Committee

---

Heidi Mix, MS, RN

Date

Director of LPN and CNA Programs

Genesee Valley Educational Partnership

Reader

---

Karlee Haschmann

Date

Clinical Analyst, Rochester General Hospital

Reader

### Acknowledgements

In addition to my thesis committee, without whose patience and encouragement this study could not have even begun, I'd like to thank the following:

- All of the hospital staff members who helped me recruit participants, allowed me access to facilities, and invited me to present about my research in progress.
- Ashley Cook of SRA International and Jennifer Herout, Ph.D., of the Veterans Health Administration, for their easy and concise usability ranking system.
- Nicholas DiFonzo, Ph.D., of RIT, for survey design advice.

### **Abstract**

Research into improving the safety of healthcare systems has recently focused on learning how incidents of harm to patients happen and how to prevent them. Although it is acknowledged that low participation in incident-reporting systems contributes to the problem of poor error prevention, little research has focused on improving participation. This research is focused on how both participation in and use of incident-reporting systems can be improved by examining the usability of the reporting tools. A large private hospital in the northeastern United States and the incident reporting app used there were examined as a case study. A mixed-methods approach using Critical Decision Method interviews, heuristic usability tests, and surveys was used. Seemingly minor usability issues like inconvenient and hard to read menus were found to inhibit both the quantity and quality of incident reports. Additionally, despite the organization having a generally strong safety culture, there were organizational obstacles to the reporting of incidents and the adoption of useful interventions in response to incidents beyond what is normally encompassed by the term “safety culture”. Specific recommendations for hospital incident reporting process improvement are included.

## Contents

<b>Acknowledgements</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>Introduction</b>	<b>1</b>
Using Incident Reporting . . . . .	4
What incident reporting can and cannot do . . . . .	5
Safety culture: Encouraging reporting . . . . .	6
Usability: Making reporting easy . . . . .	8
Precedents . . . . .	9
Successful incident reporting in another industry: NASA ASRS . . . . .	9
Current healthcare reporting practices . . . . .	11
Purpose of this Research . . . . .	12
<b>Approach 1: Interviews</b>	<b>13</b>
Method . . . . .	13
Participants . . . . .	13
Materials and procedure . . . . .	13
Results . . . . .	15
Factors affecting reporting . . . . .	15
Factors affecting analysis . . . . .	18
Ancillary findings . . . . .	21
<b>Approach 2: Usability Tests</b>	<b>21</b>
Method . . . . .	23
Participants . . . . .	23
Materials and procedure . . . . .	23
Results . . . . .	25

Bargas-Avila guideline issues . . . . .	25
Nielsen heuristics issues . . . . .	27
IOM guideline issues . . . . .	28
Ancillary findings . . . . .	30
<b>Approach 3: Surveys</b>	<b>31</b>
Method . . . . .	31
Participants . . . . .	31
Materials and procedure . . . . .	32
Results . . . . .	34
Free response questions . . . . .	34
Pre-analysis for rating scale questions: Demographic trends . . . . .	36
Rating scale questions . . . . .	37
<b>General Discussion</b>	<b>42</b>
Answering Research Question 1 . . . . .	42
Answering Research Question 2 . . . . .	44
Answering Research Question 3 . . . . .	45
<b>Limitations and Recommendations for Future Use of This Method</b>	<b>47</b>
Interviews . . . . .	47
Usability Tests . . . . .	48
Surveys . . . . .	50
<b>Conclusions: Specific Recommendations for Improvement of the Reporting</b>	
<b>App and Organization</b>	<b>50</b>
Recommendation 1: Revising reporting interface menu options . . . . .	50
Recommendation 2: Investigating and remedying punitive use . . . . .	52
Recommendation 3: Expanding and improving the use of time-saving shortcuts . . . . .	53
Recommendation 4: Improving the flow of information . . . . .	55

INCIDENT REPORTING vi

A final general advisory . . . . . 56

**Appendix A: Interview Questions 59**

**Appendix B: Usability Test Scenario 64**

**Appendix C: Survey Questions 66**

**Appendix D: RIT IRB Approval of Project 72**

**References 73**



**List of Figures**

1	The types of incidents . . . . .	3
2	Figure 1, expanded to show how the types of incidents relate to the types of incident reporting systems . . . . .	5
3	Flowchart showing the cycle of incident report information at the hospital .	22
4	Flowchart showing the goal of improving the cycle of incident report information	57
5	The flowchart seen in Figure 4, recolored to emphasize the proposed creation of a <i>Callback</i> -style newsletter . . . . .	58

**List of Tables**

1	Reporting App Compliance with Bargas-Avila and Colleagues' (2010) Guidelines for Web Forms . . . . .	25
2	Reporting App Compliance with Nielsen's (1994) General Usability Heuristics	27
3	Reporting App Compliance with Institute of Medicine (IOM) Guidelines . .	29
4	Participant Demographics: Role . . . . .	32
5	Participant Demographics: Incident Reporting Frequency . . . . .	33
6	Risk Matrix Used for Prioritization . . . . .	39
7	Priorities for UPs' Usability Problems . . . . .	40
8	Priorities for APs' Usability Problems . . . . .	41
9	Priorities for UPs' Culture/Organizational Problems . . . . .	42
10	Priorities for APs' Culture/Organizational Problems . . . . .	43

## Introduction

Beginning in 1999, the Institute of Medicine (IOM) of the National Academies of Science released a series consisting of reports, *To Err is Human*, *Crossing the Quality Chasm*, and *Patient Safety*, collectively called the “Quality Chasm series”, detailing the state of patient safety in the American healthcare system and recommending benchmarks and procedures to improve the system. A major focus of these and other improvement efforts has been reducing “adverse events”, a broad term encompassing any harm to patients that does not result from the patient’s condition (La Pietra, Calligaris, Molendini, Quattrin, & Brusaferrro, 2005). Adverse events due to human error are of particular interest to those who study the healthcare system, since they suggest the possibility of improving healthcare quality by eliminating or compensating for errors through policy changes. The origin in human error means these adverse events are presumably preventable.

Adverse events include the more urgent subcategory “sentinel events”, defined by the Joint Commission on Accreditation of Healthcare Organizations as instances of death, permanent harm, or temporary but life-threatening harm to patients (Joint Commission, 2016)<sup>1</sup>. Along with these, there are also “near misses”, errors which could reasonably have resulted in adverse events but whose effects were averted. Adverse events and near misses together make up what the IOM (1999, p. 101) calls “potential errors” and which are more commonly referred to in the current research literature as “incidents”. Incidents provide a large pool of information that may be tracked and learned from, but the IOM’s repeated recommendations to track and learn from these incidents (IOM, 1999, 2001, 2004) often go unfulfilled and are plagued by incomplete data.

The data that exist on the topic suggest frequent difficulty in detecting incidents. About half of adverse events that occurred during one study on complications from surgery were preventable (Healey, Shackford, Osler, Rogers, & Burns, 2002). In the case of medi-

---

<sup>1</sup>This is the official regulatory definition of a sentinel event, but it is worth noting that this may not match how all healthcare organizations use that term. The Joint Commission does allow some flexibility in how its policies are implemented, and has previously instructed organizations to modify the definition of a sentinel event for their own internal purposes, as long as such definitions still cover deaths and serious injuries (Joint Commission, 2011).

cation errors, a type of incident that is particularly difficult to detect except in retrospect, manual and computerized reviews of medical charts found 617 “adverse drug events”, only 27 of which had been reported at the time by staff (Jha et al., 1998). Those adverse events which are reported are estimated to represent, at best, half of the adverse events that actually occur (Barach & Small, 2000).

The further goals of explaining and predicting incidents do not fare well either. A study comparing different methods of predicting and explaining adverse events for 1,395 patients who underwent a risky surgery at 15 facilities in 4 countries concluded that root cause analyses (RCAs) failed to find explanations for 24.8% of the resulting deaths, and none of the various systems for predicting future complications which were applied at time of surgery successfully predicted more than 21.7% of the deaths (Vollmer et al., 2011). Even the prediction method that made the largest number of accurate predictions, POSSUM, seems to suffer from a different problem, a bias towards producing a large number of false alarms, since it also predicted death for 17.6% of patients who did not die during the period examined. Such inaccurate predictions imply either faulty models or a lack of detailed and accurate information from which to make predictions about patients’ future well-being. A recent review by Marchon and Mendes (2014) of the patient safety literature as it pertains to primary care turned up similarly troubling results: studies in a number of different countries identified a lack of communication of “lessons learned” from past incidents, as well as vast differences in practices from facility to facility making policy change recommendations very difficult.

If this is representative of the state of human error analysis in modern medicine, researchers must study and improve the way incidents are assessed and analyzed. One potential improvement to incident reporting is to track not only adverse events but also the much larger population of near misses. Research in other high-risk industries suggests that the circumstances surrounding near-misses are not substantially different from those that cause actual harm (IOM, 2004). This observation means that a failure to know about near-misses is an especially worrisome situation for an organization to be in. An organization

analyzing the circumstances around near-misses would naturally have access to a much larger data set with which to understand the circumstances that can lead to harm than is one which only analyzes the particular set of circumstances that led to any one adverse or sentinel event (IOM, 2004). Indeed, Marchon and Mendes (2014) found estimates that between 41% and 61% of known reported incidents were near misses by some definition, either not reaching patients at all or reaching patients but not harming them. We have, then, three types of “incidents” to consider. The categories of incidents and their relative frequencies are summarized in Figure 1.

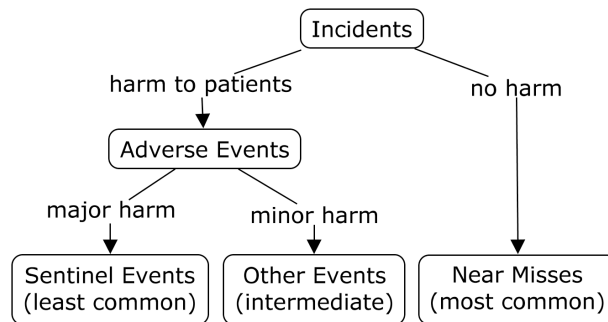


Figure 1. The types of incidents.

Studies on the behavior of people who use incident reporting systems (for the purposes of this study, “systems” are any of the various procedures and tools by which people may report incidents) suggest that the true number of near misses is even larger than what Marchon and Mendes (2014) found. One survey of American nurses found that barely over half of them would make a formal report about an incident that does not harm the patient, and barely over two thirds would share information about a harmless incident with colleagues, compared to rates of 89% to 99% for reporting or sharing incidents that caused various amounts of harm (Throckmorton & Etchegaray, 2007). In Norway, Flaaten and Hevr (1999) introduced an incident reporting system for an intensive care unit using anonymously-submitted forms for describing incidents. Although they found that the majority of incidents reported were near misses, the total number of reports made was so small as to suggest that there was overall drastic underreporting by staff. Comparison of incidents reported by staff against those detected by observers suggests that there is also a problem of certain types of

incidents that do not harm patients going unreported because staff members do not notice them, or do not recognize them as incidents (Capuzzo et al., 2005). By not emphasizing near miss reporting, and by not achieving a high enough reporting rate of harmful incidents, large pools of potentially-useful data are being overlooked by the healthcare system as a whole.

There has been some progress in reporting near misses in other contexts, and documents encouraging the inclusion of near misses in incident reporting systems are becoming common and readily accessible in diverse contexts. cursory Google searches for terms like “near miss” and “near miss report” quickly turn up hundreds of thousands of hits, including online reporting systems for firefighters, explanatory web pages from several federal and state government agencies, and construction industry employee training materials. The prevalence of smartphones has even allowed mobile and convenient systems for reporting incidents to the relevant authorities, such as Snap Send Solve for reporting Australian public works incidents (Outware Mobile, 2013) or ioSweep for companies to keep track of their own commercial vehicle accidents (ioSweep, 2015). Of course, even the best system is of no use if it is not adopted; this is why it is important to the improvement of healthcare that researchers, policy makers, and medical professionals seem to have taken the IOM reports as a “call to arms” (Barach & Small, 2000, for example). In the five years after the first of the reports was published, the proportion of many different kinds of patient safety publications within the medical literature increased, including an increase in original research on patient safety issues from 23.7 to 40.8 per 100,000, and the most commonly-discussed patient safety topic became organizational culture rather than malpractice (Stelfox, Palmisani, Scurlock, Orav, & Bates, 2006).

### **Using Incident Reporting**

Before changes can be made to address incidents with any certainty of success, accurate information must be obtained about the rate and manner in which they occur. In the words of a brief presented to Congress on the topic, due to inconsistent standards across

states and unknown response rates, “it is not possible to quantify the number of errors today and therefore impossible to determine if the goal has been met” in reducing their occurrence (Bleich, 2005, p. 9). It is certain, however, that not nearly all of the incidents that do occur are reported. Currently, only 26 states and the District of Columbia have policies in place to report their adverse events (Joint Commission, 2013), and even those places with reporting systems do not make full use of them. The New York Patient Occurrence Reporting and Tracking System (NYPORTS) requires a variety of adverse events to be reported to the state within one business day of their occurrence, but there are severe problems of underreporting throughout the state, especially in New York City (State of New York Department of Health, 2007; Pear, 2012, January 6). Further, NYPORTS, like the Joint Commission’s voluntary reporting system (Joint Commission, 2013), mainly tracks sentinel events, not the large and information-rich selection of near misses or even certain minor incidents of harm. A summary breakdown of the types of incidents and how they are (or may be) reported is shown in Figure 2.

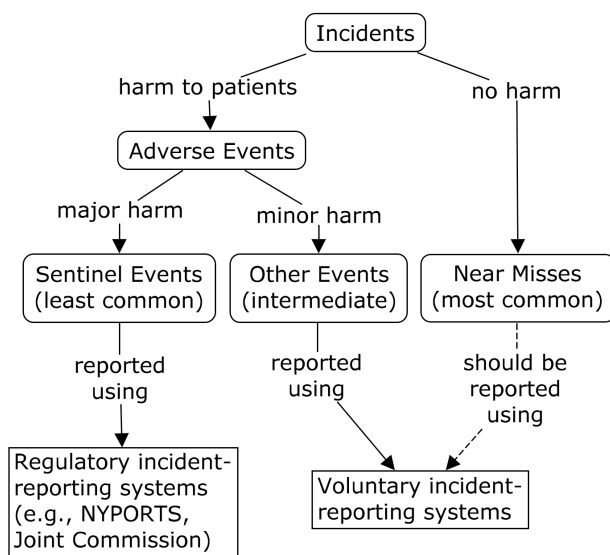


Figure 2. Figure 1, expanded to show how the types of incidents relate to the types of incident reporting systems.

**What incident reporting can and cannot do.** Information like that gathered on a voluntary basis by the Joint Commission can only be used to make rough estimates or

to compile a general ranking of the kinds of incidents that are common, or at least those that are commonly reported. The Joint Commission's data also only contains sentinel events, not incidents in general, and can only provide a limited picture of the types of errors that are made and how they can be avoided or mitigated even if the sentinel event data were complete.

However, it may not be necessary to collect reports of all incidents, or even a representative sample of them. Strictly speaking, if "incidents" are events due to human error that cause or could have caused harm to patients as the definitions above suggest, any time a person in the hospital coughs could be reported as a near-miss of an exposure to a contagious disease. Such reports would not conceivably contribute to the overarching goal of improving the safety of the healthcare system, and would instead waste the staff's time. The purpose of research on incident reporting and interventions to improve it should not necessarily be to draw the incident reporting rate closer to the true rate of all incidents but to ensure that staff are willing and able to consistently report those incidents that would be most useful for reducing patient harm. We should not necessarily expect reports to be proportionally-distributed throughout all departments or even for reports of a given type of incident to match the frequency with which that type of incident occurs. However, this means that it is important to determine whether those incidents that are considered important are being reported, and whether those reports are useful for changing procedures to avoid or mitigate such incidents in the future.

**Safety culture: Encouraging reporting.** The term "safety culture" encompasses the attitudes and practices of members of an organization that contribute to avoid or reduce harm caused by the service provided by understanding and learning from incidents instead of punishing people for those incidents or merely complying with outside regulations (Milstead, 2005). This contrasts with "blame cultures" that seek to identify and punish specific people instead of understanding incidents as resulting from situational factors (Waring, 2005), and is closely related to the idea of "generative cultures" based on cooperation and innovation (Westrum, 2004). An organization with a weak safety culture will treat employees reporting



problems as incompetent or out of line, while one with a strong safety culture will treat them as providing challenging and productive information (Taylor, 2012).

Some regulatory and advisory groups in healthcare have recognized the importance of a strong safety culture. The IOM (1999) emphasized the value of creating a non-punitive environment even in the original *To Err is Human* report. The Commission on Accreditation of Medical Transport Systems regularly surveys its accredited organizations to measure the current state of their safety cultures (Frazer, 2011), and the Agency for Healthcare Research and Quality (AHRQ), part of the United States Department of Health and Human Services, conducts a number of regular safety culture surveys specifically tailored for hospitals, medical offices, nursing homes, and pharmacies (AHRQ, 2004).

The attitudes described as safety culture correspond to helpful behaviors. Interventions to improve safety culture reduce the rate of observed harm to patients (Timmel et al., 2010; Braddock et al., 2014). Beyond this, though, safety culture indicators predict incident reporting rate, especially those relating to teamwork (Erler et al., 2013). Safety culture at the level of the entire facility appears most closely related to incident reporting, although the attitudes of the facility, department, and individual are all correlated negatively with incident occurrence rate and positively with incident reporting rate (Kagan & Barnoy, 2013). It is worth noting that in one of the examples of underreporting of incidents noted above, only about half of nurses surveyed rated their workplace as having a positive safety culture (Throckmorton & Etchegaray, 2007), and that only 44% of all respondents to a major nationwide safety culture survey had made an incident report in the previous year (Sorra et al., 2014). Therefore, in addition to usability, safety culture both overall and at the unit level should also be measured and considered when attempting to explain the hospital's incident reporting methods and determining how they may be improved. It is also worth investigating whether the very high proportion of reports made by nurses is due to particularly lax reporting by other hospital staff (explainable by differences in safety culture) or to nurses actually observing more incidents due to the nature of their jobs (Busse & Wright, 2000; Johnson, 2002).

**Usability: Making reporting easy.** “Usability” is the extent to which a particular tool can be used for its intended purpose (ISO, 2010). This is important to consider when assessing a reporting system because, if the task of writing an incident report is sufficiently unpleasant, confusing, or difficult, staff members could be discouraged or even prevented from completing those reports. Changes in the design of web forms in other applications such as surveys and online shopping has increased the rate of completion of those forms by up to 40% (Wroblewski, 2008). “Usability” refers not only to how easy something is to use, however, but specifically to how easy it is to use *in the intended manner* and *for the intended goal*. In this case, to be considered “usable”, the form should not only encourage people to complete it, but also collect the amount and kind of information about an incident that is needed to implement a response that will make future incidents less frequent or less severe in the future. In designing a military incident reporting form, Pettersson (2013a) not only focused on features such as the form’s generality and readability, but also including prompts for specific details (e.g., weather conditions, previously-existing issues that influenced the incident) that are important for truly understanding the incident.

Less obvious but also quite important is that the same principles of usability could determine whether the information gathered by the system is intelligible to analysts, and whether it can be shared easily when needed. The full details of Pettersson’s form redesign process are illuminating for this, as she identified incident reports made using the older, more open-ended form as being frequently incomplete, and determined through testing of new form designs that prompts for key details resulted in reports that more often identified plausible causes and solutions for the incidents in question than reports filed with the old form had (Pettersson, 2013b). In the hospital setting with which the present study is concerned, an incident report and its analysis are typically not provided by the same person, and are certainly not provided at the same time using the same form, but the general idea remains that certain context must be included in incident reports if a useful solution is to be found.

The hospital involved in this study uses a reporting app with three interfaces of inter-

est: one for reporting incidents, one for reviewing the reports and describing the measures taken to respond to them, and one for viewing aggregate data about the reports that have been filed. For the purposes of this paper, I will refer to them as the reporting, followup, and aggregate interfaces, respectively. Aspects of the design of all three of these interfaces could influence the hospital's ability to learn from incidents, and in particular, the reporting interface's design could contribute to the problem of incident underreporting. Designing new reporting software is outside the scope of the present study, but recommendations to the hospital or the software developer for how they may, respectively, customize or redesign the interfaces is. Various types of usability guidelines may be applied to the way the reporting app solicits input to find potential flaws with the design of the system itself that discourage or prevent incident reporting from proceeding as intended.

### **Precedents**

To understand how safety culture and usability can impact reporting, and to identify what likely areas for improvement exist at a particular facility or with a particular reporting tool, consider the successful examples of such systems that already exist.

**Successful incident reporting in another industry: NASA ASRS.** One precedent for a successfully-implemented incident reporting system in an industry highly sensitive to dangerous errors is the Aviation Safety Reporting System (ASRS). This system was developed by the National Aeronautical and Space Administration (NASA) in response to a pair of nearly identical incidents on different airlines in 1974, one a near-miss and one a deadly crash, which could have been avoided had safety information been shared between the two different airlines involved (Reynard, Billings, Cheaney, & Hardy, 1986). Although information about the actual administration of ASRS is difficult to come by (Beaubien & Baker, 2002), it is at least known that ASRS was designed to collect and store data in such a way that it may be searched for key information about "any arbitrarily defined topic related to aviation safety that may arise" while preserving the quality of information that would be available by accessing the original reports from which the database is constructed

(Reynard et al., 1986).

ASRS is still improving, and encouragingly, has shown change over time in who reports incidents. From 2003 to 2012, the proportion of incidents reported by air traffic control and cabin crews increased (NASA, 2012). It is hopefully uncontroversial to say that the very different environments involved would expose reporters with different jobs to different types of incidents, and that they therefore find problems that would otherwise go unnoticed. However, it is still not a perfect system. The actual data collection process consists of voluntary submissions from pilots and other professionals and amateurs involved in aviation, who report aviation's equivalents of the types of adverse events and near-misses (that is, accidents that actually did cause harm to people and incidents that could have but were somehow averted). Due to this voluntary nature, it is unlikely that ASRS provides for a representative sample of incidents, but it at least has the capacity to identify the causes of frequently-reported types of incidents (Reynard et al., 1986). It does not directly result in regulation, although analysts' judgements about urgent hazards are forwarded to appropriate regulatory authorities. Rather, it is credited with improving the safety of aviation by improving the willingness and ability of its participants themselves to identify and prevent the circumstances that commonly lead to incidents (Beaubien & Baker, 2002). Two key features responsible for this are confidentiality and feedback.

Confidentiality in ASRS includes anonymization of the reports themselves and legal immunity when incidents are neither deliberate nor due to a person being unqualified for their duty. If it is clear that an incident was not deliberate or due to incompetence, NASA analysts remove identifying information from reports and release them to the public, making them practically anonymous (Reynard et al., 1986; Billings, 1998). The offer of conditional immunity from prosecution, according to the designer of the system, has resulted in pilots providing large quantities of information which precisely explain their incidents and show why they were not at fault (Billings, 1998).

Feedback in ASRS comes mainly in the form of a regular newsletter, *Callback*<sup>2</sup>, which

---

<sup>2</sup>Callback is freely available to the public at <http://asrs.arc.nasa.gov/publications/callback.html>

presents selected incidents and the results of analyses of them. Each issue focuses on particular themes which are reflected in issue titles, such as “Winter Weather Operations” (January 2014) or “When Autopilots Go Bad” (October 2012). Incident reporters are also acknowledged and thanked for their reports in a very specific way, with a letter that includes a timeframe within which the incident report will be analyzed, and a form for the reporter to send to NASA to request a description of how their particular report was handled (Reynard et al., 1986). Studies on feedback interventions in other contexts imply that this is a likely area for improvement of incident reporting practices. Feedback on how survey results are used has been found to increase intent to respond to future surveys (Thompson & Surface, 2009) and to attend follow-up events (Hutchinson et al., 2012). Most promisingly, Kluger and DeNisi (1996) found via a meta-analysis that feedback generally improves motivation and performance in a variety of tasks.

**Current healthcare reporting practices.** The American healthcare system lacks a nationwide reporting and analysis system as large and comprehensive as ASRS, despite having a variety of limited-use voluntary and statutory systems in use for many years (Leape, 2002) to draw from for inspiration. The recent Partnership for Patients initiative is somewhat of an example, a limited analogue to ASRS used only within the Medicare/Medicaid system, which has been credited with reducing the number of hospital-acquired conditions (AHRQ, 2013). However, many tools are available with which healthcare providers can collect data on different indicators of their safety and efficiency. Billing data may be used to measure factors like time of stay, which act as indirect measures of the quality of care provided. Nursing departments may submit data to the National Database of Nursing Quality Indicators, which records many different types of incidents but only in a nursing context (Montalvo, 2007). Hospital staff in general report adverse events and near misses by various methods for their internal records as suggested by IOM (1999) and Joint Commission (2011). The hospital with which the present study is concerned uses all of these methods.

Incidents are reported there through an online reporting application, the records from which are then used for both internal and external reporting purposes. It is alone among

the above-mentioned methods in that it identifies both adverse events and near misses, is used by all departments, and contains details about not just what kinds of incidents occur but also how they occur. Reporting apps seem very promising as a method of learning from incidents – but only if they are implemented in such a way that people are both willing and able to provide the needed information. The present study focuses on the usability and organizational practices surrounding this app.

### **Purpose of this Research**

The present study examined incident reporting as it is practiced at a large private hospital in the northeastern United States. The purpose of this research was to determine whether, and to what extent, the plausible factors reviewed above contribute to the underreporting of important incidents, including both harmful incidents and near misses, and how well the resulting reports are used to improve patient safety. This information was then used to advise the hospital on how they can expect to gather more and better incident reports. This is formalized in the following research questions:

- *Question 1:* To what extent is underreporting of incidents due to problems in the organizational procedures surrounding the use of the reporting app?
- *Question 2:* To what extent is underreporting of incidents due to problems in the usability of this particular reporting app?
- *Question 3:* What issues are analysts facing in their use of incident reports to improve patient safety, and to what extent?

The present study addressed the above questions in three ways. The first two approaches, semi-structured interviews and usability testing, focused on examples of organizational and usability features of the reporting system. These “depth” methods examined specific attitudes and decisions related to incident reporting and observations of the reporting app to identify specific strengths and weaknesses of the app and the way the organization

applies it. The last approach, surveys, focused on the experiences of the hospital staff overall. It was a “breadth” method, using questions based on past surveys and the results of the interview and usability test approaches to determine the commonness and severity of potential problems in the system, for the sake of prioritizing policy changes that this hospital can take and in order to provide a model for generalizing these results.

### **Approach 1: Interviews**

#### **Method**

**Participants.** Six participants were recruited for in-depth interviews, five regarding normal use of the reporting app and one regarding the analysis and use of aggregate data. Three participants were recruited via a series of e-mails sent by staff in the hospital’s Patient Safety department to specific staff members they believed would be interested in participating. Two participants were recruited via contact cards given out at a quality improvement event at the hospital. One participant was recruited via e-mail at the suggestion of a family member of the participant. All participants were asked to recruit others in a snowball sample, but no additional participants were recruited this way. Of the five participants who took part in the interviews about normal use of the reporting app, one was a nurse manager, two were non-manager nurses, one was a manager in a non-medical position, and one was a non-manager in a non-medical position. Participants’ ages and genders were not collected due to concerns that this information could be used to personally identify them.

**Materials and procedure.** The interview procedure for the five regular app users was semi-structured and consisted of three sections. In the first, I asked introductory questions about job position and whether the participant was responsible for analyzing submitted incident reports. In the second, I asked an assortment of questions about teamwork and attitudes adapted from various safety culture surveys, along with a small number of usability-related questions. In the third, I used the Critical Decision Method (CDM), a tool of Cognitive Task Analysis, as described by Crandall, Klein, and Hoffman (2006) to

explore specific examples.

Following the lead of Erler et al. (2013), the general safety culture and usability questions were drawn from, or based on questions found in, established sources: two overlapping but slightly different safety culture surveys used by regulatory bodies (Commission on Accreditation of Medical Transport Systems, 2012; AHRQ, 2004), one study that explored the implementation of a health information exchange system (Feldman, Schooley, & Bhavsar, 2014), and one study that recommended a set of questions for getting explanations of a proposed system's user requirements in advance of designing the system (Browne & Rogich, 2001). For analysts, who are regular users of the completed reports others have submitted rather than its input interface, questions were further modified to ask about potential organizational influence on their analyses rather than on their input.

The CDM (Crandall et al., 2006) consists of identifying an example of the activity being studied, then asking a series of questions about the details of the decision-making process used in that example, including whether the participant consulted others for advice, how long it took to make the decision, and whether the example event the participant describes is typical or atypical within their work experience. Answers are then verified by creating a timeline of the event and asking the participant to verify that it is an accurate description. Some additional questions were drawn from the above-mentioned study on eliciting user requirements (Browne & Rogich, 2001) and a study on the general usability of incident reports across industries (Snowdon & Johnson, 1999). Counterfactual questions are typically also included in the CDM, asking what the participant would have done if certain key factors had been different, but this line of questioning was omitted. In the present study, I asked for an example of one time each participant successfully made an incident report and one time each participant decided not to make a report or aborted a report in progress.

In all sections, participants were also asked some unscripted followup questions for clarification and to explore unexpected tangential topics, but for the most part, the interviews followed the scripts prescribed by the safety culture surveys and the CDM. The



complete set of scripted interview questions is included as Appendix A.

The trend analyst was interviewed using a separate procedure in a substantially less-structured way about other issues unique to that participant's job responsibilities, which still incorporated some of the safety culture and CDM questions when appropriate. The analyst uniquely asked not to be recorded, so notes were taken throughout the interview.

## Results

**Factors affecting reporting.** These factors can be classified broadly as dealing with usability or organizational culture and practices.

*Organizational factors.* CDM walkthroughs did not provide as much detail as expected about the examples they described because participants were very slow to offer specific examples and instead typically opted to describe general types of incidents they do or do not report. When asked for examples of situations in which they chose not to make reports, participants frequently mentioned problems encountered when making reports on behalf of others. One participant gave the example of simply forgetting to make a report requested by someone else, and implied that the person who requested the report could have made the report instead but was unwilling rather than unable to do so. One participant gave a very different answer, an example of a problem judged to be not appropriate to report using the reporting system. In this example, the participant described intervening to settle a personal dispute involving a member of their unit being rude to another member, an issue which does not meet the criteria of an incident but which is occasionally reported anyway (see "Approach 3: Surveys", below).

When asked for examples of situations in which they chose to make reports, participants once again described scenarios in which they entered reports on behalf of others because the person who committed or witnessed the incident did not have time. Taken with the above, this suggests that a key difference between the reports made and reports not made on behalf of others is simply a matter of who finds time during a busy workday to make the report. One participant described a situation in which they made a report that

another staff member neglected to make because they judged it to be unimportant. This example serves as a counterpoint to the above example of declining to report a staff conflict, as it described a situation in which someone's decision that something was a "non-incident" was incorrect. Such misses, if not picked up by another staff member as occurred in this example, are a likely cause of underreporting, especially since the participant indicated that this was a "very typical" circumstance.

In almost all cases of both types of examples, participants indicated making the decision to report or not to report quickly, immediately, and with little or no outside guidance. When guidance was sought, it was to confirm the accuracy of details with those involved in the incident. Incident reporting, except when reporting on behalf of someone else, appears to be a very individual process.

Four participants indicated they believed that staff were afraid of being punished for their involvement in incidents when the incident reporting system was first introduced. They said that that fear became less widespread as staff became used to the reporting process, but two participants went on to say that they were aware of staff members who currently believe anonymous reports are not actually anonymous. Although confidentiality is built into the reporting app itself because the system never requires incident reporters to identify themselves, there is at least some fear that reports could contain enough information to be personally identifiable. One unit manager clarified that this varies from department to department, because some unit managers' followups to incidents are very harsh and blame-centered while others handle incidents more gently. Despite the confidentiality the app is designed to allow, participants reported having to reassure others that the reports are supposed to be used for learning rather than punishment. The other unit manager who participated even reported having to disregard reports that were apparently filed with the intent of getting specific staff members in trouble.

Certain issues related to the organization and job responsibilities may interact with the app in such a way that they create new usability problems that otherwise would not occur. The above-mentioned problem of unhelpfully vague reports resulting from problems

that do not meet the definition of “incident” results from the way the organization treats the reporting system. If a staff member pauses or is interrupted while writing a report, they may lose their report due to the app automatically “timing out”. Most often cited by participants, however, was the problem of sheer time required. Participants from two units described shortcuts created to reduce the time needed to make reports. Although all participants in this study completed their example incident reports in less than 7 minutes (see “Approach 2: Usability Tests”, below) they warned that staff may take up to 15 minutes per report, which is prohibitive when it competes with important job responsibilities like making patient rounds or when staff typically only have time to report at the end of the workday rather than immediately upon committing, witnessing, or otherwise learning of the incident they intend to report. Delaying reports to late in the workday makes it more likely that important details or entire incidents will be forgotten.

Participants also reported making judgements that particular problems did not qualify as “incidents” and therefore should not be reported. Some of the specific examples given in interviews, dirty bathrooms and interpersonal disputes between staff members, seemed to be unambiguously correct rejections which should be handled by another method rather than reported via the incident reporting system. Two participants whose job duties included transcribing reports filed via time-saving shortcut methods noted that they had decided to not report non-incidents that other staff had attempted to report via those shortcuts, both of which appeared to be correct rejections. If they had not been in the position to judge those as inappropriate topics for incident reports, those non-incidents would have been reported. This suggests that, in addition to underreporting, there is a problem of “false alarms”, that is, non-incidents being reported as incidents. Misses are plausible here as well, for example if a staff member mistakenly believed some piece of faulty surgical equipment to be a low-priority maintenance issue rather than an urgent danger, but none of the participants provided examples that were clearly their own failures to recognize true incidents.

The hospital organization also seems to have a strong safety culture, judging by

participants' statements that there is much less fear of punishment now than there was when the reporting app was introduced. One participant who handled reports submitted through a shortcut method mentioned singling out staff who made near miss reports in order to praise them for their "good catches". Although doing so compromises confidentiality, positive feedback for participating in the incident reporting system encourages future use.

*Usability factors.* Although a separate procedure was used to detect usability issues (see "Approach 2: Usability Tests" below), participants in the structured interview also reported that either they or their colleagues encountered issues with the app's usability. These mostly concerned the reporting app's drop-down menus. These menus, which are used for classifying incidents, were described as containing too many options to be used quickly, including some options which were ambiguous or seemed irrelevant to the particular category of incident being reported. However, issues with the specifics of the drop-down menus were not simply a matter of there being "too many" options. One participant reported that certain types of common incidents were left off of the menus, and suggested that the app could be used much more quickly if additional incident categories were added.

One unit manager mentioned an ongoing problem that impacted the usability of the submitted incident reports. In some incidents involving equipment, reports did not specify the situation or manner in which equipment was malfunctioning, making it impossible to know what diagnostics or repairs are needed if the problem is not obvious upon seeing the equipment. The participant attributed this to a belief among the staff of the unit that submitted the report that the equipment was not actually broken but had just been used incorrectly. Finally, the analyst who took part in the unstructured interview identified one issue with the usability of the aggregate interface: that selecting the followups made by unit managers does not consistently include the followups in the generated aggregate reports. It was not clear whether this issue impeded analysis or was merely irritating.

**Factors affecting analysis.** These factors are broadly concerned with the flow and use of information about incidents within the hospital. Most are organizational, although participants reported that two usability issues not addressed above also influenced this

organizational information flow.

Participants' descriptions of the reporting process, along with informal conversations with other hospital staff members who aided in arranging the present study, were used to create a flowchart, seen in Figure 3, of the hospital's incident report cycle. Staff members use the reporting interface of the app or the appropriate workarounds to make incident reports. Specialized staff members called "Quality Coordinators" (QCs) audit the submitted reports to ensure that they are classified correctly (e.g., that the "level of harm" selected is consistent with the narrative description of the incident) before forwarding them to the appropriate unit manager(s) for analysis. Unit managers review the reports, decide appropriate responses, implement them, and ask leadership for coordination if the response to the incident requires the action of multiple units. They then describe their responses using the app's followup interface and submit them for record-keeping purposes.

Patient safety analysts use the app's aggregate interface to obtain "ad hoc reports", customizable tables of aggregate data about incidents. These aggregate reports are used by unit managers, quality coordinators, and administrators to examine the trends in the occurrence of incidents. Only reports about the frequency of specific incident types deemed to be high priority are created on a regular basis, but analysts may generate reports as requested on a wide variety of topics, including any category of incident listed in the reporting interface and what, if any, response was made to a given incident. These reports do not always contain the requested information. The analyst participant reported that updates to the app are issued, but may change default settings without prior notice, and that even if the settings are correct, the aggregate interface intermittently fails to export the descriptions of the followup actions taken.

The time required to make a thorough report was a common theme in the interviews. Intriguingly, two participants in the structured interview described workarounds already in place in their units to reduce the time needed to report incidents. In one unit, a hotline was implemented, so that the person who commits or witnesses an incident may call an internal phone number and leave a voice message describing the incident. In another, a

“reporting book”, essentially a notepad of brief forms to fill out, was introduced, which achieves the same effect in a different medium by allowing staff to write rather than speak their freeform descriptions. In both workarounds, a designated staff member (a supervisor for the hotline, and the unit manager for the book) transcribes the freeform descriptions into the appropriate box in the reporting interface of the app and enters as much of the other prompted information as possible from the description available. The participants who described these workarounds credited them with reducing the average time to report an incident; in the case of the hotline, the estimated reduction was from 10–15 minutes to as little as one minute.

Feedback is provided in “group huddles”, meetings between unit managers and their units where recent incidents are reviewed and unit managers give reminders or introduce unit-level policy changes when needed, and broad information on trends in certain incident rates is provided in aggregate reports. However, the unit managers interviewed expressed difficulty coordinating with other units. When a unit manager’s followup concludes that multiple units contributed to an incident, it can be difficult to get the proper policy changes instituted in all units through the proper organizational channels. Units are also somewhat isolated in the sense that, if one unit has already solved a particular recurring problem, other units may not be aware of it and therefore will miss the opportunity to adopt the same solution. However, like the fear of punishment, this likely varies throughout the hospital depending on the overall attitudes of and relationships between specific units. One unit manager participant praised another specific unit for their degree of cooperation.

There is presently no method for providing widespread feedback on participation in the incident reporting process, and interviewees report difficulty communicating with other units, both for coordinating with them and learning from them. A newsletter analogous to the ASRS’s *Callback* previously existed but was discontinued; as of the end of interview data collection, a new procedure is currently being set up consisting of short daily meetings of unit managers to share recent important incidents and followups. Formal records of these followups, however, are not necessarily available due to an apparent oversight in the design

of the followup interface: unit managers may submit their forms blank, leaving it unclear in the report record whether an incident was addressed or, if so, how. This likely is a strong contributor to the problem of units being unable to learn from each other.

**Ancillary findings.** Some details about the app and the organization turned up during interviews that are not clearly classifiable but are still interesting and potentially relevant to improving the reporting process. One participant offered two suggestions for specific improvements to the process which are not already addressed by the above items. The participant suggested that the aggregate reports in use should track the rates of the causes of incidents and that the reporting interface could be improved by adding prompts for appropriate actions to take in response to the incident type, such as reminding a staffmember who is accidentally stuck with a needle to go to Employee Health after entering the incident report. Two participants mentioned that users attempt, erroneously, to use the reporting app to report ongoing problems such as dirty facilities or faulty alarms. Although certainly issues worthy of attention that could have safety impacts, they are not considered incidents because they do not involve *specific* instances of harm or potential harm to *specific* patients or staff. This seems to be an inevitable effect of attempting to report a particular kind of problem that the system is not designed to handle.

Although they were not asked to, two participants incidentally mentioned the amount of work experience they had. One participant had only worked at the hospital for about a year, which raised the possibility that hospital employees may be insufficiently experienced using the system or insufficiently assimilated into the organizational culture. As this is a potential confound but not directly part of any research question, this question was added to later data collection (see “Approach 3: Surveys”).

## **Approach 2: Usability Tests**

For the second phase of data collection, I observed natural-like use of three different interfaces contained within the reporting app.

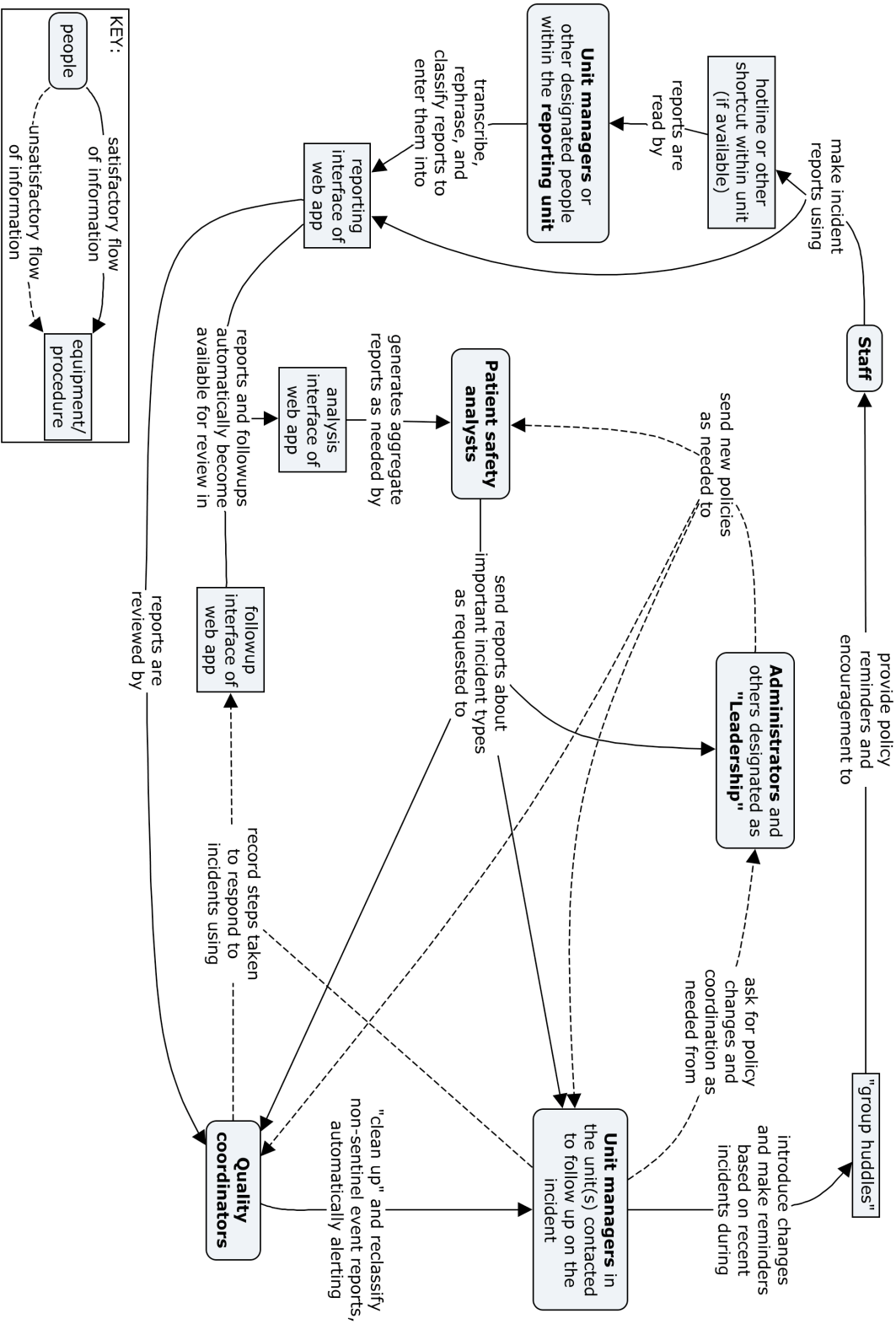


Figure 3. Flowchart showing the cycle of incident report information at the hospital.



## Method

**Participants.** Participants in this portion of the study were four of the six interview participants, the principal investigator (PI), and the thesis advisor (TA). For assessing the interface used for creating incident reports, the two nurse non-managers, the nurse manager, the PI, and the TA participated. The non-medical manager was also asked to participate in this task, but declined, reporting having never entered a single incident report. The non-medical non-manager was not asked to participate due to technical difficulties. Two participants (one unit manager and the analyst) participated in assessment of the followup interface, and one (the trend analyst) participated in assessment of the aggregate interface. As in Approach 1, participants' ages and genders were not collected for privacy.

**Materials and procedure.** Assessment of the report-creation interface used the "training module", which replicates the interface of the fully-functional version in every way except that no information entered is actually saved as an incident report. Preliminary testing was performed by the PI exploring the options of the interface unstructured, followed by a pilot test of the scenario by the TA. All participants including the PI and TA were presented with the training module on a Dell Latitude D820 laptop computer running Microsoft Windows 7. All participants other than the PI also received a paper copy of a narrative describing an incident based on a published case report (Gupta & Cook, 2013). They were instructed to read the narrative and enter an incident report into the reporting app as if they had been asked to do so by an overworked colleague.

The incident narrative described a medication error in which a patient is given an incorrect treatment at first, but the treatment is changed to the correct one with no noticeable harm done. This scenario was chosen because medication errors are uncontroversially worthy of reporting, and the choice to present it as if it were a request from a colleague rather than a first-hand experience was made so that participants employed in non-medical fields could also perform the task without objecting that it would be implausible for them to tend to a patient. The incident narrative is included as Appendix B.

Discussions with various authorities in the hospital and with the participant ruled out

direct access to actual incident reports as examples due to legal and privacy concerns, so the followup interface was assessed informally without exploring the interface or a natural-like observation of use via a brief examination of the actual interface and screenshots of it used by the hospital for training.

The aggregate interface was assessed by asking the participant to arbitrarily choose items to export in an example aggregate report, so that the normal procedure of selecting options and using them to generate a spreadsheet could be observed.

For all participants who were interviewed, the usability test immediately followed the interview, so audio recording for the non-analyst participants and note-taking for the analyst participant continued from the previous task. Due to the unavailability of satisfactory screen-capture software (see “Limitations and Recommendations for Future Use of This Method” below), the think-aloud method (Rubin & Chisnell, 2008, p 204–206) was applied instead. Participants in the report scenario task were asked to narrate their use of the reporting app step-by-step and to volunteer any additional comments that came to mind about the interface or process as they entered their incident reports. Additionally, I took notes to ensure that behaviors or obstacles not noted aloud by the participants were still recorded.

The observed interfaces and user behaviors were compared against three sets of guidelines: those of Bargas-Avila et al. (2010), Nielsen (1994), and the IOM (2004). Bargas-Avila and colleagues created a set of domain-specific guidelines intended to address not just general usability issues but also the specific problems of filling out forms. Nielsen’s heuristics somewhat overlap with these, but are more well-established and may be used to detect and describe problems not anticipated in this specific context. They are comparable to competing methods in their effectiveness (Liljegren, 2006; Hvannberg, Law, & Lárusdóttir, 2007). Finally, the IOM’s criteria for a good reporting system are used to determine whether the reporting app includes ways to collect all of the information that is necessary for learning from incidents, which may be used to evaluate both the existing prompts and the complete record of the incident report and the organizational response that they are used to obtain.

## Results

Comparisons of the reporting app interfaces to Bargas-Avila and colleagues' guidelines are described in Table 1, Nielsen's in Table 2, and the IOM's in Table 3. Where violations were found, or where additional explanation of a specific guideline is necessary, they are described in greater detail in long form. The incident reporting and followup interfaces were rated according to compliance with all three sets of guidelines. Because the Bargas-Avila guidelines only apply to forms to be filled out, the aggregate interface was only rated on the relevant Nielsen heuristics and IOM guidelines.

Table 1. *Reporting App Compliance with Bargas-Avila and Colleagues' (2010) Guidelines for Web Forms.*

Category	Description	Violations
Form content	Forms should be short and intuitively-ordered, allow flexibility in answers, and clearly distinguish between required and optional questions	Yes
Form layout	Fields for answers should be labeled above, listed one per row, and sized appropriately for the length of the answer	No
Input types	Menus, checkboxes, buttons, or open response fields are appropriate for different numbers and types of answers. Options should be in an intuitive order and limited in number if possible	Yes
Error handling	Answers should have their expected formats clearly indicated and should not be cleared by errors. Error messages should be polite, informative, embedded in the form, and easily noticed	Yes
Form submission	Submission buttons should only be usable once and not be confusable with "reset" buttons. Submissions and how they will be used should be confirmed	Yes

**Bargas-Avila guideline issues.** Violations in four of the five categories were found. These violation are described in detail below.

*Form content-related.* One section of the reporting interface form produces a pop-up window with its own "save" and "close" options separate from the rest of the form, while others cause additional fields to appear conditionally within the same window. This inconsistency violates the guideline to keep the form as simple as possible and could distract

or confuse the user. Although the form allows the user to save and resume making a report at any stage, the time taken to complete a report may itself discourage the use of the reporting interface. On the followup interface, despite all fields being marked with the asterisk that indicates they are mandatory, it is possible to submit a completely blank report.

*Input type-related.* The reporting interface consists largely of drop-down menus used for categorizing incidents by unit, severity, type, and so on. Participants had some difficulty finding the appropriate items both on longer menus and from lists generated by search boxes. One selected the wrong item from the search results list when trying to describe the medication error in the scenario, failed to notice this error, and completed the incident report with the incorrect information still entered. Participants also disagreed on the classification of the incident in the scenario, with one calling it “wrong treatment or therapy” while the others called it a “medication event”. Participants even disagreed on whether the incident should be classified as an actual incident (one participant) or a near miss (all others).

Additional menu problems were detected in the preliminary test, but did not occur during any of the tests involving the incident scenario. Menus sometimes lack responses equivalent to “no” or “not applicable” in situations where they would be expected (e.g., as a response to “was another area involved?” when indicating the department where the incident took place). Not all date fields have format templates.

*Error handling-related* Screenshots of error messages obtained from one of the participants indicate that error messages are extremely vague. In one example, an error message indicated that a report could not be submitted because the user needed to “correct data” but failed to explain what entered information specifically needed to be changed. Another error message reviewed not only failed to explain to the user what the problem was, but did not give an error name or code to be passed on to someone else using the app’s technical support.

*Form submission-related.* While it is unclear whether the reporting interface is set up to prevent accidental multiple submissions of the same report, the unclearness is because it directs the user almost instantly to a guideline-compliant confirmation screen. The system’s

response to submission lacks any indication of what, if anything, will be done with the report.

Table 2. *Reporting App Compliance with Nielsen's (1994) General Usability Heuristics.*

Heuristic	Description	Violations
Visibility of system status	The system should provide appropriate and timely feedback	Yes
Match between system and the real world	The system should present things in familiar terms similar to those encountered in other contexts	Yes
User control and freedom	The system should allow users to undo and redo actions	Yes
Consistency and standards	Terms should use consistent and conventional definitions	Yes
Error prevention	The system should be designed to avoid errors rather than simply address them	Yes
Recognition rather than recall	The user should have easy access to references and not need to store and recall information from one part to another	Yes
Flexibility and efficiency of use	There should be customization and shortcut tools available for experienced users to adapt the system to their needs	Unclear
Aesthetic and minimalist design	Dialogues should avoid displaying unnecessary information	No
Help users recognize, diagnose, and recover from errors	Error messages should be presented in terms the user can understand and suggest ways to address the error	No
Help and documentation	Help tools should be readily available, searchable, and explain things step by step	Unclear

**Nielsen heuristics issues.** Violations of seven of the ten heuristics were found, with evidence ambiguous about one additional heuristic.

*Visibility of system status.* When an unsubmitted report is left untouched for approximately three minutes, the contents of the form are briefly replaced with a hard-to-read redirection message before returning to its previous state, which may startle or interrupt a user who is still at the computer but not actively typing.

*Match between system and the real world.* For the reporting and followup interfaces, this is addressed by the Bargas-Avila form content guidelines. In addition to the issues listed

there, the list of possible fields on the aggregate interface is not in an easily-recognizable order.

*User control and freedom.* Changing a response to such a question clears responses to all “downstream” items, e.g., changing the answer to “What is the nature of the event?” for a medication-related incident will clear all entered medications and doses even though the same medication information is still relevant for different types of medication incidents.

*Consistency and standards.* This is addressed by the Bargas-Avila input type guidelines.

*Error prevention.* The reporting app generally provides no confirmation or instruction to review entered items when the form is saved. In the aggregate interface, certain characters cannot be copied-and-pasted into text boxes, but the app does not indicate this.

*Recognition rather than recall.* By definition, the reporting interface’s instruction to “provide a brief factual narrative” relies on recall. The user answers most questions using drop-down menus, templates, or search features that suggest possible matches while the user searches, allowing at least some use of recognition. The followup interface consists entirely of free response questions, relying entirely on recall over recognition.

*Flexibility and efficiency of use.* It was unclear from testing whether the reporting and followup interfaces had any opportunities for customization or shortcuts. The list of fields on the aggregate interface is customizable, and may be reordered for easy use, according to a followup e-mail exchange with the analyst participant.

*Help users recognize, diagnose, and recover from errors.* This is addressed by the Bargas-Avila error handling guidelines.

*Help and documentation.* Other than a single tutorial video which features an overview, acquiring any help using the system requires non-anonymously accessing a support center by logging in to it or contacting a person via e-mail or phone.

**IOM guideline issues.** The total stored data about the incident, including both the incident report and the record of any followup measures taken, are referred to in Table 3 as the “report record”. Comparison of the information requested by the reporting app

Table 3. *Reporting App Compliance with Institute of Medicine (IOM) Guidelines.*

IOM Guideline	Reporting Interface Prompts	Report Records Contain
<b>Discovery</b>		
1. Who discovered/reported the event? (roles, not names)	Yes	Yes
2. How?	Yes	Yes
<b>The Event Itself</b>		
3. What happened? (Type of event)	Yes	Yes
4. Where in the care process was it discovered/did it occur?	Yes	Yes
5. When did the event occur?	Yes	Yes
6. Who was involved? (functions, not names)	Yes	Yes
7. Why did it occur? (Dominant cause based on preliminary analysis)	Yes	Yes
8. Risk assessment:		
8a. Severity	Yes	Yes
8b. Preventability	No	No
8c. Likelihood of recurrence	No	No
9. Narrative of the event, including contributing factors.	Yes	Yes
<b>Ancillary Information</b>		
10. Product information (blood, devices, drugs, etc. that were involved).	[1]	[1]
11. Patient information (age, gender, ethnicity, diagnosis, procedures, and comorbid conditions)	[2]	[2]
<b>Detailed Causal Analysis (if deemed necessary)</b>		
12. Technical, organizational, and human factors associated with the Eindhoven model (RCA).	NA	[3]
13. Recovery factors that can occur at each point for near misses.	NA	No
14. Corrective actions that were taken to recover from the incident.	NA	No
15. Patient outcome as a result of the corrective actions taken.	NA	No
16. Whether a similar case has recently been investigated.	NA	Yes
<b>Lessons Learned</b>	NA	No

Notes: [1] Depends on the indicated type of incident and/or answer to “was any equipment involved?” [2] Identifying the patient by name and medical record number is mandatory, so this information can be retrieved if thought relevant. [3] Root Cause Analyses (RCAs) are only conducted for sentinel events, which are outside the scope of this study. See Deeter and Rantanen (2012) for a discussion of RCA as it is applied in this particular hospital and comparison to a possible alternate analysis method.

against the IOM's guidelines for a good incident reporting system revealed compliance with the vast majority of the relevant features. Items labeled "no" in the table are those for which the reporting app lacks a specific prompt for the user to provide the type of information described by the guideline. Items labeled "NA" are those that are not concerned with the original incident report and would therefore not be expected to occur in the reporting interface. It should be cautioned that there are many optional fields on the reporting interface, and users of the followup interface may even submit a completely blank followup, so a "complete" report is not guaranteed to contain all of the prompted information, merely to contain prompts for it. The version of the report displayed in the followup interface contains all information that was entered by the original reporter, in the same order and with the same labels (excepting, of course, any changes made by QCs based on the incident narrative details).

Neither the reporting nor followup interface asks users to rate the preventability or likelihood of recurrence of an incident as part of their incident report or followup. There are no required fields in the followup interface, so unit managers also do not need to identify recovery factors, describe the patient outcome, or even necessarily report what followup they made to the incident. As long as the incident report is properly labeled by the original reporter, however, the aggregate interface may be used to determine whether similar incidents have recently been reported and addressed by generating a table of whatever criteria are of interest.

**Ancillary findings.** Although not strictly part of any of the above usability guidelines, the reporting app also has some additional worrisome features. The app is web-based and only compatible with certain specific versions of Microsoft Internet Explorer (IE), which may seriously limit software compatibility as support for older browsers is gradually dropped. This proved to be a problem when setting up the laptop that was used for usability testing, as the appropriate version of IE had to be manually installed and the automatic update feature of Microsoft Windows had to be manually turned off to avoid automatically downloading and reinstalling a newer, incompatible version of IE. In most of



the usability tests, the app loaded quickly and without serious problems, but during one test, the simple act of accessing the app took about two minutes. The participant for that test confirmed that such delays are common. This implies that some unknown issue(s) with the app, the browser, or the internet service at the hospital is/are artificially extending the time required to report each incident.

### **Approach 3: Surveys**

Although the interviews and usability tests identified a number of likely obstacles and a few genuine benefits in the current system, their very small sample size means they cannot be relied upon to determine how common issues are, and may have missed other issues completely. For the third phase of data collection, I administered two hospital-wide surveys to judge the frequency of the various obstacles and helpful features identified in the previous approaches, as well as to answer questions that the previous approaches raised or left unanswered.

#### **Method**

**Participants.** The hospital's Chief Medical Officer recruited 198 participants via an e-mail drafted by the authors. Participants were staff members of the hospital involved in the above approaches and a second, smaller hospital in the region owned by the same parent company. We expected that participants would indicate which of the two hospitals they worked at using the "primary unit or work area" field, but very few did. However, because both hospitals use the same incident-reporting app and are part of the same larger administrative structure, we did not attempt to separate data by facility.

Of the original 198 participants, 142 (72%) indicated that they did not analyze incidents, and were directed to the "user version" of the survey and the remaining 56 (28%) indicated that they did, and were directed to the "analyst version" of the survey. For clarity and brevity, user version participants are called "UPs" and analyst version users are called "APs" below. Nineteen UPs and 2 APs declined to answer any further questions after the

demographics, and were excluded from analysis. UPs who answered some questions but not others were excluded from analysis on the specific questions they did not answer. One AP indicated in the free response questions that they did not actually review submitted incident reports; this AP's answers to two of the free response questions (asking for an example of an underreported incident type and for "any other comments about incident reporting") were coded and included below in the analysts' free response answers, but this participant was excluded from analysis on all other questions. This yielded a total of 123 UPs and 53 APs who gave valid responses to some or all of the multiple choice questions. Valid responses to free response questions varied, and are described individually in "Results", below.

Breakdowns of participants' responses to the demographic questions are found in Tables 4, and 5. APs had a median of 13 years of experience at the hospital (range: 0.08 – 37 years), while UPs had a median of 8.5 years of experience (range: 0.42 – 41 years). As in Approaches 1 and 2, participants' ages and genders were not collected for privacy.

Table 4. *Participant Demographics: Role.*

Role	Number in User Survey (% of UPs)	Number in Analyst Survey (% of APs)	Total (% of All Participants)
Manager of a Medical Unit	1 (< 1%)	13 (24%)	14 (7.9%)
Manager of a Non-Medical Unit	0 (0%)	2 (3.8%)	2 (1.1%)
Nurse	118 (96%)	28 (52%)	146 (82%)
Physician	1 (< 1%)	8 (15%)	9 (5.1%)
Other Medical Role	1 (< 1%)	1 (1.9%)	2 (1.1%)
Other Non-Medical Role	1 (< 1%)	0 (0%)	1 (< 1%)
Other Unclassified Role*	1 (< 1%)	1 (1.9%)	2 (1.1%)
No Answer	0 (0%)	1 (1.9%)	1 (< 1%)

\*Self-reported roles that are difficult to classify into the other categories.

**Materials and procedure.** Results from the observation and interviews were used, along with other likely problems predicted based on the successes of ASRS and on the safety culture surveys and usability guidelines cited above, to generate questions for online surveys. Pilot testing with five students in a graduate Research Methods class was used to revise the phrasing of certain questions to be more easily understood. Participants were asked

Table 5. *Participant Demographics: Incident Reporting Frequency.*

Incidents Reported in the Previous Year	Number in User Survey (% of UPs)	Number in Analyst Survey (% of APs)	Total (% of All Participants)
0	11 (8.9%)	10 (19%)	21 (12%)
1–2	43 (35%)	15 (28%)	58 (33%)
3–5	37 (30%)	7 (13%)	44 (25%)
6–10	15 (12%)	14 (26%)	29 (16%)
11–20	11 (8.9%)	6 (11%)	17 (9.6%)
21 or more	6 (4.9%)	2 (3.7%)	8 (4.5%)

first about their job categories and work units (as was the case in the interviews), then about their years of work experience and frequency of use of the reporting app. They then indicated whether or not they analyzed/followed up on incidents that others reported. If they said they did, they were directed to the analyst version of the survey. If they said they did not, they were directed to the regular user version. Participants took the survey via an online survey tool called “Clipboard” developed and hosted by RIT. The specific questions used in each version were different as appropriate to the context, but each version consisted of two blocks of Likert-type scale questions and a small number of followup free response questions. Questions were ordered so that those considered less personal or potentially embarrassing were presented first, in accordance with survey design recommendations around “objectionable questions” (Dillman, 2000, p 73–77). The complete survey questions are included as Appendix C.

In the first block of each survey, participants were asked to indicate the extent to which they agreed with various statements using 5-point Likert-type scales anchored at all points with “strongly disagree”, “somewhat disagree”, “neither agree nor disagree”, “somewhat agree”, and “strongly agree”, then to provide specific examples of the types of situations being asked about (e.g., to provide an example of a category of incident they would like to see added to the drop-down menus). In the second block, they were asked to indicate the frequency with which specific types of events occurred while they were trying to report or analyze an incident. These also used 5-point Likert-type scales, with the anchors being

“never”, “rarely”, “sometimes”, “often”, and “always”.

## Results

**Free response questions.** Free response questions to surveys were analyzed by grouping conceptually-similar answers together to create codes, then grouping those coded answers together into categories. Most free response questions asked for specific examples or improvement ideas, and so were coded simply by summarizing the specific example or improvement offered by the participant. The other questions asked for “any additional comments”. Note that the number of responses falling into a particular category do not necessarily sum to the number of usable answers for that category, as some participants touched on multiple topics in their answers, and that the numbers of participants who chose to answer each question are quite different.

UPs were asked for examples of confusing incident classification categories and categories that would be helpful to have but are currently not included on menus. When asked about confusing incident classifications, 11 of the 61 UPs who answered the question suggested simplifications, including 5 who said it was difficult to find the appropriate items on menus and 4 who suggested that the app should be made entirely free-response. Nine said that it was not always possible to determine the level of harm to a patient. Similarly, 9 said they were unsure of the meanings of specific options, including 3 who had difficulty understanding the levels of harm to patients and 5 who said terms used in the app were generally hard to understand or poorly-defined. The most common response for confusing incident classifications, given by 26 UPs, was that particular options were missing; however, half of such answers did not give a specific example, and no specific type of incident that should be added to the menus was mentioned by more than three UPs.

When asked specifically about missing menu options, 11 of 38 UPs gave responses which included at least one specific item they would like to see added. This yielded 20 specific option suggestions, each mentioned by exactly one UP, and 7 general categories of missing options, only one of which (skin issues) was mentioned by more than one UP. Ten

UPs said they could not think of a specific problem, and 4 simply said “see above” because they had already answered this question in their response to the previous question. One UP suggested making more room for users to enter “other” responses, seemingly anticipating the large variety of other responses.

APs were asked for examples of problems they could not fix due to underreporting, confusing classifications in reports they receive, and reports that did not contain enough information to address the incidents described. As above, there were many unique responses and therefore little confidence in what constituted an ongoing problem. When asked about underreporting, 4 of 17 APs gave examples of specific types of incidents, 4 gave problems in patient record-keeping, and 5 gave problems in general information flow throughout the hospital. The remaining answers dealt with general quality issues that do not qualify as incidents, or did not give a relevant example at all.

When asked about confusing classifications in reports, responses were even more scattered. Out of 13 APs who answered, only 9 gave specific examples. Their answers covered 12 different issues, only one of which was identified by more than one AP: 3 APs answered that reports were being sent to the wrong people for followup.

When asked about missing information in reports, 10 of 19 APs identified specific missing information, including 8 who said the free response narratives to describe incidents were too short, missing numerous details. Nine APs indicated having to seek additional information about incidents in order to decide how to address them, including 6 who consulted patient charts and 6 who asked for additional explanation directly from the staff involved in the incidents. One AP specifically singled out their unit’s reporting shortcut as the cause of the un-detailed reports.

UPs’ and APs’ responses to “any other comments?” included many specific suggestions for improving the app (17 of 40 UPs, 9 of 20 APs) or organization (24 UPs, 6 APs), but these suggestions were, again, widely scattered and those suggestions given by more than one participant were often reiterations of their responses to earlier questions (both free response and rating scale questions), such as indicating that reporting was time-consuming

(5 UPs and 2 APs, as well as 6 UPs who cited high workload taking precedence over reporting). The other most common responses were that the system is used punitively or in some way results in fear or resentment (10 UPs, 2 APs), complaints about a lack of feedback on how reports are used (5 UPs, 5 APs), calls to shorten or simplify tedious aspects of the form (4 UPs, 4 APs), some form of praise for shortcut reporting methods either already in place or that the participant believes should be introduced (3 UPs, 2 APs), and that near-miss reporting should be encouraged more (4 UPs).

**Pre-analysis for rating scale questions: Demographic trends.** Demographics were not considered when coding and categorizing free responses. However, they were considered prior to analyzing the rating scale (agreement and frequency) questions because they could serve as explanations or confounds for particular trends.

Unsurprisingly, years of experience at the hospital and years of experience in health-care overall were fairly highly correlated (Pearson's  $r = 0.656$  for the analyst survey,  $r = 0.633$  for the user survey, both  $ps < 0.001$ ). To avoid confusion due to covariation, and because years of experience at the hospital was conceptually more relevant to questions about the hospital's organizational culture, only years of experience at the hospital were included in the further analyses. Consequently, references to "experience" below refer to years of experience at the hospital.

For each survey version, role and experience were examined as possible predictors of reporting rate, using PLUM method ordinal regressions with role as the factor, experience as the covariate, and number of reports in the past year as the DV. Among APs, nurses filed fewer incident reports in the past year than did participants in other roles, odds ratio 0.120 compared to non-nurses (95% CI 0.026 to 0.553), Wald  $\chi^2 = 7.395$ ,  $p = 0.007$ . Other roles and years of experience did not appear related to number of reports filed for either APs or UPs, all Wald  $\chi^2 < 1.9$ , all  $ps > 0.2$ .

Participants came from a very large number of units, with a very small number of participants working in each unit, so participant unit was not included in statistical tests due to the high risk of false positive results. Instead, the only survey responses examined

on the level of individual units were those from the units identified using the free response answers as likely locations of endemic problems. Almost all of the reports of punitive use in the free responses came from three closely-related units representing one specialty. Although only one AP singled out the shortcut in use in their unit as a cause of missing information in narratives, it is important to check the other responses from that unit to determine whether this is a plausible cause of problems there because, despite support among staff for the existing shortcuts and the introduction of more, there is good reason from the literature to think they do not work as well as other methods because the lack of prompts would result in frequent omission of helpful information.

**Rating scale questions.** Some of the user survey items (e.g., “I tend to report as much as I can, even if it means making some reports that aren’t really necessary”) were ambiguous as to whether they indicated helpful or harmful conditions. For these items, answers indicating agreement were coded as negative and disagreement as positive because the purpose of all such questions was to gauge the prevalence of potential sources of inappropriate reports and underreporting identified in the interview phase. I then proceeded to create a way to rank these items by the approximate strength of this disagreement. Likert scale responses are, strictly speaking, ordinal data, so although the following data transformation is reasonable because ordinal attitude data consistently seems to behave very much like linear interval data (Norman, 2010), I caution that this approach is only used here, and should only be used, to generally determine whether participants’ attitudes were positive or negative and to rank overall attitudes in order of their relative strengths, not to assign precise numerical values to the absolute strengths of those attitudes.

To determine the highest-priority issues, responses were recoded according to valence, so that frequency and agreement data can be compared directly by interpreting them as indicators of the presence or absence of particular problems or helpful features. For all items rated on level of agreement with the item, answers of “neither agree nor disagree” were excluded as being difficult to interpret and other answers were recoded according to whether they indicated a harmful condition (“negative”) or a helpful condition (“positive”).

For example, the answer “agree” would be coded as positive in response to the item “my unit is good at avoiding mistakes” but as negative in response to the item “using the reporting app is tedious”. Similarly, for items rated on the frequency of the item, answers that indicated a helpful condition occurring more often than not were coded “positive” and those that indicated a harmful condition occurring at least as often as not were coded “negative”. That is, the middle option of “sometimes” was coded as negative because it indicated that a positive condition occurred no more than half the time. The extreme answers, “always”, “never”, “strongly agree”, and “strongly disagree”, were weighted as twice as important as others in order to generate an approximate measure, valence, of how strongly participants felt a problem was present or absent.

For items that were possibly related to the unit-specific issues identified above, a score of “net valence” was constructed. This allowed the valences of all responses by participants in the unit(s) in question to be compared to determine whether participants overall agreed that those issues were present, as well as to find possible causes of them. Net valence  $V$  was calculated as:

$$V = (2VP + P) - (N + 2VN)$$

where  $VP$  = the number of answers coded “very positive”,  $P$  = number coded “positive”,  $N$  = number coded “negative”, and  $VN$  = number coded “very negative”.

The units reporting punitive use of the system were examined first. For most of the relevant items (e.g., “mistakes lead to positive change”, “I get positive feedback when I follow safety procedures”),  $V$  ranged from 2 to -2, with most individual participants’ scores either neutral or weakly positive, indicating that staff attitudes towards their supervisors and general confidence in the system is not seriously lacking. The remaining items examined all had  $V \leq -3$ . This included two items asking about punitiveness and one about anonymity, showing a general trend towards distrusting the way the system is applied.

The unit that used a shortcut reporting method was then assessed the same way, with responses that might be relevant to the question of scored for net valence. Although UPs



indicated on three different measures that they were rushed for time ( $V \leq -5$ ) and often handled incidents without reporting them ( $V = -11$ ), they seemed satisfied with the speed of the shortcut reporting method ( $V = 1$ ). APs, however, indicated that they were not receiving enough reports ( $V = -8$ ) and that the reports they did receive tended to not have enough detail ( $V = -3$ ).

Net valence may be a helpful tool for comparing similar numbers of responses, but for samples with large differences in size, it can be very misleading based on wildly unequal numbers of responses to different questions. Therefore, in order to rank all rating scale questions by how urgently UPs believe they could be improved, all rating scale responses were standardized based on valence and number of responses as follows:

$$U = \frac{2VN + N}{2(VP + P + N + VN)}$$

yielding a general rating of perceived urgency,  $U$ , which is the weighted proportion of negative responses. This makeshift “urgency index” is theoretically able to range from 0 (if nobody believed it was a problem) to 1 (if everyone believed it was a severe problem).

Table 6. *Risk Matrix Used for Prioritization.*

Severity	Urgency Index:			
	$\leq 0.09$ (Rare)	0.10–0.24 (Occasional)	0.25–0.49 (Common)	$\geq 0.50$ (Chronic)
Serious				High
Moderate			Medium	
Minor	Low			

This index was then used along with the criteria put forward by Cook and Herout (2015) for ranking usability problems by severity. Problems are ranked “Serious”, “Moderate”, “Minor”, or “Not Applicable” according to the effect of a usability problem on use of the tool being assessed. “Serious” problems are those which cause a user to abandon a task or render the task impossible. “Moderate” problems are those which cause setbacks or delays that are not severe enough to prevent use, but which may make the user so dissat-

Table 7. *Priorities for UPs' Usability Problems.*

Question	Urgency Index	Severity	Priority
Some types or details of incidents are not included on [the app]'s drop-down menus, but should be.	0.53	Moderate	High
The way [the app] asks me to classify incidents is confusing.	0.52	Moderate	High
My incident reports get ruined by [the app] having an error or crashing.	0.10	Serious	High
I give up on reporting a particular incident because of [the app] having an error or crashing.	0.10	Serious	High
I wish my unit had a faster way of reporting incidents.	0.50	Minor	Medium
Using [the app] is tedious.	0.44	Moderate	Medium
If something goes wrong with [the app], I can easily find out how to fix it or work around it.	0.44	Moderate	Medium
When I use [the app], I have trouble finding the items I need on menus.	0.40	Minor	Medium
Anonymous reports aren't really anonymous.	0.23	Moderate	Medium
[The app] asks me for information I am not comfortable providing.	0.07	Moderate	Low

ified that they avoid using the tool. “Minor” problems are those which cause a task to be irritating, confusing, or inefficient but which do not prevent the task from being completed or lead users to stop using the tool. Some items clearly mapped to severity levels, such as “I give up on reporting a particular incident because of [the app] having an error or crashing,” which falls in the serious category because it entails the app encountering an error or crashing, one of the possible criteria that can lead a problem to be rated serious. For those questions that were more ambiguous, severity was determined with the aid of comments made in the interview and free response answers to clarify whether encountering particular problems can result in particular relevant consequences, like users avoiding the app entirely. Although these severity criteria are intended for assessment of usability features of software, they are phrased in such a way that they may be generalized to apply to safety culture features that affect willingness and ability to make incident reports. Since each question in the survey measures the prevalence of a problem, no items were coded as

Table 8. *Priorities for APs' Usability Problems.*

Question	Urgency Index	Severity	Priority
There are certain things I am sure are problems, but I can't fix them because they don't get reported.	0.38	Serious	High
I don't get enough information in incident reports to figure out how to respond to them.	0.27	Serious	High
My analyses get ruined by [the app] having an error or crashing.	0.10	Serious	High
Using [the app] is tedious.	0.41	Moderate	Medium
If something goes wrong with [the app], I can easily find out how to fix it or work around it.	0.38	Moderate	Medium
When I use [the app], I have trouble finding the items I need on menus.	0.41	Minor	Medium
I get confused by the way incidents are classified in the reports I receive.	0.29	Minor	Medium
People are making too many reports about things that don't really count as "incidents".	0.14	Moderate	Medium
Incident reports are phrased like the reporter is hiding something.	0.14	Moderate	Medium
I give up on analyzing an incident report because of [the app] having an error or crashing.	0.09	Serious	Medium
People are making too many reports about things that are not in my power to fix.	0.06	Minor	Low

"Not Applicable".

Problem rankings and usability index scores were then used to create a risk matrix, which allows for each problem to be coded as high, medium, or low-priority for the hospital or software developer to address. The risk matrix is shown in Table 6, and the survey items with their urgency, severity, and overall priority scores in Tables 7, 8, 9, and 10. Usability issues were the most likely to be classified as high priority. This may be attributed to the overall positive scores participants gave to most aspects of hospital safety culture, except for those related to punitive use of the system. In addition to being rare, safety culture problems also tended to be ranked as less severe because few of them would lead directly to the user being unable to complete a task. Time pressure and tradeoffs between reporting or analysis and other tasks, which are broader organizational issues not necessarily indicative

Table 9. *Priorities for UPs’ Culture/Organizational Problems.*

Question	Urgency Index	Severity	Priority
I get interrupted while making an incident report and end up losing the report.	0.19	Serious	High
If I made more reports, I would not have time to do other important parts of my job.	0.48	Moderate	Medium
I feel like I have to rush through work.	0.43	Moderate	Medium
I try to address incidents without reporting them by talking directly with whomever was involved.	0.31	Moderate	Medium
When we change how we do something in response to an incident, the change makes sense to me.	0.26	Minor	Medium
People make incident reports to try to get each other in trouble.	0.19	Moderate	Medium
I tend to focus on the incidents I think are most important, even if it means not reporting some other incidents.	0.18	Moderate	Medium
I get positive feedback when I follow safety procedures.	0.17	Moderate	Medium
My mistakes are held against me.	0.12	Moderate	Medium
I tend to report as much as I can, even if it means making some reports that aren’t really necessary.	0.08	Minor	Low
I am comfortable reporting incidents.	0.06	Moderate	Low
If nobody has addressed an ongoing problem, I will make sure to report it the next time it happens to try to get someone to fix it.	0.05	Moderate	Low
My supervisor/manager responds helpfully to suggestions for how to improve safety.	0.05	Moderate	Low
Mistakes lead to positive change.	0.04	Moderate	Low
My unit is good at avoiding mistakes.	0.03	Minor	Low
It’s helpful to report near misses.	0.02	Moderate	Low

of safety culture problems, were classified as medium priority for both UPs and APs.

**General Discussion**

**Answering Research Question 1**

*To what extent is underreporting of incidents due to problems in the usability of this particular reporting app?*

The particular reporting app in use here is promising, but has some serious flaws that

Table 10. *Priorities for APs' Culture/Organizational Problems.*

Question	Urgency Index	Severity	Priority
I wish I knew more about how other departments are addressing incidents.	0.57	Moderate	High
When another department solves a problem my department is also having, I can easily find out about it and try using their solution.	0.53	Moderate	High
I can come up with ways to address incidents that make similar incidents less common or less harmful.	0.30	Serious	High
I get interrupted during incident report analyses, and end up not finishing them	0.22	Serious	High
If an incident is due to problems in more than one department, I can get everyone to make the necessary changes.	0.46	Moderate	Medium
I feel like I have to rush through work.	0.36	Moderate	Medium
If I made more thorough analyses, I would not have time to do other important parts of my job.	0.28	Moderate	Medium
People make incident reports to try to get each other in trouble.	0.23	Minor	Low
My unit is good at avoiding mistakes.	0.04	Minor	Low
Mistakes lead to positive change.	0.03	Moderate	Low
My supervisor/manager responds helpfully to suggestions for how to improve safety.	0.03	Moderate	Low
It's helpful to report near misses.	0.01	Moderate	Low

are likely the largest obstacles to reporting in this hospital. Survey results confirmed that usability issues were the most severe impediments to making incident reports. Problems related to menus and error handling were the most notable.

Consistently throughout all stages of this study, participants encountered problems with navigating drop-down menus. These menus are customizable, meaning that the content and arrangement options is within the hospital's power to improve, but the way menus are currently customized is a serious issue. In usability testing, they found it difficult to find appropriate menu items to describe incidents, both because appropriate items were not always present and because the menus themselves were long and hard to navigate. Interview and survey answers bore this out, with the two most urgent high-priority problems

by far being a lack of useful menu options and complaints that existing menu options are confusing. These may not prevent reports from being made, but they certainly discourage reports from being made by making the system so confusing and inconvenient that users may avoid reporting in the future. However, there was no consensus about how the menus should be changed, even to the point that participants disagreed on whether there should be more or fewer options.

Although much less common than menu-related issues (no related problems occurred in usability testing), error handling was also a common theme throughout the interviews and surveys. The app occasionally encounters software errors that ruin reports in progress, and uninformative error messages occasionally leave users unable to figure out why a report can not be submitted. The result is the loss of a small number of incident reports, but a problem that prevents reporting entirely rather than merely impeding it still should be taken seriously. Although this did not come up in interviews or surveys, this problem could be exacerbated by the fact that little technical support help is available without contacting the makers of the software non-anonymously.

### **Answering Research Question 2**

*To what extent is underreporting of incidents due to problems in the organizational procedures surrounding the use of the reporting app?*

Overall, the safety culture of the hospital seems to be quite sound. With the exception of some questions about punitiveness, which received a rating of “medium priority”, participants had very few doubts that the system was, in principle, a good idea or that it could help improve patient safety. Other hospitals seeking to learn from this case study should take into account that staffmembers’ positive attitudes towards their coworkers and the practice of incident reporting do not guarantee that they are actually able to make reports, and other features of the workplace have to be taken into account when planning policies for incident reporting.

As suggested by the interview results, survey results support the ideas that safety

culture varies by unit and that punitive use is generally uncommon. The specific cases of punitive use mentioned by participants in the survey suggest a problem of distrust between coworkers, and may explain the two most negatively-rated items on the survey among members of those units, “I try to address incidents without reporting them by talking directly with whomever was involved” and “I don’t get enough information in incident reports to figure out how to respond to them”, because fear of punishment is expected to lead to less use of the incident reporting system.

The issue of addressing incidents without reporting them, however, may also be caused by time pressure. The unit investigated more closely to examine its shortcut reporting method still has shortcut and task tradeoff problems despite the presence of a satisfying shortcut. UPs consider their reporting system adequate but handle incidents informally while APs receive insufficient incident reports and insufficient details in the reports they do receive, suggesting that attempting to further reduce the time needed to report will not help boost reporting rate, but that reducing or managing other drivers of workload might<sup>3</sup>. It is also worth considering that the shortcut itself, although making the reporting process more convenient, also makes the reports themselves less detailed, which is discussed in more detail below (see “Expanding and improving the use of time-saving shortcuts”).

Organizational problems that were not necessarily related to safety culture *per se* were the highest-priority in this category. The only problem rated high-priority dealt with interruptions, and other than punitiveness and feedback, all medium-priority issues also appeared to be related to finding the time to make reports, through explicitly mentioning it or hinting at it through informal incident handling and judgement calls about which incidents are most important.

### Answering Research Question 3

*What issues are analysts facing in their use of incident reports to improve patient safety, and to what extent?*

---

<sup>3</sup>For more information on workload in the same hospital discussed in this study, see Umansky (2016), *Workload in nursing*. (Unpublished Master’s thesis) Rochester Institute of Technology.

APs faced several of the same usability issues that UPs did, and rated them as similarly urgent. Unlike UPs, APs encountered usability obstacles in the data provided to them rather than only in the format and content of the reporting app. APs' responses throughout all stages of this study have agreed that incidents reports can often not be analyzed because they simply do not contain the necessary information. The survey results in particular showed that, with the exception of rare but severe software glitches, the APs' most pressing usability problems were related to a lack of information received from incident reports. This is an interesting and unusual type of usability problem to have, because it cannot be remedied directly by altering the format of the report or the user interface, but instead must be remedied by correcting the organizational and interface problems that led to the omission of useful information from the reports in the first place.

Although not as explicitly clear, the various categorization issues suggest an interrater reliability problem. UPs reported confusion about which category labels were the most appropriate descriptions for various incidents, including disagreement about how to classify the example scenario itself, which was expected to be labeled uncontroversially as a "near miss of a medication error". This not only further supports the usability problem of bad menu options mentioned above under "Answering Research Question 1", but also directly leads to a usability problem for the analysts. APs reported finding the categorization of incidents confusing in the reports they received, which may, along with the general lack of information, explain the high degree of uncertainty analysts reported about their own ability to determine useful responses to incidents. In both of these cases, it is reasonable to say that analysts' problems can be addressed incidentally by addressing users' problems.

Interview participants voiced concerns that reports were being filed about inappropriate topics, or that they were being misdirected. The implication of this was that analysts' time was being wasted by having to reject or redirect a large number of "false alarm" incident reports. Survey results, however, suggest that rather than being common, this may just be an occasional problem that analysts find particularly annoying but not particularly troublesome. The failures in the incident reporting process that hinder analysis appear to



be much more due to misses than to false alarms.

On the organizational side, APs were (perhaps unsurprisingly given the nature of their jobs) again similar to UPs in believing that learning from mistakes was useful, positive, and often successful. However, a higher proportion of their organizational issues were high priority than UPs' organizational issues were<sup>4</sup>, and the issues they had in common with UPs were rated as much less urgent by analysts. Instead, APs' organizational issues tended to be those dealing with the structure of the organization itself and the procedures in place. They encountered unique challenges in interpreting and communicating reports, especially in coordinating different departments to reform their procedures together, and a lack of easy access to others' solutions to similar problems. The latter problem in particular is reminiscent of the problems that prompted the creation of ASRS and Partnership for Patients. This should serve as a "take-away message" that those seeking to improve the healthcare system should think about solutions on the scale of individual facilities in addition to state-, nation-, or industry-wide solutions, because the individual organization may experience internal small-scale versions of the same problems that are present industry-wide.

### **Limitations and Recommendations for Future Use of This Method**

#### **Interviews**

Although recommended as part of the CDM, the interview procedure did not include counterfactual scenarios. It is possible that this caused me to overlook unusual problems present in the organization or system by virtue of not asking the right questions. Future researchers and human factors professionals attempting to use CDM in this context should come prepared to offer counterfactuals, although such questions are admittedly difficult to generate beforehand. Perhaps it would be useful to identify example incidents reported or witnessed but not reported to discuss in advance, then to create alternative scenarios to explore based on them, or to (as Crandall, Klein, and Hoffman suggest themselves) conduct

---

<sup>4</sup>This includes a common doubt ( $U = 0.30$ ) that APs had about their own abilities to address incidents constructively, which seems to contradict that confidence in learning from mistakes.

the interview over multiple sessions to extract more information. However, either of these should be done with care to avoid prompting participants about what kinds of answers they “should” give. Although retelling an event can be reasonably expected to elicit more details (Nadel, Campbell, & Ryan, 2007), doing so with particular goals and contexts in mind will tend to alter the memory of the event itself, leading to those extra details being biased or embellished (Tversky & Marsh, 2000; Mazzoni & Vannucci, 2007, e.g.). Another approach could be to recruit others who have experience at the specific task being studied to assist in conducting the CDM interviews, which might also add the comfort of a peer being present, where an outside interviewer might make the participant uncomfortable and self-censoring.

### **Usability Tests**

Due to difficulty recruiting and scheduling participants, evaluation of the reporting interface was based on the experience of only five users total, including two who did not work for the hospital (myself and my advisor), and evaluation of the aggregate interface was based on only one user’s experience. Due to privacy and legal concerns, the followup interface was only evaluated based on training screenshots rather than actual use, essentially making it not only analysis of a single user but a cursory and unrepresentative analysis at that. It is likely that there are many usability issues that were not detected by this study, either because they were simply not noticed or because the correct condition needed to trigger a particular problem was not met. Keeping in mind that this was, strictly speaking, a user test using heuristics as the method for classifying problems rather than a heuristic evaluation *per se*, we can look to the model by Tan, Liu, and Bishu (2009) based on different methods of evaluating websites. This model would suggest that only about 10% of the followup and aggregate interfaces’ problems and about 35% of the reporting interface’s problems were detected. Applying the more generous claims made by Nielsen (1994), this method would have been sufficient to identify 29% of followup and aggregate interface problems and 75% of reporting interface problems. In either case, the surveys verified which of those that were detected were common problems and allowed participants to volunteer any overlooked

problems they recalled encountering in the past.

An alternate method for increasing the proportion of usability problems found would be to use a different method of recording the data. Initially, I had intended to record the usability tests using screen capture software specifically designed for usability testing; however, the only product that could be located and purchased in time for the scheduled tests did not function as promised (notably, it was a trial version which only recorded about a quarter of the screen area, and the registration code provided at purchase failed to “unlock” the full version of the software) and repeated requests to the manufacturer for technical assistance yielded no help other than a download link for a different trial version of the same software, which also did not work. Therefore, data on topics like the number of clicks needed to perform a task or precisely how long it took users to find desired menu options were not recorded. I strongly encourage future researchers and designers of incident-reporting software to explore it with other usability testing methods, especially in light of the interview results indicating a common dissatisfaction with the time requirements and confusing menu items of the reporting interface. Design changes to the menus that reduce the number of steps required to classify incidents would likely reduce both time requirements and confusion.

Additionally, lacking the opportunity to directly test the incident followup interface means there may be other usability problems unique to that interface which were not evident. Certainly, screenshots used for training purposes would be unlikely to include error messages, for example. However, it is assumed for the purpose of this analysis that, because all three interfaces are different “modes” of the same piece of software, it is likely that they share similar issues with error handling, error prevention, and visibility of system status. Future studies on the usability of incident reporting systems should include a sample complete incident so that realistic followups may be created and entered in the same manner in which the present study presented a sample incident for the sake of a realistic incident report.

## Surveys

The surveys did not gather a large enough number of responses to examine all roles or units thoroughly, and as we recruited participants through intermediaries, we did not collect the total number of people contacted to participate and could not calculate response rates.

One question outside of the scope of this study which may be interesting to future researchers or patient safety professionals is that of priorities. Interviews raised the suggestions that a large enough proportion of incident reports are false positives (inappropriate reports, such as those that cover topics that are not technically incidents) or misses (failures to report important topics). For the purposes of this study, it was assumed that failure to report true incidents was potentially dangerous while reporting non-incidents was merely inconvenient, hence the former being rated “Serious” severity while the latter was rated “Moderate”, but the very concept of usability gives us reason to expect that inconvenience is enough to impede learning from past incidents. The results of this survey suggest that entire missing categories of incidents and entire categories of false positives are both relatively minor problems, but still do not answer the question of which of these attitudes, if either, is preferable in general. Certainly, reporting of a large number of important incidents is ideal, but if an organization *must* choose between false alarms and misses, which situation would be more favorable to learning from incidents?

## **Conclusions: Specific Recommendations for Improvement of the Reporting App and Organization**

### **Recommendation 1: Revising reporting interface menu options**

Interviews and usability testing provided evidence that reporting interface menus contain confusing items that impact interrater reliability and such a large total number of items that navigation is difficult. It seems obvious that the path to addressing this is to examine users’ suggestions for better menu options, but there are serious obstacles and constraints to this approach. Legal or accreditation issues preclude altering the menu

options too much, as it would impair the hospital's ability to report accurate statistics to regulatory bodies and quality tracking projects like NDNQI, the Joint Commission's reporting system, and NYPORTS. A review of the information required by those bodies would produce a list of "necessary menu items", and others may be combined, split, or renamed as appropriate.

However, the survey data suggest that there is no consensus among staff about what additional menu options would actually be helpful for their reporting, and adding all of the options suggested would likely overall reduce reporting, considering that about half of UPs (52.46%) and APs (47.17%) already said the app was tedious, and a large majority of UPs (77.78%) said they at least sometimes have difficulty finding the existing menu items. Adding additional items may not make navigation more difficult. Indeed, there is very little difference in the time needed to locate the fifth or the hundredth item on a drop-down menu as long as the contents of the menu are familiar (Cheng & Patterson, 2002; Fischer & Schwan, 2008). However, adding items based on only a single request is a questionable use of the time and effort of those who can revise the menus. It may be better to follow the lead of the one survey participant who called for more opportunities to offer answers of "other", or perhaps even to give users the explicit option to leave the classification of incidents to the quality coordinators if they cannot decide themselves how to classify them.

If altering existing menu options for clarity is chosen as a means for improving, perhaps it would be useful to create a reference chart of "translations" between internally-used categories and regulatorily-required ones. In one variation on this idea, users could use such a chart to look up what the official category label is for the type of incident they are trying to describe. Alternatively, the menu options themselves could be changed so that users could describe incidents on their own terms, and the charts would be used by patient safety analysts and quality coordinators to convert internal reports to a format intelligible to regulatory and other outside organizations. Either of these options, however, entails not only additional time to report or reclassify incidents, but a particularly tedious kind of additional work. A more modest alternative, then, is to examine the common types of

changes already being made by quality coordinators and creating a reference chart of rules for distinguishing between the most commonly-confused labels.

Given the importance of time constraints as a source of underreporting, I recommend adopting approaches that seem least likely to increase the time needed to make reports. First, to gather potential new options, introduce an “other” option where one is not currently available, review the reports marked “other” for common trends, and add new menu options that adequately describe the common types of “other” reports. Second, to aid users in distinguishing between confusing categories (i.e., those categories which are most frequently erroneously chosen and changed by quality coordinators), track quality coordinators’ changes to categorization of incoming incident reports, determine the most confusing categories, and create a reference chart with rules for distinguishing between them. Grouping together the most commonly-used items at the top of each menu may additionally be helpful for reducing the number of users who cannot find appropriate items, and will likely have little practical effect on time demands (Omanson, Miller, & Joseph, 2014), but users will require some time to adapt to the new format and will likely have a temporarily higher error rate as they learn the new format, so the menu should not be changed frequently to avoid user confusion.

A final option, which would be directed at the developer of the app as a consideration for future versions of the app rather than at those hospital staff empowered to customize the app already in use, would be a change to the menu layout itself. Given that some of the drop-down menus contain a very large number of items, an alphabetized grid like that used by Cheng and Patterson (2002) would be a welcome improvement. Such a menu setup would make the task of scanning a long list of items easy and eliminate the potentially annoying need to scroll down a long single column.

### **Recommendation 2: Investigating and remedying punitive use**

Overall, punitive use scored near the middle of the priority rankings, but it stood out as an unusual exception to a general trend of positive safety culture, and especially so

because it was concentrated in units engaged in a particular specialty. Not only because it is corrosive to a positive work environment and discourages reporting, but also because it encourages poor-quality reporting and therefore hinders attempts to improve, allegations of punitive use should be examined seriously and carefully by the hospital administration.

Punitive use of the reporting system may be considered a form of workplace aggression, and people who engage in this kind of behavior are somewhat more likely to have a host of psychological factors such as highly negative attitudes about others' perceived advantages and poor "political skills", the ability to use one's influence to further organizational goals (Harris, Harvey, & Booth, 2010). In this case, a possible method for correction could be to consistently respond to misuses of the system with reminders (directed at the entire unit, not the individual, or else the anonymity of the reporting system would be compromised) about the importance of responsible reporting, focusing on the fact that punitive reports can erode safety and other long-term goals of the organization by eroding cooperation and information-sharing.

If, on the other hand, punitive use doesn't seem to consist of targeted acts of retaliation but is more like ordinary incivility rather than like aggression, there is another avenue for improvement. While workplace aggression is apparently due to a mixture of interpersonal conflicts and predispositions for confrontational behavior, incivility may not even be intentionally directed at anyone in particular. Especially since staff in the affected departments already reported generally positive opinions of their supervisors, supervisors can step in to improve staff morale by reminding them that they are valued and cared for. Although this approach is unlikely to prevent punitive use, it could at least provide staff with the additional encouragement needed to prevent the victims of punitive use from becoming unproductive or quitting (Sakurai, 2011).

### **Recommendation 3: Expanding and improving the use of time-saving shortcuts**

Users and analysts alike identified time constraints, trade-offs, and interruptions as important discouragements or barriers to using the incident reporting app. Ideally, such a

problem would be remedied by setting aside time during the workday for reporting, but this intervention would be best implemented by hiring additional staff to reduce the individual's workload, which would be expensive, or by extending the workday itself, which would not only be expensive but on top of it irritating to the staff. Instead, I recommend that the hospital examine ways to reduce the time needed to make each individual report.

Although the participants interviewed gave positive reviews of the the shortcuts implemented to reduce time demands, determining whether the shortcuts are actually beneficial would require comparing the followups done on incidents that were otherwise similar except for how they were reported – which was not allowed in this study due to legal and privacy concerns about allowing access to complete incident reports. Hospital staff who are allowed to access such information should identify incident reports from the units that use shortcuts and compare those made with and without shortcuts to determine whether shortcut reports lack necessary information for following up appropriately as was the case in the freeform incident report forms examined by Pettersson (2013a, 2013b).

If the shortcut reports are still sufficiently detailed for followup, the existing shortcuts should remain in place, and other units that have the highest workloads should adopt similar shortcuts of their own to reduce the time demands of the reporting process. If the shortcut reports are less detailed in ways that impede followup (as seems likely based on form design guidelines and Pettersson's incident report redesign), any existing and future shortcuts should be revised so that all necessary information is prompted. It is probably safe, for example, to omit the "incident category" prompt currently present in the reporting interface from a shortcut form, as it can be easily determined from the details of the narrative, but omitting details so seemingly simple as the time of the incident or the age of the patient could lead unit managers to make uninformed, incomplete, or irrelevant decisions based on mistaken assumptions about the conditions of the particular incident. The hotline in use in one department is particularly worrisome as far as potential for lost information, as prompts on a written or computerized form may be revisited at any time or addressed out of order if needed, while voice prompts cannot.



**Recommendation 4: Improving the flow of information**

The flow of information about incidents and the followups to them has some limitations, largely involving a lack of feedback and coordination. At the time of data collection, little information was available about the newly-introduced feedback system, but it centers around sharing information about recent incidents directly between unit managers. This directly remedies analysts' highest-priority problems, which dealt with learning from others units' successful interventions, and a lower-priority problem, making changes multiple units at once.

Such a system seems highly promising and should be given a chance to work before any additional changes are made, but it is clear immediately that certain aspects of the hospital's information flow are beyond its scope. Unlike a *Callback*-style newsletter, it would not provide specific feedback to those who reported incidents or allow for long-term overviews of how a particular category of incidents has been addressed. This means that, even if coordination and information-sharing between departments is successful, it will not necessarily be obvious to ordinary staff that their incident reports contributed to the success. A lack of clear examples of how their input helped plausibly explains why users reported commonly ( $U = 0.26$ ) not understanding the interventions taken in response to incidents. In other words, this is currently a system with inconsistent process feedback and low face validity, which reduces users' trust that the system is useful for making decisions at all (de Vries, van den Berg, & Midden, 2015). Another approach is necessary to provide users with the kind of feedback that makes them feel like they are contributing to a useful project.

The feedback system also seems geared only towards addressing recent and notable incidents. Though this may be very helpful for new or currently-critical issues, time will tell how useful it is for infrequent or ongoing ones. In either case, good record-keeping would be helpful for the overall goal of finding successful ideas and spreading them quickly. Analysts should be required to enter descriptions of their responses to incidents into the reporting app's followup interface not for their own future reference for the future reference of anyone who needs to address a similar incident in the future.

A newsletter themed around specific topics would fill in the gaps, allowing everyone from administration to care providers access to information about the way incidents are being handled and how incident reporting has contributed to it. Flowcharts showing the proposed improved flow of information may be seen in Figures 4 and 5. I admit uncertainty about whose responsibility the newsletter should be, but due to their relevant background knowledge and access to the necessary information, quality coordinators are an obvious choice. This newsletter can make its point briefly – *Callback* was originally only a single page (Reynard et al., 1986, p. 76) and is still only two pages long – and releasing it on a monthly schedule allows for a longer perspective than the recently-introduced unit manager meetings seem designed for. The “highlights” of those meetings or appeals to suggest solutions to unsolved problems could also be included, demonstrating clearly to the typical user how their incident reports are used.

#### **A final general advisory**

Both the hospital with which I worked and any healthcare professionals seeking to apply the lessons learned here to their own facilities should apply the above proposed improvements together, not pick and choose among them. Better flow of information is not useful if the information itself is not useful, and useful reports are not likely when staff are too busy, confused, or guarded to make reports that contain all the information analysts want. Increased sharing of incomplete or misleading information will not improve patient safety, but instead simply be the classic problem of “garbage in, garbage out”.

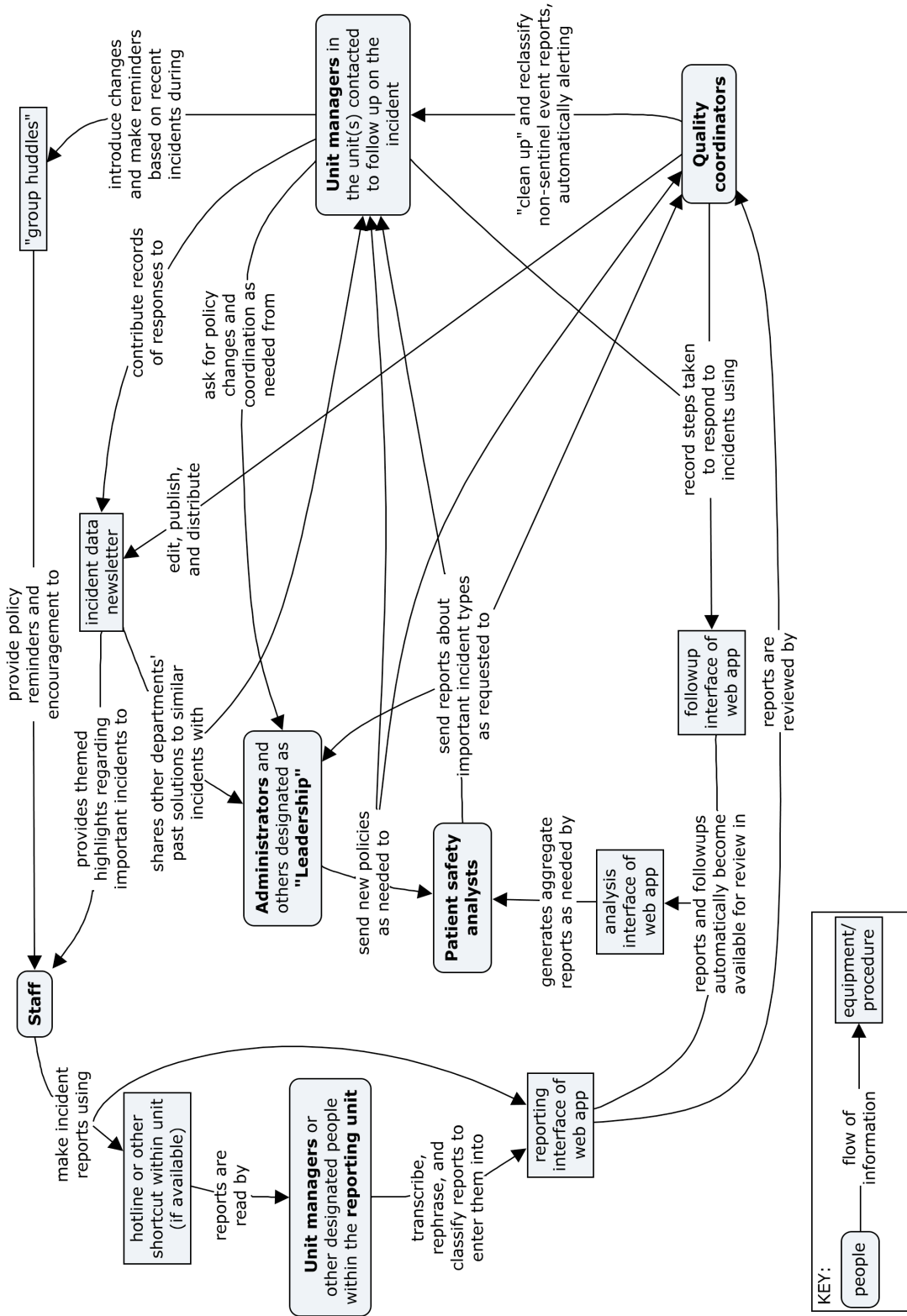


Figure 4. Flowchart showing the goal of improving the cycle of incident report information.

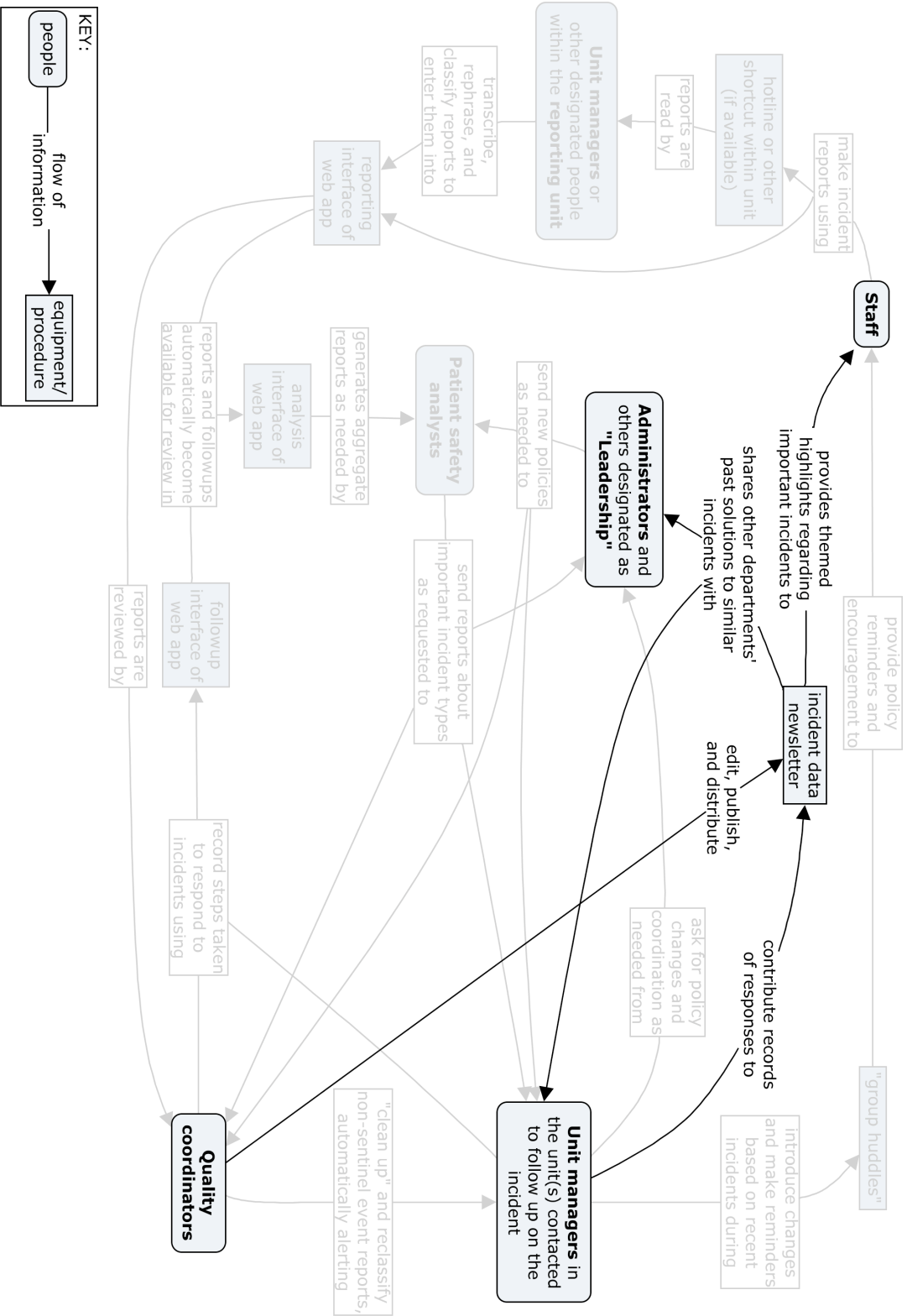


Figure 5. The flowchart seen in Figure 4, recolored to emphasize the proposed creation of a Callback-style newsletter.

## Appendix A: Interview Questions

*Notes: In the form used to collect interview data, each set of questions was followed by blank space for notes, in case the interview was not recorded or information was best conveyed in a visual format, e.g. a timeline. The last section, labeled “Post-Test”, was administered after the participant completed the usability test. If the participant did not participate in that portion of the study, they were asked only the very last question from that section.*

### Preliminary

Before we begin, I'd like to ask about your specific job responsibilities in case this is relevant to how incidents get reported.

#### Demographics

- What is your primary work area or unit in this hospital?
- What is your job title?

*(request further clarification if needed for either of the above)*

- How often do you use [the reporting app]?

### Part 1

Now I'd like to move on to some questions about how you and your colleagues feel about incident reporting.

#### Effective Use

- Do you think mistakes have led to positive change?
- What does your supervisor or manager do about suggestions to improve safety?
- Is your unit good at avoiding errors?

- What makes people not want to use [the reporting app]?

*FOR ANALYSTS ONLY:*

- What do your colleagues want you and the incident reporting system to do for them?
- What do you want out of a good incident reporting system?

### **Encouragement**

- Do people generally support their coworkers here?
- Are you comfortable reporting incidents?
- Do you get positive feedback when you follow safety procedures?
- Do you get informed when incident reports lead to policy changes?
- To what extent does management succeed at getting people to use [the reporting app]?
- What kind of support do people need to report incidents, and from whom?
- What feedback do people need from the system to use it effectively?

### **Punishment**

- Do staff feel like their mistakes are held against them?
- Does it feel like the problem or the person is being reported?
- What information does [the reporting app] require that might make someone not want to use it?

### **Time Pressure**

- Do you feel expected to do too much too quickly to have time for incident reporting?
- What might people not be able to do anymore if they made more incident reports?

Now let's move on to a specific case. I want to know about your experience with this.

### **Selection**

- Give me an example of a time you've decided not to report an incident.

*If needed, clarify that this can mean:*

- *avoiding or cancelling a report*
- *an incident committed or just observed*
- *a near miss or actual harm (but not a sentinel event)*
- *unreported for absolutely any reason*

### **Timeline**

- Ask what led to the decision to avoid or cancel the report.
- Ask for additional decisions and arrange chronologically.
- Confirm that timeline is accurate.

### **Deepening**

- Confirm how typical/atypical these decisions were.
- Confirm how typical/atypical these circumstances were.
- How long did it take you to make this decision?
- Did you get any guidance or advice about this? Why?
- What other possible responses did you consider? How did you go about choosing not to report?
- Did you get any guidance or advice about this? Why?
- What other goals were competing with this? What was most important to you at the time?

*Again, don't forget to ask for clarification as needed!*

### **Part 2b**

Let's try a different example.

#### **Selection**

- Give me an example of a time you **have** reported an incident.

#### **Timeline**

- Ask what led to the decision to make the report.
- Ask for additional decisions and arrange chronologically.
- Confirm that timeline is accurate.

#### **Deepening**

- Confirm how typical/atypical these decisions were.
- Confirm how typical/atypical these circumstances were.
- How long did it take you to make this decision?
- Did you get any guidance or advice about this? Why?
- What other possible responses did you consider? How did you go about choosing to report?
- Did you get any guidance or advice about this? Why?
- What other goals were competing with this? What was most important to you at the time?

*Again, don't forget to ask for clarification as needed!*

#### **Post-Test**



- Do you feel like this reflects reasonably well how you would make a real incident report?
- What, if anything, is different? *esp. scenario details & time to complete task*
- Can you think of any other problems using [the reporting app] you've run into that didn't happen just now?

Is there anything else you'd like to say or ask about the reporting process or this study?

### Appendix B: Usability Test Scenario

**Please read the following narrative carefully. Then, using the [reporting app] training module, please create an incident report about it. Feel free to narrate your thoughts about the process of using [the reporting app] as you work, and to compare this scenario to actual incidents you experienced or know about if any examples come to mind.**

Recently, due to increased workload from a flu outbreak, outside staff have been brought in and existing staff have been shuffled around. This has led to a number of “near misses” and minor incidents where patients could have been harmed seriously because staff aren’t fully trained in [the hospital]’s safety procedures or used to the department in which they are currently working. You overhear some coworkers talking about the problems they’ve run into.

One of them, a nurse, explains a recent scenario that could have been very dangerous if she hadn’t noticed and fixed it. A few days ago, a 56-year-old man was admitted to the hospital with nausea and severe pain in his abdomen. It turned out he had been drinking heavily. He was diagnosed with alcohol-induced pancreatitis and has been in the ICU for several days. While he has been here, he has suffered serious complications, including failure of other organs.

As part of his treatment, he needed to have a saline solution infused directly into an artery. However, it appears that the last nurse who changed the bag before your coworker’s shift attached the wrong one. Because of similar labels on different bags of fluids, they had attached a bag containing a solution of both 0.9% sodium chloride and 5% glucose instead of the bag of only sodium chloride solution that was needed.

Luckily, when it was your coworker's turn to check up on the patient, she noticed the mention of glucose on the label for the bag, realized it was the wrong bag, and swapped it out for a bag of the correct solution. Your coworker says this decision saved the patient from what could have been additional organ damage and even death, but that she's too busy to make the report so you decide to do it yourself based on her description of the event.

### Appendix C: Survey Questions

*Notes: Because this survey was administered online, formatting of the actual survey was somewhat different from what is included here. The content of all questions, however, is identical. Questions without options listed and those asking the participant to provide specific examples were free-response.*

#### Questions asked to all participants

##### Before main survey:

- Which of these best describes your role at [the hospital]?
  - Physician
  - Nurse
  - Manager of a medical unit
  - Other non-management medical role
  - Manager of a non-medical unit
  - Other non-medical role
  - Other (please specify)
- What is your primary unit or work area?
- How long have you worked at [the hospital]?
- How long have you worked in healthcare overall?
- In the past 12 months, how many incident reports have you filed?
  - None
  - 1–2
  - 3–5
  - 6–10

11–20

21 or more

- Do you analyze incident reports that others submit?

Yes

No

**After main survey:**

- Do you have any additional comments about incident reporting?

**Main Survey: Analysts' Version**

**Please indicate how much you agree or disagree with the following.**

- My unit is good at avoiding mistakes.
- Mistakes lead to positive change.
- It's helpful to report near-misses.
- My supervisor/manager responds helpfully to suggestions for how to improve safety.
- People are making too many reports about things that don't really count as "incidents".
- People are making too many reports about things that are not in my power to fix.
- I wish I knew more about how other units are addressing incidents.
- If something goes wrong with [the reporting app], I can easily find out how to fix it or work around it.
- I wish my unit had a faster way of reporting incidents.
- Using [the reporting app] is tedious.

- I have to do too much too quickly.
- If I made more thorough analyses, I would not have time to do other important parts of my job.
- There are certain things I am sure are problems, but I can't fix them because they don't get reported.
- Think about the item "there are certain things I am sure are problems, but I can't fix them because they don't get reported" from the list above. If you said "Agree" or "Strongly Agree", please provide an example.

**Please indicate how commonly or rarely the following things happen.**

- When another unit solves a problem my unit is also having, I can easily find out about it and try using their solution.
- People make incident reports to try to get each other in trouble.
- Incident reports are phrased like the reporter is hiding something.
- When I use [the reporting app], I have trouble finding the items I need on menus.
- My analyses get ruined by [the reporting app] having an error or crashing.
- I get interrupted during incident report analyses and end up not finishing them.
- I can come up with ways to address incidents that make similar incidents less common or less harmful.
- If an incident is due to problems in more than one unit, I can get everyone to make the necessary changes.
- I give up on analyzing an incident report because of [the reporting app] having an error or crashing.
- I get confused by the way incidents are classified in the reports I receive.

- I don't get enough information in incident reports to figure out how to respond to them.
- Think about the item "I get confused by the way incidents are classified in the reports I receive" from the list above. If this has happened to you, please provide an example.
- Think about the item "I don't get enough information in incident reports to figure out how to respond to them" from the list above. If this has happened to you, please provide an example.

### **Main Survey: Users' Version**

**Please indicate how much you agree or disagree with the following.**

- My unit is good at avoiding mistakes.
- Mistakes lead to positive change.
- It's helpful to report near-misses.
- My supervisor/manager responds helpfully to suggestions for how to improve safety.
- I am comfortable reporting incidents.
- I get positive feedback when I follow safety procedures.
- If nobody has fixed an ongoing problem, I will make sure to report it the next time it happens to try to get someone to fix it.
- My mistakes are held against me.
- Anonymous reports aren't really anonymous.
- [the reporting app] asks me for information I am not comfortable providing.
- I tend to report as much as I can, even if it means making some reports that aren't really necessary.

- I tend to focus on the incidents I think are most important, even if it means not reporting some other incidents.
- If something goes wrong with [the reporting app], I can easily find out how to fix it or work around it.
- I wish my unit had a faster way of reporting incidents.
- Using [the reporting app] is tedious.
- I have to do too much too quickly.
- If I made more reports, I would not have time to do other important parts of my job.
- The way [the reporting app] asks me to classify incidents is confusing.
- Some common types or details of incidents are not included on [the reporting app]'s drop-down menus, but should be.
- Think about the item “the way [the reporting app] asks me to classify incidents is confusing” from the list above. If you said “Agree” or “Strongly Agree”, please provide an example.
- Think about the item “some types or details of incidents aren’t already included on [the reporting app]'s drop-down menus, but they are so common they should be” from the list above. If you said “Agree” or “Strongly Agree”, please provide an example.

**Please indicate how commonly or rarely the following things happen.**

- When we change how we do something because of an incident, the change makes sense to me.
- People make incident reports to try to get each other in trouble.
- When I use [the reporting app], I have trouble finding the items I need on menus.
- My incident reports get ruined by [the reporting app] having an error or crashing.



- I get interrupted while making an incident report and end up losing the report.
- I try to address incidents without reporting them by talking directly with whomever was involved.
- I give up on reporting a particular incident because of [the reporting app] having an error or crashing.

## Appendix D: RIT IRB Approval of Project

R·I·T

Form C  
IRB Decision Form

## Rochester Institute of Technology

RIT Institutional Review Board for the  
Protection of Human Subjects in Research  
141 Lomb Memorial Drive  
Rochester, New York 14623-5604  
Phone: 585-475-7673  
Fax: 585-475-7990  
Email: hmfrsr@rit.edu

**TO:** Paul Wiele  
**FROM:** RIT Institutional Review Board  
**DATE:** March 5, 2015  
**RE:** Decision of the RIT Institutional Review Board

Project Title – Incident Reporting Systems: Usability of Input Interfaces and Resulting Data

The Institutional Review Board (IRB) has taken the following action on your project named above.

Approved, no greater than minimal risk

Now that your project is approved, you may proceed as you described in the Form A. Note that this approval is only for a maximum of 12 months; you may conduct research on human subjects only between the date of this letter and March 5, 2016.

You are required to submit to the IRB any:

- Proposed modifications and wait for approval before implementing them,
- Unanticipated risks, and
- Actual injury to human subjects.

Return the Form F, at the end of your human research project or 12 months from the above date. If your project will extend more than 12 months, your project must receive continuing review by the IRB.

Continuing review of research and approval of research studies is required so long as the research study is ongoing, that is, until research-related interactions and interventions with human subjects or the obtaining and analysis of identifiable private information described in the IRB-approved research plan have been completed.

Investigators are responsible for submitting sufficient materials and information for the IRB to meet its regulatory obligations, and should follow the institutional policies and procedures for continuing IRB review of research that are required by HHS regulations at (45 CFR 46.103(b)(4), 45 CFR 46.109(e), 45 CFR 46.115(a)(1)) as appropriate to the research activity.



Heather Foti, MPH  
Associate Director  
Office of Human Subjects Research

## References

- AHRQ. (2004). *Hospital survey on patient safety. [Data collection instrument]*. Rockville, MD. Retrieved from <http://www.ahrq.gov/professionals/quality-patient-safety/patientsafetyculture/index.html>
- AHRQ. (2013). *Interim update on 2013 annual hospital-acquired condition rate and estimates of cost savings and deaths averted from 2010 to 2013 (Report No. 15-0011-EF)*. Retrieved from <http://www.ahrq.gov/professionals/quality-patient-safety/pfp/interimhacrate2013.pdf>
- Barach, P., & Small, S. (2000). Reporting and preventing medical mishaps: Lessons from non-medical near miss reporting systems. *British Medical Journal*, *320*, 759–763.
- Bargas-Avila, J., Brenzikofer, O., Roth, S., Tuch, A., Orsini, S., & Opwis, K. (2010). Simple but crucial user interfaces in the world wide web: Introducing 20 guidelines for usable web form design. In R. Mátrai (Ed.), *User interfaces* (pp. 1–10). Rijeka, Croatia: InTech.
- Beaubien, J. M., & Baker, D. P. (2002). A review of selected aviation human factors taxonomies, accident/incident reporting systems and data collection tools. *International Journal of Applied Aviation Studies*, *2*(2), 11–36.
- Billings, C. (1998). Incident reporting systems in medicine and experience with the aviation safety reporting system. In R. Cook, D. Woods, & C. Miller (Eds.), *A tale of two stories: contrasting views of patient safety. Report from a workshop on assembling the scientific basis for progress on patient safety* (pp. 52–61). McLean, VA: National Patient Safety Foundation at the AMA.
- Bleich, S. (2005). Medical errors: five years after the iom report. *Issue Brief (Commonwealth Fund)*, *830*, 1–15.
- Braddock, I., C.H., Szaflarski, N., Forsey, L., Abel, L., Hernandez-Boussard, T., & Morton, J. (2014). The TRANSFORM patient safety project: A microsystem approach to improving outcomes on inpatient units. *Journal of General Internal Medicine*, *4*, 425–433.

- Browne, G., & Rogich, M. (2001). An empirical investigation of user requirements elicitation: Comparing the effectiveness of prompting techniques. *Journal of Management Information Systems*, *17*, 223–249.
- Busse, D., & Wright, D. (2000). Classification and analysis of incidents in complex medical environments. *Topics in health information management*, *20*(4), 1–11.
- Capuzzo, M., Nawfal, I., Campi, M., Valpondi, V., Verri, M., & Alvisi, R. (2005). Reporting of unintended events in an intensive care unit: Comparison between staff and observer. *BMC Emergency Medicine*, *5*(3). Retrieved from <http://www.biomedcentral.com/1471-227X/5/3>
- Cheng, H., & Patterson, P. (2002). The grid menu: Efficient and robust selection of menu-items. In *Proceedings of the Human Factors and Ergonomics Society 46th annual meeting* (pp. 1281–1285). Thousand Oaks, CA: SAGE Publications.
- Commission on Accreditation of Medical Transport Systems. (2012). *Safety culture survey. [survey instrument]*. Retrieved from <http://www.camts.org/Safety-Culture-Survey.html>
- Cook, A., & Herout, J. (2015). Developing a usability ranking system for findings in health information technology products. In *Proceedings of the international symposium on human factors and ergonomics in health care* (Vol. 4, pp. 23–28). Thousand Oaks, CA: SAGE Publications.
- Crandall, B., Klein, G. A., & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis*. MIT Press.
- Deeter, J., & Rantanen, E. (2012). Human reliability analysis in healthcare. In *Proceedings of the 2012 symposium on human factors and ergonomics in health care: Bridging the gap* (pp. 45–51). SAGE Publications.
- de Vries, P., van den Berg, S., & Midden, C. (2015). Assessing technology in the absence of proof: Trust based on the interplay of others' opinions and the interaction process. *Human Factors*, *57*, 1378–1402.
- Dillman, D. (2000). *Mail and internet surveys: the tailored design method* (2nd ed.). New

York, NY: John Wiley & Sons.

- Erler, C., Edwards, N. E., Ritchey, S., Pesut, D. J., Sands, L., & Wu, J. (2013). Perceived patient safety culture in a critical care transport program. *Air medical journal*, *32*(4), 208–215.
- Feldman, S., Schooley, B., & Bhavsar, G. (2014). Health information exchange implementation: Lessons learned and critical success factors from a case study. *JMIR Medical Informatics*, *2*(2), e19.
- Fischer, S., & Schwan, S. (2008). Adaptively shortened pull down menus: location knowledge and selection efficiency. *Behaviour Information Technology*, *27*, 439–444.
- Flaaten, H., & Hevr, O. (1999). Errors in the intensive care unit (ICU): Experiences with an anonymous registration. *Acta Anaesthesiologica Scandinavica*, *43*, 614–617.
- Frazer, E. (2011). How do we assess safety culture? *Air medical journal*, *30*(2), 62–63.
- Gupta, K., & Cook, T. (2013). Accidental hypoglycaemia caused by an arterial drug flush error: a case report and contributory causes analysis. *Anaesthesia*, *68*, 1179–1187.
- Harris, K., Harvey, P., & Booth, S. (2010). Who abuses their coworkers? an examination of personality and situational variables. *The Journal of Social Psychology*, *150*, 608–627.
- Healey, M., Shackford, S., Osler, T., Rogers, F., & Burns, E. (2002). Complications in surgical patients. *Archives of surgery*, *137*, 611–618.
- Hutchinson, A. M., Batra-Garga, N., Cranley, L., Boström, A., Cummings, G., Norton, P., & Estabrooks, C. A. (2012). Feedback reporting of survey data to healthcare aides. *Implement Sci*, *7*, 89.
- Hvannberg, E., Law, E., & Lárusdóttir, M. (2007). Heuristic evaluation: Comparing ways of finding and reporting usability problems. *Interacting with Computers*, *19*, 225–240.
- IOM. (1999). *To Err is Human: Building a safer health system*. Washington, DC: National Academy Press.
- IOM. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: National Academies Press.
- IOM. (2004). *Patient safety*. Washington, DC: National Academies Press.

- ioSweep. (2015). *ioSweep*. Retrieved from <http://www.iosweep.com/>
- ISO. (2010). *Ergonomics of human-system interaction – part 210: Human-centred design for interactive systems (ISO 9241-210)*. Geneva, Switzerland: International Organization for Standardization.
- Jha, A., Kuperman, G., Teich, J., Leape, L., Shea, B., Rittenberg, E., ... Bates, D. W. (1998). Identifying adverse drug events: Development of a computer-based monitor and comparison with chart review and stimulated voluntary report. *Journal of the American Medical Informatics Association*, 5(3), 305–314.
- Johnson, C. (2002). Reasons for the failure of incident reporting in the healthcare and rail industries. In *Components of system safety: Proceedings of the 10th safety-critical systems symposium* (pp. 31–57). Springer.
- Joint Commission. (2011). *Facts about the sentinel event policy*.
- Joint Commission. (2013). *Sentinel event data: General information 1995-2012*.
- Joint Commission. (2016). *Sentinel event policy and procedures*. Retrieved from [http://www.jointcommission.org/sentinel\\_event\\_policy\\_and\\_procedures/](http://www.jointcommission.org/sentinel_event_policy_and_procedures/)
- Kagan, I., & Barnoy, S. (2013). Organizational safety culture and medical error reporting by Israeli nurses. *Journal of Nursing Scholarship*, 45(3), 273–280.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological bulletin*, 119(2), 254.
- La Pietra, L., Calligaris, L., Molendini, L., Quattrin, R., & Brusaferrò, S. (2005). Medical errors and clinical risk management: state of the art. *Acta Otorhinolaryngologica Italica*, 25(6), 339.
- Leape, L. (2002). Reporting of adverse events. *The New England Journal of Medicine*, 347, 1633–1638.
- Liljegren, E. (2006). Usability in a medical technology context assessment of methods for usability evaluation of medical equipment. *International Journal of Industrial Ergonomics*, 36, 345–352.

- Marchon, S., & Mendes, W. (2014). Patient safety in primary health care: A systematic review. *Cadernos de saúde pública*, *30*, 1815–1835.
- Mazzoni, G., & Vannucci, M. (2007). Hindsight bias, the misinformation effect, and false autobiographical memories. *Social Cognition*, *25*(1), 203–220.
- Milstead, J. (2005). The culture of safety. *Policy, Politics, Nursing Practice*, *6*, 51–54.
- Montalvo, I. (2007). The National Database of Nursing Quality Indicators®(NDNQI®). *The Online Journal of Issues in Nursing*, *12*(3). Retrieved from <http://www.nursingworld.org/MainMenuCategories/ANAMarketplace/ANAPeriodicals/OJIN/TableofContents/Volume122007/No3Sept07/NursingQualityIndicators.aspx>
- Nadel, L., Campbell, J., & Ryan, L. (2007). Autobiographical memory retrieval and hippocampal activation as a function of repetition and the passage of time. *Neural Plasticity*, Article ID 90472.
- NASA. (2012). *ASRS Program Briefing 2012* (Tech. Rep.). Washington, DC: National Aeronautics and Space Administration.
- Nielsen, J. (1994). Usability inspection methods. In J. Nielsen & R. Mack (Eds.), (pp. 25–62). New York, NY: John Wiley & Sons.
- Norman, G. (2010). Likert scales, levels of measurement, and the ‘laws’ of statistics. *Advances in Health Sciences Education*, *15*, 625–632.
- Omanson, R., Miller, C., & Joseph, S. (2014). Effects of item grouping on selection efficiency. In *Proceedings of the Human Factors and Ergonomics Society 58th annual meeting* (pp. 1526–1530). Thousand Oaks, CA: SAGE Publications.
- Outware Mobile. (2013). *Snap send solve*. Retrieved from [www.snapsendsolve.com](http://www.snapsendsolve.com)
- Pear, R. (2012, January 6). Report finds most errors at hospitals unreported. *New York Times*, A12.
- Petterson, U. (2013a). A form to collect incident reports: Learning from incidents in the Swedish armed forces. *The Electronic Journal of Knowledge Management*, *11*(2), 150–157.
- Petterson, U. (2013b). *Improving incident reports in the Swedish armed forces* (Unpub-

- lished doctoral dissertation). Lund University, Lund, Sweden.
- Reynard, W., Billings, C., Cheaney, E., & Hardy, R. (1986). *The development of the NASA aviation safety reporting system* (Tech. Rep. No. NASA Reference Publication No. 1114; NASA ASRS Publication No. 60). Moffett Field, CA: National Aeronautics and Space Administration.
- Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing* (Second Edition ed.). Indianapolis, IN: John Wiley & Sons.
- Sakurai, K. (2011). *Coworker incivility and incivility targets' work effort and counterproductive work behaviors: The moderating role of supervisor social support* (Unpublished doctoral dissertation). Bowling Green State University, Bowling Green, OH.
- Snowdon, P., & Johnson, C. (1999). Results of a preliminary survey into the usability of accident and incident reports. In M. Bransby & J. Noyes (Eds.), *People in control: An international conference on human interfaces in control rooms, cockpits, and command centres* (pp. 258–262). London, UK: Institution of Electrical Engineers.
- Sorra, J., Famolaro, T., Yount, N., Smith, S., Wilson, S., & Liu, H. (2014). *Hospital survey on patient safety culture: 2014 user comparative database report* (Tech. Rep.). Rockville, MD: Agency for Healthcare Research and Quality.
- State of New York Department of Health. (2007). *New York Patient Occurrence Reporting and Tracking System—NYPORTS—2005, 2006, 2007 Report* (Tech. Rep.). Albany, NY: State of New York Department of Health.
- Stelfox, H. T., Palmisani, S., Scurlock, C., Orav, E. J., & Bates, D. W. (2006). The “To Err is Human” report and the patient safety literature. *Quality and Safety in Health Care*, 15(3), 174–178.
- Tan, W., Liu, D., & Bishu, R. (2009). Web evaluation: Heuristic evaluation vs. user testing. *International Journal of Industrial Ergonomics*, 39, 621–627.
- Taylor, J. B. (2012). *Safety culture: Assessing and changing the behaviour of organisations*. Gower Publishing, Ltd.
- Thompson, L. F., & Surface, E. A. (2009). Promoting favorable attitudes toward personnel



- surveys: The role of follow-up. *Military Psychology*, *21*(2), 139.
- Throckmorton, T., & Etchegaray, J. (2007). Factors affecting incident reporting by registered nurses: The relationship of perceptions of the environment for reporting errors, knowledge of the Nursing Practice Act, and demographics on intent to report errors. *Journal of PeriAnesthesia Nursing*, *22*(6), 400–412.
- Timmel, J., Kent, P., Holzmueller, C., Paine, L., Schulick, R., & Pronovost, P. (2010). Impact of the Comprehensive Unit-Based Safety Program (CUSP) on safety culture in a surgical inpatient unit. *The Joint Commission Journal on Quality and Patient Safety*, *36*(6), 252–260.
- Tversky, B., & Marsh, E. (2000). Biased retellings of events yield biased memories. *Cognitive Psychology*, *40*, 1–38.
- Umansky, J. (2016). *Workload in nursing* (Unpublished master's thesis). Rochester Institute of Technology.
- Vollmer, C., Sanchez, N., Gondek, S., McAuliffe, J., Kent, T., Christein, J., . . . The Pancreatic Surgery Mortality Study Group (2011). A root-cause analysis of mortality following major pancreatectomy. *Journal of gastrointestinal surgery*, *16*, 89–103.
- Waring, J. (2005). Beyond blame: Cultural barriers to medical incident reporting. *Social Science Medicine*, *60*, 1927–1935.
- Westrum, R. (2004). A typology of organisational cultures. *Quality and Safety in Health Care*, *13*(Suppl. II), ii22–ii27.
- Wroblewski, L. (2008). *Web form design: Filling in the blanks*. New York, NY: Rosenfeld Media.